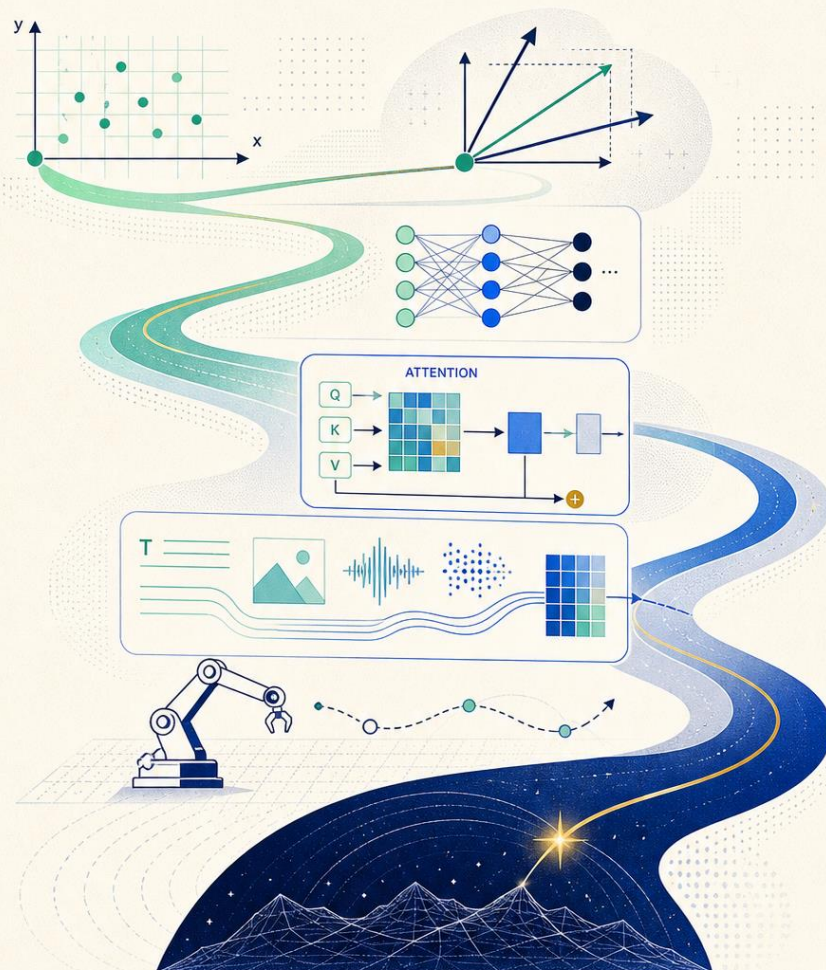


从零开始学AI

普通人与高中生的24个月互动教材

2026.08 Beta



NEWVAR

第一卷：计算与数学地基

版本、边界与使用方法

结论先行：这是一套面向普通人与高中生的系统教材，不是速成承诺。当前为 **2026.08 Beta**；只有技术评审、事实核查和领域专家签字全部完成后，才能进入 stable。

项目	内容
版本	2026.08 Beta
卷册	第一卷：计算与数学地基
本卷章节	24
全书结构	4 卷 · 24 单元 · 96 章
实践资产	96 个互动实验 · 24 个项目 · 288 道练习
正文规模	约 689,651 个汉字
核实日期	2026-08-06
版权	Copyright © 2026 NEWVAR LAB. All rights reserved.

能力终点

本书以独立开发、部署、论文阅读、实验复现、失败分析和研究问题提出为终点能力；不承诺仅靠一本书自动成为世界顶尖研究者。

全面覆盖的可审计边界

“全面”指对主要 AI 知识域、能力框架、工程链路和研究方法进行可审计覆盖，不等于收录互联网上每一页资料。

每章学习契约

- 先看结论、价值和完成后能做什么。
- 完成先修检查，用生活类比获得方向，再主动找出类比失效处。
- 一次只改变一个核心变量；网页提供键盘操作和静态替代说明。
- 按直觉、数学、最小实现、PyTorch、实验、失败、迁移的顺序推进。
- 工程线与研究线至少选择一条；三级练习先独立完成，再看答案。
- 所有结果必须标成实际测量、教学模拟或文献引用；不能混写。
- 遇到高风险决策、隐私数据或证据不足时停止自动化并交给给人。

隐私与材料规则

全书案例只使用虚构或公开数据。不要在练习、截图、Notebook 或网站笔记中输入姓名、账号、联系方式、密钥、医疗记录、成绩等个人信息。网页学习记录仅保存在当前设备，可匿名导入导出；不建账号、不做行为追踪。

目录

目录由 Word 字段生成。打开文件后若页码尚未刷新，请按 Ctrl+A，再按 F9。

版本、边界与使用方法..... 2

 能力终点..... 2

 全面覆盖的可审计边界..... 3

 每章学习契约..... 3

 隐私与材料规则..... 3

目录..... 4

本卷 24 章路线图..... 35

 月 1 · AI 全景、证据与伦理..... 35

 月 2 · Python 与计算思维..... 35

 月 3 · Git、数据、SQL 与可复现性..... 36

 月 4 · 线性代数..... 36

 月 5 · 概率统计与信息论..... 36

 月 6 · 微积分、优化和数值计算..... 37

第 1 章 AI 不是魔法：任务、模型与系统..... 38

 1. 为什么重要，现在能做什么..... 38

 2. 先修检查与生活类比..... 39

 3. 互动实验：只改变一个变量..... 39

 4. 直觉到数学..... 40

 5. 最小实现..... 40

6. 五轮系统化理解	41
1. 问题边界	41
3. 机制与因果链	42
7. 评估与不确定性	43
9. 工程与复现	45
11. 迁移与研究	46
7. PyTorch 实现路线	47
8. 实验设计与结果记录	47
9. 常见失败与排查顺序	48
10. 迁移任务	48
11. 三级练习	49
12. 本章能力检查	49
13. 来源与核实	49
第 2 章 从符号主义到基础模型：范式如何更替	51
1. 为什么重要，现在能做什么	51
2. 先修检查与生活类比	52
3. 互动实验：只改变一个变量	52
4. 直觉到数学	53
5. 最小实现	53
6. 五轮系统化理解	54
1. 问题边界	54

3. 机制与因果链.....	55
7. 评估与不确定性	56
9. 工程与复现	58
11. 迁移与研究	59
7. PyTorch 实现路线	60
8. 实验设计与结果记录	61
9. 常见失败与排查顺序	61
10. 迁移任务.....	62
11. 三级练习.....	62
12. 本章能力检查.....	62
13. 来源与核实	63
第 3 章 先问证据：怎样判断一个 AI 说法.....	64
1. 为什么重要，现在能做什么.....	64
2. 先修检查与生活类比	65
3. 互动实验：只改变一个变量.....	65
4. 直觉到数学	66
5. 最小实现.....	66
6. 五轮系统化理解	67
1. 问题边界.....	67
3. 机制与因果链.....	68
7. 评估与不确定性	69

9. 工程与复现	71
11. 迁移与研究	72
7. PyTorch 实现路线	73
8. 实验设计与结果记录	73
9. 常见失败与排查顺序	74
10. 迁移任务	74
11. 三级练习	75
12. 本章能力检查	75
13. 来源与核实	75
第 4 章 价值不是附录：伦理、劳动与公共影响	77
1. 为什么重要，现在能做什么	77
2. 先修检查与生活类比	78
3. 互动实验：只改变一个变量	78
4. 直觉到数学	79
5. 最小实现	79
6. 五轮系统化理解	80
1. 问题边界	80
3. 机制与因果链	81
7. 评估与不确定性	82
9. 工程与复现	83
11. 迁移与研究	85

7. PyTorch 实现路线	86
8. 实验设计与结果记录	86
9. 常见失败与排查顺序	87
10. 迁移任务	87
11. 三级练习	87
12. 本章能力检查	88
13. 来源与核实	88
月 1 阶段项目：给校园社团做一份 AI 机会与风险地图	90
交付物	90
验收方法	90
第 5 章 变量、类型与表达式：让计算机保存意义	91
1. 为什么重要，现在能做什么	91
2. 先修检查与生活类比	92
3. 互动实验：只改变一个变量	92
4. 直觉到数学	93
5. 最小实现	93
6. 五轮系统化理解	94
1. 问题边界	94
3. 机制与因果链	95
7. 评估与不确定性	96
9. 工程与复现	98

11. 迁移与研究	99
7. PyTorch 实现路线	100
8. 实验设计与结果记录	100
9. 常见失败与排查顺序	101
10. 迁移任务	101
11. 三级练习	102
12. 本章能力检查	102
13. 来源与核实	102
第 6 章 控制流、函数与测试：把大问题切小	104
1. 为什么重要，现在能做什么	104
2. 先修检查与生活类比	105
3. 互动实验：只改变一个变量	105
4. 直觉到数学	106
5. 最小实现	106
6. 五轮系统化理解	107
1. 问题边界	107
3. 机制与因果链	108
7. 评估与不确定性	109
9. 工程与复现	110
11. 迁移与研究	112
7. PyTorch 实现路线	113

8. 实验设计与结果记录	113
9. 常见失败与排查顺序	114
10. 迁移任务	114
11. 三级练习	114
12. 本章能力检查	115
13. 来源与核实	115
第 7 章 列表、字典、文件与模块：组织一批数据	116
1. 为什么重要，现在能做什么	116
2. 先修检查与生活类比	117
3. 互动实验：只改变一个变量	117
4. 直觉到数学	118
5. 最小实现	118
6. 五轮系统化理解	119
1. 问题边界	119
3. 机制与因果链	120
7. 评估与不确定性	121
9. 工程与复现	123
11. 迁移与研究	124
7. PyTorch 实现路线	125
8. 实验设计与结果记录	125
9. 常见失败与排查顺序	126

10. 迁移任务	126
11. 三级练习	127
12. 本章能力检查	127
13. 来源与核实	127
第 8 章 算法、复杂度与调试：找到慢和错的原因	129
1. 为什么重要，现在能做什么	129
2. 先修检查与生活类比	130
3. 互动实验：只改变一个变量	130
4. 直觉到数学	131
5. 最小实现	131
6. 五轮系统化理解	132
1. 问题边界	132
3. 机制与因果链	133
7. 评估与不确定性	134
9. 工程与复现	135
11. 迁移与研究	137
7. PyTorch 实现路线	138
8. 实验设计与结果记录	138
9. 常见失败与排查顺序	139
10. 迁移任务	139
11. 三级练习	139

12. 本章能力检查	140
13. 来源与核实	140
月 2 阶段项目：制作一个可测试的学习时间分配器	141
交付物	141
验收方法	141
第 9 章 命令行与 Git：给每次改变留下可回溯证据	142
1. 为什么重要，现在能做什么	142
2. 先修检查与生活类比	143
3. 互动实验：只改变一个变量	143
4. 直觉到数学	144
5. 最小实现	144
6. 五轮系统化理解	145
1. 问题边界	145
3. 机制与因果链	146
7. 评估与不确定性	147
9. 工程与复现	149
11. 迁移与研究	150
7. PyTorch 实现路线	151
8. 实验设计与结果记录	151
9. 常见失败与排查顺序	152
10. 迁移任务	152

11. 三级练习	153
12. 本章能力检查	153
13. 来源与核实	153
第 10 章 数据生命周期：采集、清洗、字典与许可	155
1. 为什么重要，现在能做什么	155
2. 先修检查与生活类比	156
3. 互动实验：只改变一个变量	156
4. 直觉到数学	157
5. 最小实现	157
6. 五轮系统化理解	158
1. 问题边界	158
3. 机制与因果链	159
7. 评估与不确定性	160
9. 工程与复现	162
11. 迁移与研究	163
7. PyTorch 实现路线	164
8. 实验设计与结果记录	164
9. 常见失败与排查顺序	165
10. 迁移任务	165
11. 三级练习	166
12. 本章能力检查	166

13. 来源与核实	166
第 11 章 SQL 与关系模型：让问题变成可核对查询.....	168
1. 为什么重要，现在能做什么.....	168
2. 先修检查与生活类比	169
3. 互动实验：只改变一个变量.....	169
4. 直觉到数学	170
5. 最小实现.....	170
6. 五轮系统化理解	171
1. 问题边界.....	171
3. 机制与因果链.....	172
7. 评估与不确定性	173
9. 工程与复现	174
11. 迁移与研究	175
7. PyTorch 实现路线	177
8. 实验设计与结果记录	177
9. 常见失败与排查顺序	177
10. 迁移任务.....	178
11. 三级练习.....	178
12. 本章能力检查.....	179
13. 来源与核实	179
第 12 章 环境、随机种子与实验记录.....	180

1. 为什么重要，现在能做什么	180
2. 先修检查与生活类比	181
3. 互动实验：只改变一个变量	181
4. 直觉到数学	182
5. 最小实现	182
6. 五轮系统化理解	183
1. 问题边界	183
3. 机制与因果链	184
7. 评估与不确定性	185
9. 工程与复现	186
11. 迁移与研究	188
7. PyTorch 实现路线	189
8. 实验设计与结果记录	189
9. 常见失败与排查顺序	190
10. 迁移任务	190
11. 三级练习	190
12. 本章能力检查	191
13. 来源与核实	191
月 3 阶段项目：建立一个可复现的公开数据小仓库	193
交付物	193
验收方法	193

第 13 章 向量：把方向、特征和相似度放进坐标.....	194
1. 为什么重要，现在能做什么.....	194
2. 先修检查与生活类比	195
3. 互动实验：只改变一个变量.....	195
4. 直觉到数学	196
5. 最小实现.....	196
6. 五轮系统化理解	197
1. 问题边界.....	197
3. 机制与因果链.....	198
7. 评估与不确定性	199
9. 工程与复现	200
11. 迁移与研究	201
7. PyTorch 实现路线	203
8. 实验设计与结果记录	203
9. 常见失败与排查顺序	203
10. 迁移任务.....	204
11. 三级练习.....	204
12. 本章能力检查.....	205
13. 来源与核实	205
第 14 章 矩阵：一次描述许多线性关系.....	206
1. 为什么重要，现在能做什么.....	206

2. 先修检查与生活类比	207
3. 互动实验：只改变一个变量.....	207
4. 直觉到数学	208
5. 最小实现.....	208
6. 五轮系统化理解	209
1. 问题边界.....	209
3. 机制与因果链.....	210
7. 评估与不确定性	211
9. 工程与复现	212
11. 迁移与研究	213
7. PyTorch 实现路线	215
8. 实验设计与结果记录	215
9. 常见失败与排查顺序	215
10. 迁移任务.....	216
11. 三级练习.....	216
12. 本章能力检查.....	217
13. 来源与核实	217
第 15 章 投影、特征方向与 PCA.....	218
1. 为什么重要，现在能做什么.....	218
2. 先修检查与生活类比	219
3. 互动实验：只改变一个变量.....	219

4. 直觉到数学	220
5. 最小实现.....	220
6. 五轮系统化理解	221
1. 问题边界.....	221
3. 机制与因果链.....	222
7. 评估与不确定性	223
9. 工程与复现	224
11. 迁移与研究	225
7. PyTorch 实现路线	226
8. 实验设计与结果记录	227
9. 常见失败与排查顺序	227
10. 迁移任务.....	228
11. 三级练习.....	228
12. 本章能力检查.....	228
13. 来源与核实	229
第 16 章 张量、广播与数值线性代数.....	230
1. 为什么重要，现在能做什么.....	230
2. 先修检查与生活类比	231
3. 互动实验：只改变一个变量.....	231
4. 直觉到数学	232
5. 最小实现.....	232

6. 五轮系统化理解	233
1. 问题边界	233
3. 机制与因果链	234
7. 评估与不确定性	235
9. 工程与复现	236
11. 迁移与研究	238
7. PyTorch 实现路线	239
8. 实验设计与结果记录	239
9. 常见失败与排查顺序	240
10. 迁移任务	240
11. 三级练习	240
12. 本章能力检查	241
13. 来源与核实	241
月 4 阶段项目：从零实现一个二维图像变换游乐场	242
交付物	242
验收方法	242
第 17 章 概率与随机变量：描述不知道的部分	243
1. 为什么重要，现在能做什么	243
2. 先修检查与生活类比	244
3. 互动实验：只改变一个变量	244
4. 直觉到数学	245

5. 最小实现.....	245
6. 五轮系统化理解	246
1. 问题边界.....	246
3. 机制与因果链.....	247
7. 评估与不确定性	248
9. 工程与复现	249
11. 迁移与研究	250
7. PyTorch 实现路线	252
8. 实验设计与结果记录	252
9. 常见失败与排查顺序	252
10. 迁移任务.....	253
11. 三级练习.....	253
12. 本章能力检查.....	254
13. 来源与核实	254
第 18 章 分布与贝叶斯更新	255
1. 为什么重要，现在能做什么.....	255
2. 先修检查与生活类比	256
3. 互动实验：只改变一个变量.....	256
4. 直觉到数学	257
5. 最小实现.....	257
6. 五轮系统化理解	258

1. 问题边界	258
3. 机制与因果链	259
7. 评估与不确定性	260
9. 工程与复现	261
11. 迁移与研究	262
7. PyTorch 实现路线	264
8. 实验设计与结果记录	264
9. 常见失败与排查顺序	264
10. 迁移任务	265
11. 三级练习	265
12. 本章能力检查	266
13. 来源与核实	266
第 19 章 估计、区间与假设检验	267
1. 为什么重要，现在能做什么	267
2. 先修检查与生活类比	268
3. 互动实验：只改变一个变量	268
4. 直觉到数学	269
5. 最小实现	269
6. 五轮系统化理解	270
1. 问题边界	270
3. 机制与因果链	271

7. 评估与不确定性	272
9. 工程与复现	273
11. 迁移与研究	274
7. PyTorch 实现路线	276
8. 实验设计与结果记录	276
9. 常见失败与排查顺序	276
10. 迁移任务	277
11. 三级练习	277
12. 本章能力检查	278
13. 来源与核实	278
第 20 章 熵、交叉熵、KL 与互信息	279
1. 为什么重要，现在能做什么	279
2. 先修检查与生活类比	280
3. 互动实验：只改变一个变量	280
4. 直觉到数学	281
5. 最小实现	281
6. 五轮系统化理解	282
1. 问题边界	282
3. 机制与因果链	283
7. 评估与不确定性	284
9. 工程与复现	285

11. 迁移与研究	286
7. PyTorch 实现路线	288
8. 实验设计与结果记录	288
9. 常见失败与排查顺序	288
10. 迁移任务	289
11. 三级练习	289
12. 本章能力检查	290
13. 来源与核实	290
月 5 阶段项目：做一个不确定性诚实的天气建议器	291
交付物	291
验收方法	291
第 21 章 导数、梯度与链式法则	292
1. 为什么重要，现在能做什么	292
2. 先修检查与生活类比	293
3. 互动实验：只改变一个变量	293
4. 直觉到数学	294
5. 最小实现	294
6. 五轮系统化理解	295
1. 问题边界	295
3. 机制与因果链	296
7. 评估与不确定性	297

9. 工程与复现	298
11. 迁移与研究	299
7. PyTorch 实现路线	301
8. 实验设计与结果记录	301
9. 常见失败与排查顺序	301
10. 迁移任务	302
11. 三级练习	302
12. 本章能力检查	303
13. 来源与核实	303
第 22 章 雅可比、海森与自动微分	304
1. 为什么重要，现在能做什么	304
2. 先修检查与生活类比	305
3. 互动实验：只改变一个变量	305
4. 直觉到数学	306
5. 最小实现	306
6. 五轮系统化理解	307
1. 问题边界	307
3. 机制与因果链	308
7. 评估与不确定性	309
9. 工程与复现	310
11. 迁移与研究	311

7. PyTorch 实现路线	313
8. 实验设计与结果记录	313
9. 常见失败与排查顺序	313
10. 迁移任务	314
11. 三级练习	314
12. 本章能力检查	315
13. 来源与核实	315
第 23 章 梯度下降、动量与 Adam	316
1. 为什么重要，现在能做什么	316
2. 先修检查与生活类比	317
3. 互动实验：只改变一个变量	317
4. 直觉到数学	318
5. 最小实现	318
6. 五轮系统化理解	319
1. 问题边界	319
3. 机制与因果链	320
7. 评估与不确定性	321
9. 工程与复现	322
11. 迁移与研究	323
7. PyTorch 实现路线	325
8. 实验设计与结果记录	325

9. 常见失败与排查顺序	325
10. 迁移任务	326
11. 三级练习	326
12. 本章能力检查	327
13. 来源与核实	327
第 24 章 数值稳定、条件数与非凸地形	328
1. 为什么重要，现在能做什么	328
2. 先修检查与生活类比	329
3. 互动实验：只改变一个变量	329
4. 直觉到数学	330
5. 最小实现	330
6. 五轮系统化理解	331
1. 问题边界	331
3. 机制与因果链	332
7. 评估与不确定性	333
9. 工程与复现	334
11. 迁移与研究	336
7. PyTorch 实现路线	337
8. 实验设计与结果记录	337
9. 常见失败与排查顺序	338
10. 迁移任务	338

11. 三级练习	338
12. 本章能力检查	339
13. 来源与核实	339
月 6 阶段项目：训练一条可视化的微型回归曲线	341
交付物	341
验收方法	341
独立练习答案	342
第 1 章答案 · AI 不是魔法：任务、模型与系统	342
基础题参考	342
应用题参考	342
挑战题参考	342
自评量规	343
第 2 章答案 · 从符号主义到基础模型：范式如何更替	343
基础题参考	343
应用题参考	343
挑战题参考	343
自评量规	344
第 3 章答案 · 先问证据：怎样判断一个 AI 说法	344
基础题参考	344
应用题参考	344
挑战题参考	345

自评量规.....	345
第 4 章答案 · 价值不是附录：伦理、劳动与公共影响	345
基础题参考	345
应用题参考	345
挑战题参考	346
自评量规.....	346
第 5 章答案 · 变量、类型与表达式：让计算机保存意义.....	346
基础题参考	346
应用题参考	346
挑战题参考	347
自评量规.....	347
第 6 章答案 · 控制流、函数与测试：把大问题切小.....	347
基础题参考	347
应用题参考	347
挑战题参考	348
自评量规.....	348
第 7 章答案 · 列表、字典、文件与模块：组织一批数据.....	348
基础题参考	348
应用题参考	348
挑战题参考	349
自评量规.....	349

第 8 章答案 · 算法、复杂度与调试：找到慢和错的原因.....	349
基础题参考	349
应用题参考	349
挑战题参考	350
自评量规.....	350
第 9 章答案 · 命令行与 Git：给每次改变留下可回溯证据.....	350
基础题参考	350
应用题参考	351
挑战题参考	351
自评量规.....	351
第 10 章答案 · 数据生命周期：采集、清洗、字典与许可.....	351
基础题参考	351
应用题参考	352
挑战题参考	352
自评量规.....	352
第 11 章答案 · SQL 与关系模型：让问题变成可核对查询	352
基础题参考	352
应用题参考	353
挑战题参考	353
自评量规.....	353
第 12 章答案 · 环境、随机种子与实验记录.....	353

基础题参考	353
应用题参考	354
挑战题参考	354
自评量规	354
第 13 章答案 · 向量：把方向、特征和相似度放进坐标	354
基础题参考	354
应用题参考	355
挑战题参考	355
自评量规	355
第 14 章答案 · 矩阵：一次描述许多线性关系	355
基础题参考	355
应用题参考	356
挑战题参考	356
自评量规	356
第 15 章答案 · 投影、特征方向与 PCA	356
基础题参考	356
应用题参考	357
挑战题参考	357
自评量规	357
第 16 章答案 · 张量、广播与数值线性代数	357
基础题参考	357

应用题参考	358
挑战题参考	358
自评量规.....	358
第 17 章答案 · 概率与随机变量：描述不知道的部分.....	358
基础题参考	358
应用题参考	359
挑战题参考	359
自评量规.....	359
第 18 章答案 · 分布与贝叶斯更新.....	359
基础题参考	359
应用题参考	360
挑战题参考	360
自评量规.....	360
第 19 章答案 · 估计、区间与假设检验.....	360
基础题参考	360
应用题参考	361
挑战题参考	361
自评量规.....	361
第 20 章答案 · 熵、交叉熵、KL 与互信息.....	361
基础题参考	361
应用题参考	362

挑战题参考	362
自评量规.....	362
第 21 章答案 · 导数、梯度与链式法则.....	362
基础题参考	362
应用题参考	363
挑战题参考	363
自评量规.....	363
第 22 章答案 · 雅可比、海森与自动微分	363
基础题参考	363
应用题参考	364
挑战题参考	364
自评量规.....	364
第 23 章答案 · 梯度下降、动量与 Adam	364
基础题参考	364
应用题参考	365
挑战题参考	365
自评量规.....	365
第 24 章答案 · 数值稳定、条件数与非凸地形	365
基础题参考	366
应用题参考	366
挑战题参考	366

自评量规.....	366
本卷来源台账.....	367
[S001] AI competency framework for students	367
[S002] Five Big Ideas in AI	367
[S003] Computer Science Curricula 2023	368
[S004] Artificial Intelligence Risk Management Framework (AI RMF 1.0)	369
[S005] AI Index Report 2026	369
[S006] Python 3 Documentation.....	370
[S007] Pro Git	370
[S008] SQLite Documentation	371
[S009] NumPy Documentation	371
[S010] Dive into Deep Learning.....	372
[S011] 18.06 Linear Algebra	373
[S012] e-Handbook of Statistical Methods.....	373
[S013] A Mathematical Theory of Communication.....	374
[S014] Convex Optimization	374
[S015] Automatic Differentiation with torch.autograd	375
[S019] Probabilistic Graphical Models.....	375
[S020] scikit-learn User Guide.....	376
[S021] Machine Learning Crash Course	376
[S025] PyTorch Tutorials.....	377

[S072] Datasheets for Datasets.....	378
[S080] Artifact Review and Badging.....	378
[S082] Colaboratory FAQ	379
[S086] Recommendation on the Ethics of Artificial Intelligence	379
全书术语表	381
版权、版本与勘误.....	391

本卷 24 章路线图

月 1 · AI 全景、证据与伦理

区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。

1. [第 1 章 AI 不是魔法：任务、模型与系统 — 区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。](#)
2. [第 2 章 从符号主义到基础模型：范式如何更替 — 用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。](#)
3. [第 3 章 先问证据：怎样判断一个 AI 说法 — 区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。](#)
4. [第 4 章 价值不是附录：伦理、劳动与公共影响 — 把受影响者、利益、伤害、申诉和替代方案放进设计起点。](#)

月 2 · Python 与计算思维

理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。

5. [第 5 章 变量、类型与表达式：让计算机保存意义 — 理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。](#)
6. [第 6 章 控制流、函数与测试：把大问题切小 — 用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。](#)
7. [第 7 章 列表、字典、文件与模块：组织一批数据 — 按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。](#)
8. [第 8 章 算法、复杂度与调试：找到慢和错的原因 — 从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。](#)

月 3 · Git、数据、SQL 与可复现性

理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。

9. [第 9 章 命令行与 Git：给每次改变留下可回溯证据 — 理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。](#)
10. [第 10 章 数据生命周期：采集、清洗、字典与许可 — 从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。](#)
11. [第 11 章 SQL 与关系模型：让问题变成可核对查询 — 理解表、键、连接、聚合和事务，用查询结果回答明确问题。](#)
12. [第 12 章 环境、随机种子与实验记录 — 区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。](#)

月 4 · 线性代数

理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。

13. [第 13 章 向量：把方向、特征和相似度放进坐标 — 理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。](#)
14. [第 14 章 矩阵：一次描述许多线性关系 — 把矩阵理解为线性变换和批量方程，而不只是数字表格。](#)
15. [第 15 章 投影、特征方向与 PCA — 理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。](#)
16. [第 16 章 张量、广播与数值线性代数 — 连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。](#)

月 5 · 概率统计与信息论

区分频率、主观不确定性、事件、条件概率、期望和方差。

17. [第 17 章 概率与随机变量：描述不知道的部分 — 区分频率、主观不确定性、事件、条件概率、期望和方差。](#)
18. [第 18 章 分布与贝叶斯更新 — 认识常用分布、似然、先验和后验，理解证据如何更新信念。](#)
19. [第 19 章 估计、区间与假设检验 — 区分参数、估计量、抽样误差、置信区间、效应量和 p 值。](#)
20. [第 20 章 熵、交叉熵、KL 与互信息 — 把信息量连接到不确定性压缩、分类损失和表示学习。](#)

月 6 · 微积分、优化和数值计算

从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。

21. [第 21 章 导数、梯度与链式法则 — 从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。](#)
22. [第 22 章 雅可比、海森与自动微分 — 理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。](#)
23. [第 23 章 梯度下降、动量与 Adam — 比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。](#)
24. [第 24 章 数值稳定、条件数与非凸地形 — 理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。](#)

第 1 章 AI 不是魔法：任务、模型与系统

AI as a system

项目	内容
学习单元	月 1 · AI 全景、证据与伦理
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：零基础入口，无前置章节。

结论先行： 区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。

版本边界： 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“AI 不是魔法：任务、模型与系统”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评

估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 本书导读。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 1-1 AI 不是魔法：任务、模型与系统：单变量实验（教学模拟）

项目	内容
只改变量	人工复核比例，范围 0–100%，步长 5
实验目标	只改变“人工复核比例”，观察结果、代价和风险如何一起变化，理解区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把人工复核比例分别设为 0%、50%和 100%，

项目	内容
	手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“人工复核比例”，观察结果、代价和风险如何一起变化，理解区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。

静态替代说明：静态替代：把人工复核比例分别设为 0%、50%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：`system = data + objective + model + evaluation + deployment + governance`。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 1-1 python 最小实现

```

from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="ai-is-a-system", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“AI 不是魔法：任务、模型与系统”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“人工复核比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{system} = \text{data} + \text{objective} + \text{model} + \text{evaluation} + \text{deployment} + \text{governance}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当人工复核比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“AI 不是魔法：任务、模型与系统”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“人工复核比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{system} = \text{data} + \text{objective} + \text{model} + \text{evaluation} + \text{deployment} + \text{governance}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当人工复核比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“AI 不是魔法：任务、模型与系统”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。。接着逐项标记观察到的事实、由

公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“人工复核比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{system} = \text{data} + \text{objective} + \text{model} + \text{evaluation} + \text{deployment} + \text{governance}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当人工复核比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“AI 不是魔法：任务、模型与系统”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“人工复核比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{system} = \text{data} + \text{objective} + \text{model} + \text{evaluation} + \text{deployment} + \text{governance}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么把 AI 系统想成学校食堂：菜谱像算法，

厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当人工复核比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“AI 不是魔法：任务、模型与系统”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“人工复核比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{system} = \text{data} + \text{objective} + \text{model} + \text{evaluation} + \text{deployment} + \text{governance}$ ，逐个

说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当人工复核比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“人工复核比例”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 25. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 26. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 27. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 28. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 29. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 30. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 31. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择不含个人信息的教学模拟天气记录作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开天文观测的缩小样例，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

32. **基础题：** 用自己的话解释“AI 不是魔法：任务、模型与系统”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
33. **应用题：** 在合成的植物叶片特征中设计一个只改变“人工复核比例”的实验，列出基线、三档设置、指标和失败样例。
34. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/ai-is-a-system.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S001] UNESCO, 《AI competency framework for students》, [访问 unesco.org](https://unesco.org))
- [S002] AI4K12 Initiative, 《Five Big Ideas in AI》, [访问 ai4k12.org](https://ai4k12.org))
- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》, [访问 csed.acm.org](https://csed.acm.org))

- [S004] NIST, 《Artificial Intelligence Risk Management Framework (AI RMF 1.0)》, [访问 nist.gov](https://nist.gov))
- [S005] Stanford HAI, 《AI Index Report 2026》, [访问 hai.stanford.edu](https://hai.stanford.edu))
- [S086] UNESCO, 《Recommendation on the Ethics of Artificial Intelligence》, [访问 unesco.org](https://unesco.org))

核实日期：2026-08-06。评审状态：**draft**。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 2 章 从符号主义到基础模型：范式如何更替

History and paradigms

项目	内容
学习单元	月 1 · AI 全景、证据与伦理
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 1 章 AI 不是魔法：任务、模型与系统](#)

结论先行：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“从符号主义到基础模型：范式如何更替”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条

件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 ai-is-a-system。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 2-1 从符号主义到基础模型：范式如何更替：单变量实验（教学模拟）

项目	内容
只改变量	资源投入，范围 0–100%，步长 5
实验目标	只改变“资源投入”，观察结果、代价和风险如何一起变化，理解用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把资源投入分别设为 0%、50%和 100%，手工

项目	内容
	比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“资源投入”，观察结果、代价和风险如何一起变化，理解用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。

静态替代说明：静态替代：把资源投入分别设为 0%、50%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $capability = method \times data \times compute \times feedback$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 2-1 python 最小实现

```

from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="history-and-paradigms", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“从符号主义到基础模型：范式如何更替”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“资源投入”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{capability} = \text{method} \times \text{data} \times \text{compute} \times \text{feedback}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当资源投入变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“从符号主义到基础模型：范式如何更替”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“资源投入”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{capability} = \text{method} \times \text{data} \times \text{compute} \times \text{feedback}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当资源投入变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“从符号主义到基础模型：范式如何更替”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用问题、

表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“资源投入”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{capability} = \text{method} \times \text{data} \times \text{compute} \times \text{feedback}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。

最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当资源投入变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“从符号主义到基础模型：范式如何更替”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“资源投入”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{capability} = \text{method} \times \text{data} \times \text{compute} \times \text{feedback}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当资源投入变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“从符号主义到基础模型：范式如何更替”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“资源投入”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{capability} = \text{method} \times \text{data} \times \text{compute} \times \text{feedback}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当资源投入变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“资源投入”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 35. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 36. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 37. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 38. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 39. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 40. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 41. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开道路网络与虚构出行任务作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成音频波形与环境声标签，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

42. **基础题**：用自己的话解释“从符号主义到基础模型：范式如何更替”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

43. **应用题**：在虚构游戏角色的状态日志中设计一个只改变“资源投入”的实验，列出基线、三档设置、指标和失败样例。

44. **挑战题**：设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/history-and-paradigms.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。

- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S001] UNESCO, 《AI competency framework for students》, [访问 unesco.org](https://unesco.org))
- [S002] AI4K12 Initiative, 《Five Big Ideas in AI》, [访问 ai4k12.org](https://ai4k12.org))
- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》, [访问 csed.acm.org](https://csed.acm.org))
- [S004] NIST, 《Artificial Intelligence Risk Management Framework (AI RMF 1.0)》, [访问 nist.gov](https://nist.gov))
- [S005] Stanford HAI, 《AI Index Report 2026》, [访问 hai.stanford.edu](https://hai.stanford.edu))
- [S086] UNESCO, 《Recommendation on the Ethics of Artificial Intelligence》, [访问 unesco.org](https://unesco.org))

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 draft → technical_review → fact_checked → stable 前进。

第 3 章 先问证据：怎样判断一个 AI 说法

Evidence and claims

项目	内容
学习单元	月 1 · AI 全景、证据与伦理
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 2 章 从符号主义到基础模型：范式如何更替](#)

结论先行： 区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。
版本边界： 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“先问证据：怎样判断一个 AI 说法”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、

评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 history-and-paradigms。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 3-1 先问证据：怎样判断一个 AI 说法：单变量实验（教学模拟）

项目	内容
只改变量	样本量，范围 10-1000 个，步长 10
实验目标	只改变“样本量”，观察结果、代价和风险如何一起变化，理解区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把样本量分别设为 10 个、100 个和 1000 个，

项目	内容
	手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“样本量”，观察结果、代价和风险如何一起变化，理解区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。

静态替代说明：静态替代：把样本量分别设为 10 个、100 个和 1000 个，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\text{claim strength} \leq \text{evidence strength}$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 3-1 python 最小实现

```

from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="evidence-and-claims", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“先问证据：怎样判断一个 AI 说法”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“样本量”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{claim strength} \leq \text{evidence strength}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当样本量变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“先问证据：怎样判断一个 AI 说法”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“样本量”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{claim strength} \leq \text{evidence strength}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当样本量变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“先问证据：怎样判断一个 AI 说法”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。。接着逐项标记观察到的事

实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“样本量”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{claim strength} \leq \text{evidence strength}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当样本量变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“先问证据：怎样判断一个 AI 说法”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“样本量”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{claim strength} \leq \text{evidence strength}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像判断一双跑鞋是否更快：一次试跑

不是证据，需要相同跑道、足够次数和清楚的计时方法。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当样本量变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“先问证据：怎样判断一个 AI 说法”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“样本量”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{claim strength} \leq \text{evidence strength}$ ，逐个说明符号的对象、形状、单位和取值范围；

把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当样本量变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“样本量”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 45. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 46. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 47. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 48. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 49. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 50. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 51. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成的植物叶片特征作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择校园能源使用的匿名汇总数据，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

52. **基础题：** 用自己的话解释“先问证据：怎样判断一个 AI 说法”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
53. **应用题：** 在公开天文观测的缩小样例中设计一个只改变“样本量”的实验，列出基线、三档设置、指标和失败样例。
54. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/evidence-and-claims.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S001] UNESCO, 《AI competency framework for students》, [访问 unesco.org](https://unesco.org))
- [S002] AI4K12 Initiative, 《Five Big Ideas in AI》, [访问 ai4k12.org](https://ai4k12.org))

- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》, [访问 csed.acm.org](https://www.csed.acm.org))
- [S004] NIST, 《Artificial Intelligence Risk Management Framework (AI RMF 1.0)》, [访问 nist.gov](https://nist.gov))
- [S005] Stanford HAI, 《AI Index Report 2026》, [访问 hai.stanford.edu](https://hai.stanford.edu))
- [S086] UNESCO, 《Recommendation on the Ethics of Artificial Intelligence》, [访问 unesco.org](https://unesco.org))

核实日期：2026-08-06。评审状态：**draft**。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 4 章 价值不是附录：伦理、劳动与公共影响

Ethics and impact

项目	内容
学习单元	月 1 · AI 全景、证据与伦理
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：第 3 章 先问证据：怎样判断一个 AI 说法

结论先行：把受影响者、利益、伤害、申诉和替代方案放进设计起点。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“价值不是附录：伦理、劳动与公共影响”从一个名词变成可操作的分析框架。

你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：把受影响者、利益、伤害、申诉和替代方案放进设计起点。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 evidence-and-claims。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。类比的¹价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 4-1 价值不是附录：伦理、劳动与公共影响：单变量实验（教学模拟）

项目	内容
只改变量	风险容忍度，范围 0–100%，步长 5
实验目标	只改变“风险容忍度”，观察结果、代价和风险如何一起变化，理解把受影响者、利益、伤害、申诉和替代方案放进设计起点。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把风险容忍度分别设为 0%、35%和 100%，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性

项目	内容
	能或现实世界因果效果。

实验目标：只改变“风险容忍度”，观察结果、代价和风险如何一起变化，理解把受影响者、利益、伤害、申诉和替代方案放进设计起点。

静态替代说明：静态替代：把风险容忍度分别设为 0%、35%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $expected\ impact = \sum stakeholder_probability \times consequence$ 。先不要急着推公式。

逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 4-1 python 最小实现

```
from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
```

```
seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="ethics-and-impact", value=0.60)
print(evaluate(case))
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“价值不是附录：伦理、劳动与公共影响”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把受影响者、利益、伤害、申诉和替代方案放进设计起点。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“风险容忍度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{expected impact} = \sum \text{stakeholder_probability} \times \text{consequence}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当风险容忍度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“价值不是附录：伦理、劳动与公共影响”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把受影响者、利益、伤害、申诉和替代方案放进设计起点。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“风险容忍度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{expected impact} = \sum \text{stakeholder_probability} \times \text{consequence}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当风险容忍度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“价值不是附录：伦理、劳动与公共影响”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把受影响者、利益、伤害、申诉和替代方案放进设计起点。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如

固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。

若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“风险容忍度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{expected impact} = \sum \text{stakeholder_probability} \times \text{consequence}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当风险容忍度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“价值不是附录：伦理、劳动与公共影响”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把受影响者、利益、伤害、申诉和替代方案放进设计起点。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。这个

类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“风险容忍度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{expected impact} = \sum \text{stakeholder_probability} \times \text{consequence}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当风险容忍度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“价值不是附录：伦理、劳动与公共影响”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把受影响者、利益、伤害、申诉和替代方案放进设计起点。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“风险容忍度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{expected impact} = \sum \text{stakeholder_probability} \times \text{consequence}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像设计校规：不能只问执行是否高效，

还要问谁被误判、能否申诉、规则是否透明。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当风险容忍度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“风险容忍度”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 55. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 56. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 57. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 58. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 59. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 60. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 61. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择虚构游戏角色的状态日志作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开古诗文本与版权已明确的材料，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 62. **基础题：** 用自己的话解释“价值不是附录：伦理、劳动与公共影响”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

63. **应用题：**在合成音频波形与环境声标签中设计一个只改变“风险容忍度”的实验，列出基线、三档设置、指标和失败样例。

64. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/ethics-and-impact.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S001] UNESCO, 《AI competency framework for students》, [访问 unesco.org](https://unesco.org))
- [S002] AI4K12 Initiative, 《Five Big Ideas in AI》, [访问 ai4k12.org](https://ai4k12.org))
- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》, [访问 csed.acm.org](https://csed.acm.org))
- [S004] NIST, 《Artificial Intelligence Risk Management Framework (AI RMF 1.0)》, [访问 nist.gov](https://nist.gov))
- [S005] Stanford HAI, 《AI Index Report 2026》, [访问 hai.stanford.edu](https://hai.stanford.edu))

- [S086] UNESCO, 《Recommendation on the Ethics of Artificial Intelligence》, [访问
unesco.org](https://unesco.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 1 阶段项目：给校园社团做一份 AI 机会与风险地



项目结论： 选择一个真实但不含个人信息的校园任务，画出数据、模型、界面、人工判断和风险边界。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

65. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
66. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
67. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
68. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

第 5 章 变量、类型与表达式：让计算机保存意义

Values and types

项目	内容
学习单元	月 2 · Python 与计算思维
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 4 章 价值不是附录：伦理、劳动与公共影响](#)

结论先行：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“变量、类型与表达式：让计算机保存意义”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署

条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 `ethics-and-impact`。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。类比的值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 5-1 变量、类型与表达式：让计算机保存意义：单变量实验（教学模拟）

项目	内容
只改变量	可转换数据比例，范围 0-100%，步长 5
实验目标	只改变“可转换数据比例”，观察结果、代价和风险如何一起变化，理解理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把可转换数据比例分别设为 0%、50%和

项目	内容
	100%，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“可转换数据比例”，观察结果、代价和风险如何一起变化，理解理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。

静态替代说明：静态替代：把可转换数据比例分别设为 0%、50%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：name → value : type。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 5-1 python 最小实现

```

from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="python-values-and-types", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“变量、类型与表达式：让计算机保存意义”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“可转换数据比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{name} \rightarrow \text{value} : \text{type}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当可转换数据比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“变量、类型与表达式：让计算机保存意义”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“可转换数据比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{name} \rightarrow \text{value} : \text{type}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当可转换数据比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“变量、类型与表达式：让计算机保存意义”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。。接着逐项标记观察到的事实、

由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“可转换数据比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{name} \rightarrow \text{value} : \text{type}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当可转换数据比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“变量、类型与表达式：让计算机保存意义”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“可转换数据比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{name} \rightarrow \text{value} : \text{type}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么变量像贴在收纳盒上的标签，标签可

以换到另一个盒子，但盒内物品的类型会影响能做什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当可转换数据比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“变量、类型与表达式：让计算机保存意义”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“可转换数据比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{name} \rightarrow \text{value} : \text{type}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大

值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当可转换数据比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“可转换数据比例”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

69. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。

70. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。

71. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

72. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

73. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

74. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

75. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开天文观测的缩小样例作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择虚构机器人在二维网格中的轨迹，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

76. **基础题：** 用自己的话解释“变量、类型与表达式：让计算机保存意义”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
77. **应用题：** 在校园能源使用的匿名汇总数据中设计一个只改变“可转换数据比例”的实验，列出基线、三档设置、指标和失败样例。
78. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/python-values-and-types.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》，[访问 csed.acm.org](https://www.acm.org/publications/curricula/2023))
- [S006] Python Software Foundation, 《Python 3 Documentation》，[访问 docs.python.org](https://docs.python.org/3/))

- [S082] Google, 《Colaboratory FAQ》, [访问 research.google.com](https://research.google.com/colab/faq/))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 6 章 控制流、函数与测试：把大问题切小

Control flow, functions and tests

项目	内容
学习单元	月 2 · Python 与计算思维
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 5 章 变量、类型与表达式：让计算机保存意义](#)

结论先行： 用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“控制流、函数与测试：把大问题切小”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条

件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 python-values-and-types。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 6-1 控制流、函数与测试：把大问题切小：单变量实验（教学模拟）

项目	内容
只改变量	分支阈值，范围 0–1，步长 0.05
实验目标	只改变“分支阈值”，观察结果、代价和风险如何一起变化，理解用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把分支阈值分别设为 0、0.5 和 1，手工比较三组教学模拟输出。

项目	内容
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“分支阈值”，观察结果、代价和风险如何一起变化，理解用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。

静态替代说明：静态替代：把分支阈值分别设为 0、0.5 和 1，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：`output = f(input); test(f(x)) = expected`。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 6-1 python 最小实现

```
from dataclasses import dataclass
@dataclass(frozen=True)
```

```
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="control-functions-tests", value=0.60)
print(evaluate(case))
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“控制流、函数与测试：把大问题切小”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分支阈值”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边

界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{output} = f(\text{input}); \text{test}(f(x)) = \text{expected}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当分支阈值变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“控制流、函数与测试：把大问题切小”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低

可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。

若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分支阈值”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{output} = f(\text{input}); \text{test}(f(x)) = \text{expected}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当分支阈值变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“控制流、函数与测试：把大问题切小”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分支阈值”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{output} = f(\text{input}); \text{test}(f(x)) = \text{expected}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当分支阈值变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“控制流、函数与测试：把大问题切小”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。。接着逐项标记观察到的事实、由

公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分支阈值”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{output} = f(\text{input}); \text{test}(f(x)) = \text{expected}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当分支阈值变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“控制流、函数与测试：把大问题切小”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分支阈值”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{output} = f(\text{input}); \text{test}(f(x)) = \text{expected}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么像组织运动会：流程表决定先后，岗

位说明隔离责任，彩排就是测试。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当分支阈值变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“分支阈值”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 79. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 80. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 81. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 82. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 83. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 84. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 85. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成音频波形与环境声标签作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开科研论文的标题与摘要元数据，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 86. **基础题：** 用自己的话解释“控制流、函数与测试：把大问题切小”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

87. **应用题：** 在公开古诗文本与版权已明确的材料中设计一个只改变“分支阈值”的实验，列出基线、三档设置、指标和失败样例。

88. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/control-functions-tests.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》，[访问 csed.acm.org](https://www.csed.acm.org/))
- [S006] Python Software Foundation, 《Python 3 Documentation》，[访问 docs.python.org](https://docs.python.org/))
- [S082] Google, 《Colaboratory FAQ》，[访问 research.google.com](https://research.google.com/))

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 7 章 列表、字典、文件与模块：组织一批数据

Data structures and files

项目	内容
学习单元	月 2 · Python 与计算思维
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 6 章 控制流、函数与测试：把大问题切小](#)

结论先行：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“列表、字典、文件与模块：组织一批数据”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条

件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 control-functions-tests。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 7-1 列表、字典、文件与模块：组织一批数据：单变量实验（教学模拟）

项目	内容
只改变量	数据条目，范围 10–1000 条，步长 10
实验目标	只改变“数据条目”，观察结果、代价和风险如何一起变化，理解按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把数据条目分别设为 10 条、100 条和 1000 条，手工比较三组教学模拟输出。

项目	内容
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“数据条目”，观察结果、代价和风险如何一起变化，理解按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。

静态替代说明：静态替代：把数据条目分别设为 10 条、100 条和 1000 条，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：structure choice ↔ access pattern。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 7-1 python 最小实现

```
from dataclasses import dataclass
```

```
@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="data-structures-files", value=0.60)
print(evaluate(case))
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“列表、字典、文件与模块：组织一批数据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“数据条目”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `structure choice` ↔ `access pattern`，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当数据条目变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“列表、字典、文件与模块：组织一批数据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“数据条目”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{structure choice} \leftrightarrow \text{access pattern}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当数据条目变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“列表、字典、文件与模块：组织一批数据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。。接着逐项标记观察到的事实、由

公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“数据条目”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{structure choice} \leftrightarrow \text{access pattern}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当数据条目变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“列表、字典、文件与模块：组织一批数据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“数据条目”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{structure choice} \leftrightarrow \text{access pattern}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么书架适合按顺序浏览，通讯录适合按

名字查找；结构应由使用方式决定。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当数据条目变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“列表、字典、文件与模块：组织一批数据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“数据条目”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 structure choice $\leftrightarrow$ access pattern，逐个说明符号的对象、形状、单位和取值范围；

把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当数据条目变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“数据条目”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

89. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。

90. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。

91. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

92. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

93. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

94. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

95. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择校园能源使用的匿名汇总数据作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成电商目录但不含真实用户行为，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

96. **基础题：** 用自己的话解释“列表、字典、文件与模块：组织一批数据”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
97. **应用题：** 在虚构机器人在二维网格中的轨迹中设计一个只改变“数据条目”的实验，列出基线、三档设置、指标和失败样例。
98. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/data-structures-files.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》，[访问 csed.acm.org](https://www.acm.org/publications/curricula/2023))
- [S006] Python Software Foundation, 《Python 3 Documentation》，[访问 docs.python.org](https://docs.python.org/3/))

- [S082] Google, 《Colaboratory FAQ》, [访问 research.google.com](https://research.google.com))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 8 章 算法、复杂度与调试：找到慢和错的原因

Algorithms, complexity and debugging

项目	内容
学习单元	月 2 · Python 与计算思维
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 7 章 列表、字典、文件与模块：组织一批数据](#)

结论先行：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“算法、复杂度与调试：找到慢和错的原因”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 data-structures-files。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 8-1 算法、复杂度与调试：找到慢和错的原因：单变量实验（教学模拟）

项目	内容
只改变量	输入规模，范围 10–1000n，步长 10
实验目标	只改变“输入规模”，观察结果、代价和风险如何一起变化，理解从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把输入规模分别设为 10n、100n 和 1000n，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“输入规模”，观察结果、代价和风险如何一起变化，理解从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。

静态替代说明：静态替代：把输入规模分别设为 10n、100n 和 1000n，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ 。先不要急着推公式。逐个给符号贴

标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：

如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；

如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 8-1 python 最小实现

```
from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="algorithms-debugging", value=0.60)
print(evaluate(case))
```


运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“算法、复杂度与调试：找到慢和错的原因”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“输入规模”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当输入规模变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“算法、复杂度与调试：找到慢和错的原因”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“输入规模”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当输入规模变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“算法、复杂度与调试：找到慢和错的原因”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规

则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“输入规模”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当输入规模变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“算法、复杂度与调试：找到慢和错的原因”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。这

个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“输入规模”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当输入规模变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“算法、复杂度与调试：找到慢和错的原因”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“输入规模”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $T(n) \in O(1), O(\log n), O(n), O(n \log n), O(n^2)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么排队逐个问人是线性查找，按学号二

分翻名册更快，但名册必须先排序。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当输入规模变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“输入规模”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 99. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 100. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 101. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 102. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 103. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 104. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 105. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开古诗文本与版权已明确的材料作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择校园图书馆的公开借阅类别，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 106. **基础题：** 用自己的话解释“算法、复杂度与调试：找到慢和错的原因”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

107. **应用题：**在公开科研论文的标题与摘要元数据中设计一个只改变“输入规模”的实验，列出基线、三档设置、指标和失败样例。

108. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/algorithms-debugging.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S003] ACM, IEEE-CS, AAAI, 《Computer Science Curricula 2023》，[访问 csed.acm.org](https://www.csed.acm.org/))
- [S006] Python Software Foundation, 《Python 3 Documentation》，[访问 docs.python.org](https://docs.python.org/))
- [S082] Google, 《Colaboratory FAQ》，[访问 research.google.com](https://research.google.com/))

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 2 阶段项目：制作一个可测试的学习时间分配器

项目结论：用虚构课程数据写命令行程序，支持输入校验、函数拆分、测试和错误提示。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

109. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
110. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
111. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
112. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

第 9 章 命令行与 Git：给每次改变留下可回溯证据

CLI and Git

项目	内容
学习单元	月 3 · Git、数据、SQL 与可复现性
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 8 章 算法、复杂度与调试：找到慢和错的原因](#)

结论先行：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“命令行与 Git：给每次改变留下可回溯证据”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署

条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 algorithms-debugging。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。类比的 价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在 笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证 据。

3. 互动实验：只改变一个变量

实验 9-1 命令行与 Git：给每次改变留下可回溯证据：单变量实验（教学模拟）

项目	内容
只改变量	单次提交改动，范围 1–100 项，步长 1
实验目标	只改变“单次提交改动”，观察结果、代价和风险如何一起变化，理解理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把单次提交改动分别设为 1 项、12 项和 100

项目	内容
	项，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“单次提交改动”，观察结果、代价和风险如何一起变化，理解理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。

静态替代说明：静态替代：把单次提交改动分别设为 1 项、12 项和 100 项，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：working tree → staging → commit graph。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 9-1 python 最小实现

```

from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="git-and-cli", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“命令行与 Git：给每次改变留下可回溯证据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“单次提交改动”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 working tree → staging → commit graph，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么 Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当单次提交改动变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“命令行与 Git：给每次改变留下可回溯证据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“单次提交改动”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 working tree → staging → commit graph，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么 Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当单次提交改动变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“命令行与 Git：给每次改变留下可回溯证据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。。接着逐项标记观察到的事实、

由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“单次提交改动”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 working tree → staging → commit graph，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么 Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当单次提交改动变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“命令行与 Git：给每次改变留下可回溯证据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“单次提交改动”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 working tree → staging → commit graph，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么 Git 像实验室的带时间戳实验记录本，

提交不是云备份，而是一组可解释的变化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当单次提交改动变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“命令行与 Git：给每次改变留下可回溯证据”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“单次提交改动”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 working tree → staging → commit graph，逐个说明符号的对象、形状、单位和取值

范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么 Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当单次提交改动变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“单次提交改动”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 113. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 114. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 115. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 116. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 117. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 118. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 119. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择虚构机器人在二维网格中的轨迹作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择不含个人信息的教学模拟天气记录，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

120. **基础题：**用自己的话解释“命令行与 Git：给每次改变留下可回溯证据”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
121. **应用题：**在合成电商目录但不含真实用户行为中设计一个只改变“单次提交改动”的实验，列出基线、三档设置、指标和失败样例。
122. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/git-and-cli.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S007] Scott Chacon, Ben Straub, 《Pro Git》，[访问 git-scm.com](https://git-scm.com)
- [S008] SQLite Project, 《SQLite Documentation》，[访问 sqlite.org](https://sqlite.org)
- [S020] scikit-learn developers, 《scikit-learn User Guide》，[访问 scikit-learn.org](https://scikit-learn.org)

- [S072] Timnit Gebru et al., 《Datasheets for Datasets》, [访问 arxiv.org](#))
- [S080] ACM, 《Artifact Review and Badging》, [访问 acm.org](#))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 10 章 数据生命周期：采集、清洗、字典与许可

Data lifecycle

项目	内容
学习单元	月 3 · Git、数据、SQL 与可复现性
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 9 章 命令行与 Git：给每次改变留下可回溯证据](#)

结论先行：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“数据生命周期：采集、清洗、字典与许可”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署

条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 `git-and-cli`。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 10-1 数据生命周期：采集、清洗、字典与许可：单变量实验（教学模拟）

项目	内容
只改变量	缺失比例，范围 0-80%，步长 5
实验目标	只改变“缺失比例”，观察结果、代价和风险如何一起变化，理解从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把缺失比例分别设为 0%、10%和 80%，手工比较三组教学模拟输出。

项目	内容
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“缺失比例”，观察结果、代价和风险如何一起变化，理解从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。

静态替代说明：静态替代：把缺失比例分别设为 0%、10%和 80%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：`usable data = raw data - errors - leakage - prohibited use`。先不要急着推公式。

逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 10-1 python 最小实现

```
from dataclasses import dataclass
@dataclass(frozen=True)
```

```

class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="data-lifecycle", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“数据生命周期：采集、清洗、字典与许可”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缺失比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{usable data} = \text{raw data} - \text{errors} - \text{leakage} - \text{prohibited use}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当缺失比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“数据生命周期：采集、清洗、字典与许可”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缺失比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{usable data} = \text{raw data} - \text{errors} - \text{leakage} - \text{prohibited use}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当缺失比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“数据生命周期：采集、清洗、字典与许可”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。。接着逐项标记观察到的事实、

由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缺失比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{usable data} = \text{raw data} - \text{errors} - \text{leakage} - \text{prohibited use}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当缺失比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“数据生命周期：采集、清洗、字典与许可”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缺失比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{usable data} = \text{raw data} - \text{errors} - \text{leakage} - \text{prohibited use}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么食材从采购到储存都有标签和保质期，

数据也需要来路、用途和退出机制。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当缺失比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“数据生命周期：采集、清洗、字典与许可”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缺失比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{usable data} = \text{raw data} - \text{errors} - \text{leakage} - \text{prohibited use}$ ，逐个说明符号的对象、

形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。

任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当缺失比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“缺失比例”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 123. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 124. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 125. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 126. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 127. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 128. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 129. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开科研论文的标题与摘要元数据作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开道路网络与虚构出行任务，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

130. **基础题：**用自己的话解释“数据生命周期：采集、清洗、字典与许可”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
131. **应用题：**在校园图书馆的公开借阅类别中设计一个只改变“缺失比例”的实验，列出基线、三档设置、指标和失败样例。
132. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/data-lifecycle.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S007] Scott Chacon, Ben Straub, 《Pro Git》，[访问 git-scm.com](https://git-scm.com)
- [S008] SQLite Project, 《SQLite Documentation》，[访问 sqlite.org](https://sqlite.org)
- [S020] scikit-learn developers, 《scikit-learn User Guide》，[访问 scikit-learn.org](https://scikit-learn.org)

- [S072] Timnit Gebru et al., 《Datasheets for Datasets》, [访问 arxiv.org](#))
- [S080] ACM, 《Artifact Review and Badging》, [访问 acm.org](#))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 11 章 SQL 与关系模型：让问题变成可核对查询

SQL and relational data

项目	内容
学习单元	月 3 · Git、数据、SQL 与可复现性
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 10 章 数据生命周期：采集、清洗、字典与许可](#)

结论先行：理解表、键、连接、聚合和事务，用查询结果回答明确问题。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“SQL 与关系模型：让问题变成可核对查询”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解表、键、连接、聚合和事务，用查询结果回答明确问题。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 data-lifecycle。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 11-1 SQL 与关系模型：让问题变成可核对查询：单变量实验（教学模拟）

项目	内容
只改变量	匹配比例，范围 0–100%，步长 5
实验目标	只改变“匹配比例”，观察结果、代价和风险如何一起变化，理解理解表、键、连接、聚合和事务，用查询结果回答明确问题。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把匹配比例分别设为 0%、30%和 100%，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“匹配比例”，观察结果、代价和风险如何一起变化，理解理解表、键、连接、聚合和事务，用查询结果回答明确问题。

静态替代说明：静态替代：把匹配比例分别设为 0%、30%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：`result = SELECT(project(join(filter.tables)))`。先不要急着推公式。逐个给符号贴

标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：

如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；

如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 11-1 python 最小实现

```
from dataclasses import dataclass

@dataclass(frozen=True)
class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="sql-relations", value=0.60)
print(evaluate(case))
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化管理

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“SQL 与关系模型：让问题变成可核对查询”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解表、键、连接、聚合和事务，用查询结果回答明确问题。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“匹配比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `result = SELECT(project(join(filter(tables))))`，逐个说明符号的对象、形状、单位和取

值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当匹配比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“SQL 与关系模型：让问题变成可核对查询”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解表、键、连接、聚合和事务，用查询结果回答明确问题。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“匹配比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边

界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `result = SELECT(project(join(filter(tables))))`，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当匹配比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“SQL 与关系模型：让问题变成可核对查询”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解表、键、连接、聚合和事务，用查询结果回答明确问题。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接

受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“匹配比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `result = SELECT(project(join(filter(tables))))`，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当匹配比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“SQL 与关系模型：让问题变成可核对查询”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解表、键、连接、聚合和事务，用查询结果回答明确问题。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“匹配比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `result = SELECT(project(join(filter(tables))))`，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当匹配比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“SQL 与关系模型：让问题变成可核对查询”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解表、键、连接、聚合和事务，用查询结果回答明确问题。。接着逐项标记观察到的事实、

由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“匹配比例”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 `result = SELECT(project(join(filter(tables))))`，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么学校里学生表和课程表通过学号、课程号连接，不必在每张表重复全部信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当匹配比例变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“匹配比例”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

133. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
134. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
135. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

136. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

137. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

138. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

139. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成电商目录但不含真实用户行为作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成的植物叶片特征，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

140. **基础题：** 用自己的话解释“SQL 与关系模型：让问题变成可核对查询”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

141. **应用题：** 在不含个人信息的教学模拟天气记录中设计一个只改变“匹配比例”的实验，列出基线、三档设置、指标和失败样例。

142. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/sql-relations.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S007] Scott Chacon, Ben Straub, 《Pro Git》，[访问 git-scm.com](https://git-scm.com))
- [S008] SQLite Project, 《SQLite Documentation》，[访问 sqlite.org](https://sqlite.org))
- [S020] scikit-learn developers, 《scikit-learn User Guide》，[访问 scikit-learn.org](https://scikit-learn.org))
- [S072] Timnit Gebru et al., 《Datasheets for Datasets》，[访问 arxiv.org](https://arxiv.org))
- [S080] ACM, 《Artifact Review and Badging》，[访问 acm.org](https://acm.org))

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 draft → technical_review → fact_checked → stable 前进。

第 12 章 环境、随机种子与实验记录

Reproducible experiments

项目	内容
学习单元	月 3 · Git、数据、SQL 与可复现性
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 11 章 SQL 与关系模型：让问题变成可核对查询](#)

结论先行： 区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“环境、随机种子与实验记录”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条

件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 sql-relations。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。类比的值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 12-1 环境、随机种子与实验记录：单变量实验（教学模拟）

项目	内容
只改变量	重复运行次数，范围 1-20 次，步长 1
实验目标	只改变“重复运行次数”，观察结果、代价和风险如何一起变化，理解区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把重复运行次数分别设为 1 次、5 次和 20 次，手工比较三组教学模拟输出。

项目	内容
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“重复运行次数”，观察结果、代价和风险如何一起变化，理解区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。

静态替代说明：静态替代：把重复运行次数分别设为 1 次、5 次和 20 次，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系：`result = f(code, data, environment, seed, hardware)`。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 12-1 python 最小实现

```
from dataclasses import dataclass
@dataclass(frozen=True)
```

```

class Experiment:
    name: str
    value: float
    seed: int = 20260806

def evaluate(exp: Experiment) -> dict[str, float]:
    # 教学模拟：真实项目应替换为可核验测量。
    bounded = max(0.0, min(1.0, exp.value))
    return {"signal": bounded, "uncertainty": 1.0 - bounded}

case = Experiment(name="reproducible-experiments", value=0.60)
print(evaluate(case))

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“环境、随机种子与实验记录”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“重复运行次数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、

边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{result} = f(\text{code}, \text{data}, \text{environment}, \text{seed}, \text{hardware})$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当重复运行次数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“环境、随机种子与实验记录”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低

可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“重复运行次数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{result} = f(\text{code}, \text{data}, \text{environment}, \text{seed}, \text{hardware})$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当重复运行次数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“环境、随机种子与实验记录”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“重复运行次数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{result} = f(\text{code}, \text{data}, \text{environment}, \text{seed}, \text{hardware})$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当重复运行次数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“环境、随机种子与实验记录”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。。接着逐项标记观察到的事实、由公式推导的

结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“重复运行次数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{result} = f(\text{code}, \text{data}, \text{environment}, \text{seed}, \text{hardware})$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当重复运行次数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“环境、随机种子与实验记录”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“重复运行次数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{result} = f(\text{code}, \text{data}, \text{environment}, \text{seed}, \text{hardware})$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么照菜谱重做仍需相同原料、火力和计量工具；只有种子相同远远不够。只是入口，并给出一个类比失效的地方。如果解释中出现“肯

定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当重复运行次数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“重复运行次数”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 143. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 144. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 145. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 146. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 147. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 148. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 149. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择校园图书馆的公开借阅类别作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择虚构游戏角色的状态日志，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 150. **基础题：** 用自己的话解释“环境、随机种子与实验记录”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

151. **应用题：**在公开道路网络与虚构出行任务中设计一个只改变“重复运行次数”的实验，列出基线、三档设置、指标和失败样例。

152. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/reproducible-experiments.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S007] Scott Chacon, Ben Straub, 《Pro Git》，[访问 git-scm.com](https://git-scm.com))
- [S008] SQLite Project, 《SQLite Documentation》，[访问 sqlite.org](https://sqlite.org))
- [S020] scikit-learn developers, 《scikit-learn User Guide》，[访问 scikit-learn.org](https://scikit-learn.org))
- [S072] Timnit Gebru et al., 《Datasheets for Datasets》，[访问 arxiv.org](https://arxiv.org))
- [S080] ACM, 《Artifact Review and Badging》，[访问 acm.org](https://acm.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 3 阶段项目：建立一个可复现的公开数据小仓库

项目结论：保存数据字典、清洗脚本、SQL 查询、环境说明、随机种子和实验日志，不保存个人信息。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

153. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
154. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
155. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
156. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

第 13 章 向量：把方向、特征和相似度放进坐标

Vectors and geometry

项目	内容
学习单元	月 4 · 线性代数
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 12 章 环境、随机种子与实验记录](#)

结论先行：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“向量：把方向、特征和相似度放进坐标”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 reproducible-experiments。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 13-1 向量：把方向、特征和相似度放进坐标：单变量实验（教学模拟）

项目	内容
只改变量	向量夹角，范围 0–180°，步长 5
实验目标	只改变“向量夹角”，观察结果、代价和风险如何一起变化，理解理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把向量夹角分别设为 0°、45°和 180°，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“向量夹角”，观察结果、代价和风险如何一起变化，理解理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。

静态替代说明：静态替代：把向量夹角分别设为 0° 、 45° 和 180° ，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\|x\|_2 = \sqrt{\sum x_i^2}$ ； $\cos \theta = x \cdot y / (\|x\| \|y\|)$ 。先不要急着推公式。逐个给符号贴

标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：

如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；

如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 13-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化管理

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“向量：把方向、特征和相似度放进坐标”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“向量夹角”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\|x\|_2 = \sqrt{\sum x_i^2}$; $\cos \theta = x \cdot y / (\|x\| \|y\|)$ ，逐个说明符号的对象、形状、单位和取值范围；

把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当向量夹角变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“向量：把方向、特征和相似度放进坐标”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“向量夹角”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\|x\|_2 = \sqrt{\sum x_i^2}$; $\cos \theta = x \cdot y / (\|x\| \|y\|)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当向量夹角变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“向量：把方向、特征和相似度放进坐标”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“向量夹角”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\|x\|_2 = \sqrt{\sum x_i^2}$; $\cos \theta = x \cdot y / (\|x\| \|y\|)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当向量夹角变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“向量：把方向、特征和相似度放进坐标”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如

固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“向量夹角”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\|x\|_2 = \sqrt{\sum x_i^2}$ ； $\cos \theta = x \cdot y / (\|x\| \|y\|)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当向量夹角变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“向量：把方向、特征和相似度放进坐标”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解向量的长度、方向、基、坐标和距离，连接到样本特征表示。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流

畅文字误当成稳定能力。地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“向量夹角”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\|x\|_2 = \sqrt{\sum x_i^2}$; $\cos \theta = x \cdot y / (\|x\| \|y\|)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么地图上的位移箭头和一名学生的多科成绩都可写成有顺序的数。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当向量夹角变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“向量夹角”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

157. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
158. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
159. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

160. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

161. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

162. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

163. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择不含个人信息的教学模拟天气记录作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开天文观测的缩小样例，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

164. **基础题：** 用自己的话解释“向量：把方向、特征和相似度放进坐标”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

165. **应用题：** 在合成的植物叶片特征中设计一个只改变“向量夹角”的实验，列出基线、三档设置、指标和失败样例。

166. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/vectors-and-geometry.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S010] Aston Zhang et al., 《Dive into Deep Learning》, [访问 d2l.ai](https://d2l.ai))
- [S011] MIT OpenCourseWare, 《18.06 Linear Algebra》, [访问 ocw.mit.edu](https://ocw.mit.edu))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 14 章 矩阵：一次描述许多线性关系

Matrices and transformations

项目	内容
学习单元	月 4 · 线性代数
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 13 章 向量：把方向、特征和相似度放进坐标](#)

结论先行：把矩阵理解为线性变换和批量方程，而不只是数字表格。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“矩阵：一次描述许多线性关系”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：把矩阵理解为线性变换和批量方程，而不只是数字表格。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 `vectors-and-geometry`。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 14-1 矩阵：一次描述许多线性关系：单变量实验（教学模拟）

项目	内容
只改变量	缩放系数，范围 $0.2\text{--}3\times$ ，步长 0.1
实验目标	只改变“缩放系数”，观察结果、代价和风险如何一起变化，理解把矩阵理解为线性变换和批量方程，而不只是数字表格。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把缩放系数分别设为 $0.2\times$ 、 $1\times$ 和 $3\times$ ，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“缩放系数”，观察结果、代价和风险如何一起变化，理解把矩阵理解为线性变换和批量方程，而不只是数字表格。

静态替代说明：静态替代：把缩放系数分别设为 $0.2\times$ 、 $1\times$ 和 $3\times$ ，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $y = Ax$ ； $C = AB$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。

若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：

如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；

如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 14-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“矩阵：一次描述许多线性关系”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把矩阵理解为线性变换和批量方程，而不只是数字表格。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缩放系数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $y = Ax$; $C = AB$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输

入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当缩放系数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“矩阵：一次描述许多线性关系”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把矩阵理解为线性变换和批量方程，而不只是数字表格。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缩放系数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边

界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $y = Ax$; $C = AB$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当缩放系数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“矩阵：一次描述许多线性关系”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把矩阵理解为线性变换和批量方程，而不只是数字表格。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设

是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缩放系数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $y = Ax$; $C = AB$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当缩放系数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“矩阵：一次描述许多线性关系”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把矩阵理解为线性变换和批量方程，而不只是数字表格。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缩放系数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $y = Ax$; $C = AB$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当缩放系数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“矩阵：一次描述许多线性关系”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把矩阵理解为线性变换和批量方程，而不只是数字表格。。接着逐项标记观察到的事实、由公式推导的结

果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“缩放系数”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $y = Ax$; $C = AB$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么同一块透明投影片会按统一规则移动所有点，矩阵就是这套规则的紧凑记录。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当缩放系数变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“缩放系数”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

167. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
168. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
169. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

170. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

171. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

172. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

173. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开道路网络与虚构出行任务作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成音频波形与环境声标签，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

174. **基础题：** 用自己的话解释“矩阵：一次描述许多线性关系”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

175. **应用题：** 在虚构游戏角色的状态日志中设计一个只改变“缩放系数”的实验，列出基线、三档设置、指标和失败样例。

176. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/matrices-transformations.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S010] Aston Zhang et al., 《Dive into Deep Learning》, [访问 d2l.ai](https://d2l.ai))
- [S011] MIT OpenCourseWare, 《18.06 Linear Algebra》, [访问 ocw.mit.edu](https://ocw.mit.edu))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 15 章 投影、特征方向与 PCA

Projection, eigenvectors and PCA

项目	内容
学习单元	月 4 · 线性代数
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 14 章 矩阵：一次描述许多线性关系](#)

结论先行：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“投影、特征方向与 PCA”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 matrices-transformations。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：手电把三维物体投到墙面；选对墙面能保留更多形状信息。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 15-1 投影、特征方向与 PCA：单变量实验（教学模拟）

项目	内容
只改变量	保留主成分，范围 1–10 个，步长 1
实验目标	只改变“保留主成分”，观察结果、代价和风险如何一起变化，理解理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把保留主成分分别设为 1 个、2 个和 10 个，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“保留主成分”，观察结果、代价和风险如何一起变化，理解理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。

静态替代说明：静态替代：把保留主成分分别设为 1 个、2 个和 10 个，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\text{projection} = QQ^T x$ ； $Av = \lambda v$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 15-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“投影、特征方向与 PCA”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。手电把三维物体投到墙面；选对墙面能保留更多形状信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“保留主成分”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{projection} = QQ^T x$; $Av = \lambda v$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、

极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么手电把三维物体投到墙面；选对墙面能保留更多形状信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当保留主成分变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“投影、特征方向与PCA”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。手电把三维物体投到墙面；选对墙面能保留更多形状信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“保留主成分”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{projection} = QQ^T x; Av = \lambda v$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么手电把三维物体投到墙面；选对墙面能保留更多形状信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当保留主成分变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“投影、特征方向与 PCA”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。手电把三维物体投到墙面；选对墙面能保留更多形状信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“保留主成分”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{projection} = QQ^T x$; $Av = \lambda v$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么手电把三维物体投到墙面；选对墙面能保留更多形状信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当保留主成分变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“投影、特征方向与 PCA”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。手电把三维物体投到墙面；选对墙面能保留更多形状信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均

值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“保留主成分”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{projection} = QQ^T x$; $Av = \lambda v$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么手电把三维物体投到墙面；选对墙面能保留更多形状信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当保留主成分变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“投影、特征方向与PCA”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。手电把三维物体投到墙面；选对墙面能保留更多形状信息。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“保留主成分”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{projection} = QQ^T x$; $Av = \lambda v$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么手电把三维物体投到墙面；选对墙面能保留更多形状信息。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当保留主成分变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()`和

`optimizer.step()`完成更新。评估阶段使用 `model.eval()`与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“保留主成分”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

177. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
178. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
179. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
180. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
181. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
182. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
183. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成的植物叶片特征作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择校园能源使用的匿名汇总数据，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

184. **基础题：** 用自己的话解释“投影、特征方向与 PCA”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

185. **应用题：** 在公开天文观测的缩小样例中设计一个只改变“保留主成分”的实验，列出基线、三档设置、指标和失败样例。

186. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/projection-eigen-pca.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。

- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S010] Aston Zhang et al., 《Dive into Deep Learning》, [访问 d2l.ai](https://d2l.ai))
- [S011] MIT OpenCourseWare, 《18.06 Linear Algebra》, [访问 ocw.mit.edu](https://ocw.mit.edu))

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 draft → technical_review → fact_checked → stable 前进。

第 16 章 张量、广播与数值线性代数

Tensors and numerical linear algebra

项目	内容
学习单元	月 4 · 线性代数
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 15 章 投影、特征方向与 PCA](#)

结论先行：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“张量、广播与数值线性代数”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条

件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 projection-eigen-pca。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 16-1 张量、广播与数值线性代数：单变量实验（教学模拟）

项目	内容
只改变量	批量大小，范围 1-128，步长 1
实验目标	只改变“批量大小”，观察结果、代价和风险如何一起变化，理解连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把批量大小分别设为 1、16 和 128，手工比较三组教学模拟输出。

项目	内容
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“批量大小”，观察结果、代价和风险如何一起变化，理解连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。

静态替代说明：静态替代：把批量大小分别设为 1、16 和 128，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\text{batch_out}[b] = A[b] \times [b]$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 16-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
```

```

    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
print({"x": round(x, 4), "loss": round(objective(x), 4)})

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“张量、广播与数值线性代数”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“批量大小”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{batch_out}[b] = A[b] \times [b]$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当批量大小变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“张量、广播与数值线性代数”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“批量大小”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{batch_out}[b] = A[b]x[b]$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当批量大小变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“张量、广播与数值线性代数”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如

固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。

若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“批量大小”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{batch_out}[b] = A[b]x[b]$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当批量大小变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“张量、广播与数值线性代数”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成

稳定能力。张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“批量大小”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{batch_out}[b] = A[b]x[b]$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当批量大小变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“张量、广播与数值线性代数”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。张量像按班级、学生、科目叠成的多层表格；形状说明每一轴代表什么。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“批量大小”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{batch_out}[b] = A[b]x[b]$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么张量像按班级、学生、科目叠成的多

层表格；形状说明每一轴代表什么。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当批量大小变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“批量大小”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 187. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 188. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 189. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 190. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 191. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 192. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 193. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择虚构游戏角色的状态日志作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开古诗文本与版权已明确的材料，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 194. **基础题：** 用自己的话解释“张量、广播与数值线性代数”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

195. **应用题：**在合成音频波形与环境声标签中设计一个只改变“批量大小”的实验，列出基线、三档设置、指标和失败样例。

196. **挑战题：**设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/tensors-and-numeric.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org)
- [S010] Aston Zhang et al., 《Dive into Deep Learning》, [访问 d2l.ai](https://d2l.ai)
- [S011] MIT OpenCourseWare, 《18.06 Linear Algebra》, [访问 ocw.mit.edu](https://ocw.mit.edu)

核实日期：2026-08-06。评审状态：draft。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 4 阶段项目：从零实现一个二维图像变换游乐场

项目结论：用向量和矩阵实现缩放、旋转、投影与组合，并解释几何意义和数值误差。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

197. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
198. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
199. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
200. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

第 17 章 概率与随机变量：描述不知道的部分

Probability and random variables

项目	内容
学习单元	月 5 · 概率统计与信息论
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：第 16 章 张量、广播与数值线性代数

结论先行： 区分频率、主观不确定性、事件、条件概率、期望和方差。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“概率与随机变量：描述不知道的部分”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：区分频率、主观不确定性、事件、条件概率、期望和方差。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 `tensors-and-numerics`。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。类比的价

值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对

应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 17-1 概率与随机变量：描述不知道的部分：单变量实验（教学模拟）

项目	内容
只改变量	事件概率，范围 0–1，步长 0.05
实验目标	只改变“事件概率”，观察结果、代价和风险如何一起变化，理解区分频率、主观不确定性、事件、条件概率、期望和方差。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把事件概率分别设为 0、0.5 和 1，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“事件概率”，观察结果、代价和风险如何一起变化，理解区分频率、主观不确定性、事件、条件概率、期望和方差。

静态替代说明：静态替代：把事件概率分别设为 0、0.5 和 1，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $P(A|B)=P(A \cap B)/P(B)$ ； $E[X]=\sum xP(x)$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 17-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“概率与随机变量：描述不知道的部分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分频率、主观不确定性、事件、条件概率、期望和方差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“事件概率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(A|B)=P(A\cap B)/P(B)$; $E[X]=\sum xP(x)$ ，逐个说明符号的对象、形状、单位和取值范围；把

零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当事件概率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“概率与随机变量：描述不知道的部分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分频率、主观不确定性、事件、条件概率、期望和方差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“事件概率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边

界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(A|B)=P(A\cap B)/P(B)$; $E[X]=\sum xP(x)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当事件概率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“概率与随机变量：描述不知道的部分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分频率、主观不确定性、事件、条件概率、期望和方差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，

最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“事件概率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(A|B)=P(A \cap B)/P(B)$; $E[X]=\sum xP(x)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当事件概率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“概率与随机变量：描述不知道的部分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分频率、主观不确定性、事件、条件概率、期望和方差。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“事件概率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(A|B)=P(A \cap B)/P(B)$; $E[X]=\sum xP(x)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当事件概率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“概率与随机变量：描述不知道的部分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分频率、主观不确定性、事件、条件概率、期望和方差。。接着逐项标记观察到的事实、由公式推

导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“事件概率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(A|B)=P(A\cap B)/P(B)$; $E[X]=\sum xP(x)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么天气预报的降雨概率不是天空下了百分之三十的雨，而是对不确定事件的量化。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当事件概率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“事件概率”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

201. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
202. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
203. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

- 204. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 205. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 206. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 207. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开天文观测的缩小样例作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择虚构机器人在二维网格中的轨迹，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 208. **基础题：** 用自己的话解释“概率与随机变量：描述不知道的部分”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。
- 209. **应用题：** 在校园能源使用的匿名汇总数据中设计一个只改变“事件概率”的实验，列出基线、三档设置、指标和失败样例。
- 210. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/probability-random-variables.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S012] NIST/SEMATECH, 《e-Handbook of Statistical Methods》, [访问 itl.nist.gov](http://itl.nist.gov))
- [S013] Claude E. Shannon, 《A Mathematical Theory of Communication》, [访问 doi.org](https://doi.org/10.7551/9780262085626_001))
- [S019] Daphne Koller, Nir Friedman, 《Probabilistic Graphical Models》, [访问 mitpress.mit.edu](http://mitpress.mit.edu))
- [S021] Google, 《Machine Learning Crash Course》, [访问 developers.google.com](https://developers.google.com/machine-learning/crash-course/))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 18 章 分布与贝叶斯更新

Distributions and Bayes

项目	内容
学习单元	月 5 · 概率统计与信息论
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 17 章 概率与随机变量：描述不知道的部分](#)

结论先行：认识常用分布、似然、先验和后验，理解证据如何更新信念。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“分布与贝叶斯更新”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：认识常用分布、似然、先验和后验，理解证据如何更新信念。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 probability-random-variables。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。类比的价
值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔
记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 18-1 分布与贝叶斯更新：单变量实验（教学模拟）

项目	内容
只改变量	先验权重，范围 0–100%，步长 5
实验目标	只改变“先验权重”，观察结果、代价和风险如何一起变化，理解认识常用分布、似然、先验和后验，理解证据如何更新信念。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把先验权重分别设为 0%、50%和 100%，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“先验权重”，观察结果、代价和风险如何一起变化，理解认识常用分布、似然、先验和后验，理解证据如何更新信念。

静态替代说明：静态替代：把先验权重分别设为 0%、50%和 100%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $P(H|D)=P(D|H)P(H)/P(D)$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 18-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“分布与贝叶斯更新”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：认识常用分布、似然、先验和后验，理解证据如何更新信念。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“先验权重”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(H|D)=P(D|H)P(H)/P(D)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极

大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当先验权重变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“分布与贝叶斯更新”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：认识常用分布、似然、先验和后验，理解证据如何更新信念。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“先验权重”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(H|D)=P(D|H)P(H)/P(D)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当先验权重变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“分布与贝叶斯更新”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：认识常用分布、似然、先验和后验，理解证据如何更新信念。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“先验权重”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(H|D)=P(D|H)P(H)/P(D)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当先验权重变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“分布与贝叶斯更新”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：认识常用分布、似然、先验和后验，理解证据如何更新信念。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规

则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“先验权重”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(H|D)=P(D|H)P(H)/P(D)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当先验权重变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“分布与贝叶斯更新”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：认识常用分布、似然、先验和后验，理解证据如何更新信念。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。

侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“先验权重”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $P(H|D)=P(D|H)P(H)/P(D)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么侦探先有多个假设，新线索会改变各假设权重，但不会凭空保证某个假设正确。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当先验权重变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“先验权重”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

211. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
212. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
213. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

214. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

215. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

216. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

217. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成音频波形与环境声标签作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开科研论文的标题与摘要元数据，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

218. **基础题：** 用自己的话解释“分布与贝叶斯更新”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

219. **应用题：** 在公开古诗文本与版权已明确的材料中设计一个只改变“先验权重”的实验，列出基线、三档设置、指标和失败样例。

220. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/distributions-and-bayes.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S012] NIST/SEMATECH, 《e-Handbook of Statistical Methods》, [访问 itl.nist.gov](http://itl.nist.gov))
- [S013] Claude E. Shannon, 《A Mathematical Theory of Communication》, [访问 doi.org](https://doi.org/10.7551/ptol/1993-1))
- [S019] Daphne Koller, Nir Friedman, 《Probabilistic Graphical Models》, [访问 mitpress.mit.edu](http://mitpress.mit.edu))
- [S021] Google, 《Machine Learning Crash Course》, [访问 developers.google.com](https://developers.google.com/machine-learning/crash-course/))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 19 章 估计、区间与假设检验

Estimation and hypothesis testing

项目	内容
学习单元	月 5 · 概率统计与信息论
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系： [第 18 章 分布与贝叶斯更新](#)

结论先行： 区分参数、估计量、抽样误差、置信区间、效应量和 p 值。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“估计、区间与假设检验”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 distributions-and-bayes。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 19-1 估计、区间与假设检验：单变量实验（教学模拟）

项目	内容
只改变量	置信水平，范围 80–99%，步长 1
实验目标	只改变“置信水平”，观察结果、代价和风险如何一起变化，理解区分参数、估计量、抽样误差、置信区间、效应量和 p 值。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把置信水平分别设为 80%、95%和 99%，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“置信水平”，观察结果、代价和风险如何一起变化，理解区分参数、估计量、抽样误差、置信区间、效应量和 p 值。

静态替代说明：静态替代：把置信水平分别设为 80%、95%和 99%，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ 。先不要急着推公式。逐个给符号贴标签：

它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：

如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；

如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 19-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“估计、区间与假设检验”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“置信水平”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ ，逐个说明符号的对象、形状、单位和

取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当置信水平变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“估计、区间与假设检验”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“置信水平”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当置信水平变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“估计、区间与假设检验”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“置信水平”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当置信水平变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“估计、区间与假设检验”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园能源使用的匿名汇总数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规

则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园图书馆的公开借阅类别作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“置信水平”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园能源使用的匿名汇总数据中，当置信水平变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“估计、区间与假设检验”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：区分参数、估计量、抽样误差、置信区间、效应量和 p 值。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能

力。量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用不含个人信息的教学模拟天气记录建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成音频波形与环境声标签作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“置信水平”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{estimate} \pm \text{critical_value} \times \text{standard_error}$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么量一张桌子多次会有微小差异；区间表达测量的不确定程度，而非隐藏误差。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在不含个人信息的教学模拟天气记录中，当置信水平变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“置信水平”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

221. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
222. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
223. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

224. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

225. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

226. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

227. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择校园能源使用的匿名汇总数据作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成电商目录但不含真实用户行为，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

228. **基础题：** 用自己的话解释“估计、区间与假设检验”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

229. **应用题：** 在虚构机器人在二维网格中的轨迹中设计一个只改变“置信水平”的实验，列出基线、三档设置、指标和失败样例。

230. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/estimation-and-testing.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S012] NIST/SEMATECH, 《e-Handbook of Statistical Methods》, [访问 itl.nist.gov](http://itl.nist.gov))
- [S013] Claude E. Shannon, 《A Mathematical Theory of Communication》, [访问 doi.org](https://doi.org/10.7551/ptol/1963))
- [S019] Daphne Koller, Nir Friedman, 《Probabilistic Graphical Models》, [访问 mitpress.mit.edu](http://mitpress.mit.edu))
- [S021] Google, 《Machine Learning Crash Course》, [访问 developers.google.com](https://developers.google.com/machine-learning/crash-course/))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 20 章 熵、交叉熵、KL 与互信息

Information theory

项目	内容
学习单元	月 5 · 概率统计与信息论
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 19 章 估计、区间与假设检验](#)

结论先行： 把信息量连接到不确定性压缩、分类损失和表示学习。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“熵、交叉熵、KL 与互信息”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：把信息量连接到不确定性压缩、分类损失和表示学习。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 estimation-and-testing。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。类比的价值观是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 20-1 熵、交叉熵、KL 与互信息：单变量实验（教学模拟）

项目	内容
只改变量	分布尖锐度，范围 0.1–5，步长 0.1
实验目标	只改变“分布尖锐度”，观察结果、代价和风险如何一起变化，理解把信息量连接到不确定性压缩、分类损失和表示学习。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把分布尖锐度分别设为 0.1、1 和 5，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“分布尖锐度”，观察结果、代价和风险如何一起变化，理解把信息量连接到不确定性压缩、分类损失和表示学习。

静态替代说明：静态替代：把分布尖锐度分别设为 0.1、1 和 5，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $H(p) = -\sum p \log p$ ； $D_{KL}(p||q) = \sum p \log(p/q)$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 20-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“熵、交叉熵、KL 与互信息”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把信息量连接到不确定性压缩、分类损失和表示学习。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分布尖锐度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失

败样例。对于 $H(p) = -\sum p \log p$; $D_{KL}(p||q) = \sum p \log(p/q)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当分布尖锐度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“熵、交叉熵、KL 与互信息”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把信息量连接到不确定性压缩、分类损失和表示学习。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分布尖锐度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、

边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $H(p) = -\sum p \log p$; $D_{\text{KL}}(p||q) = \sum p \log(p/q)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当分布尖锐度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“熵、交叉熵、KL 与互信息”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把信息量连接到不确定性压缩、分类损失和表示学习。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低

可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分布尖锐度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $H(p) = -\sum p \log p$; $D_{\text{KL}}(p||q) = \sum p \log(p/q)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当分布尖锐度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“熵、交叉熵、KL 与互信息”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把信息量连接到不确定性压缩、分类损失和表示学习。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开古诗文本与版权已明确的材料建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把不含个人信息的教学模拟天气记录作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分布尖锐度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $H(p) = -\sum p \log p$; $D_{KL}(p||q) = \sum p \log(p/q)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开古诗文本与版权已明确的材料中，当分布尖锐度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“熵、交叉熵、KL 与互信息”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：把信息

量连接到不确定性压缩、分类损失和表示学习。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开道路网络与虚构出行任务建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把校园能源使用的匿名汇总数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“分布尖锐度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $H(p)=-\sum p \log p$; $D_{KL}(p||q)=\sum p \log(p/q)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么猜一枚公平硬币需要更多信息，猜几乎总是正面的硬币需要更少惊讶。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开道路网络与虚构出行任务中，当分布尖锐度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“分布尖锐度”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

231. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
232. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
233. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

234. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

235. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

236. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

237. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开古诗文本与版权已明确的材料作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择校园图书馆的公开借阅类别，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

238. **基础题：** 用自己的话解释“熵、交叉熵、KL 与互信息”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

239. **应用题：** 在公开科研论文的标题与摘要元数据中设计一个只改变“分布尖锐度”的实验，列出基线、三档设置、指标和失败样例。

240. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/information-theory.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S012] NIST/SEMATECH, 《e-Handbook of Statistical Methods》, [访问 itl.nist.gov](http://itl.nist.gov))
- [S013] Claude E. Shannon, 《A Mathematical Theory of Communication》, [访问 doi.org](https://doi.org/10.7551/ptol/1993-1))
- [S019] Daphne Koller, Nir Friedman, 《Probabilistic Graphical Models》, [访问 mitpress.mit.edu](http://mitpress.mit.edu))
- [S021] Google, 《Machine Learning Crash Course》, [访问 developers.google.com](https://developers.google.com/machine-learning/crash-course/))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 5 阶段项目：做一个不确定性诚实的天气建议器

项目结论：使用公开或教学模拟数据，输出概率、校准图、置信区间和限制，而不是只给单点答案。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

241. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
242. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
243. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
244. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

第 21 章 导数、梯度与链式法则

Derivatives, gradients and chain rule

项目	内容
学习单元	月 6 · 微积分、优化和数值计算
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 20 章 熵、交叉熵、KL 与互信息](#)

结论先行：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“导数、梯度与链式法则”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 `information-theory`。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 21-1 导数、梯度与链式法则：单变量实验（教学模拟）

项目	内容
只改变量	横坐标，范围 $-5-5x$ ，步长 0.1
实验目标	只改变“横坐标”，观察结果、代价和风险如何一起变化，理解从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把横坐标分别设为 $-5x$ 、 $1x$ 和 $5x$ ，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“横坐标”，观察结果、代价和风险如何一起变化，理解从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。

静态替代说明：静态替代：把横坐标分别设为-5x、1x 和 5x，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $df/dx = \lim(h \rightarrow 0) (f(x+h) - f(x)) / h$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 21-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“导数、梯度与链式法则”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“横坐标”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $df/dx = \lim_{h \rightarrow 0} (f(x+h) - f(x))/h$ ，逐个说明符号的对象、形状、单位和取值范围；把零

值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当横坐标变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“导数、梯度与链式法则”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“横坐标”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $df/dx = \lim_{h \rightarrow 0} (f(x+h) - f(x))/h$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当横坐标变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“导数、梯度与链式法则”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“横坐标”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $df/dx = \lim_{h \rightarrow 0} (f(x+h) - f(x))/h$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当横坐标变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“导数、梯度与链式法则”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构机器人在二维网格中的轨迹建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固

定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开道路网络与虚构出行任务作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“横坐标”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $df/dx = \lim(h \rightarrow 0)(f(x+h)-f(x))/h$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在虚构机器人在二维网格中的轨迹中，当横坐标变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“导数、梯度与链式法则”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误

当成稳定能力。走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成的植物叶片特征建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开古诗文本与版权已明确的材料作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“横坐标”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $df/dx = \lim_{h \rightarrow 0} (f(x+h) - f(x))/h$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么走山路时坡度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成的植物叶片特征中，当横坐标变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“横坐标”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

245. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
246. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
247. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

248. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

249. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

250. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

251. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择虚构机器人在二维网格中的轨迹作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择不含个人信息的教学模拟天气记录，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

252. **基础题：** 用自己的话解释“导数、梯度与链式法则”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

253. **应用题：** 在合成电商目录但不含真实用户行为中设计一个只改变“横坐标”的实验，列出基线、三档设置、指标和失败样例。

254. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/derivatives-gradients-chain.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S014] Stephen Boyd, Lieven Vandenberghe, 《Convex Optimization》, [访问 web.stanford.edu](https://web.stanford.edu))
- [S015] PyTorch, 《Automatic Differentiation with torch.autograd》, [访问 pytorch.org](https://pytorch.org))
- [S025] PyTorch, 《PyTorch Tutorials》, [访问 pytorch.org](https://pytorch.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 22 章 雅可比、海森与自动微分

Jacobians, Hessians and autodiff

项目	内容
学习单元	月 6 · 微积分、优化和数值计算
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 21 章 导数、梯度与链式法则](#)

结论先行：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“雅可比、海森与自动微分”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 derivatives-gradients-chain。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。类比的價值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 22-1 雅可比、海森与自动微分：单变量实验（教学模拟）

项目	内容
只改变量	计算图深度，范围 1–100 层，步长 1
实验目标	只改变“计算图深度”，观察结果、代价和风险如何一起变化，理解理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把计算图深度分别设为 1 层、12 层和 100 层，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

实验目标：只改变“计算图深度”，观察结果、代价和风险如何一起变化，理解理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。

静态替代说明：静态替代：把计算图深度分别设为 1 层、12 层和 100 层，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $J_{ij} = \partial y_i / \partial x_j$; $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 22-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“雅可比、海森与自动微分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“计算图深度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $J_{ij} = \partial y_i / \partial x_j$ ； $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、

极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。只是入口，并给出一个类比失效的地方。如果解释中出现

“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当计算图深度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“雅可比、海森与自动微分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“计算图深度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $J_{ij} = \partial y_i / \partial x_j$ ； $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当计算图深度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“雅可比、海森与自动微分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“计算图深度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $J_{ij} = \partial y_i / \partial x_j$; $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当计算图深度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“雅可比、海森与自动微分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开科研论文的标题与摘要元数据建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如

固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成的植物叶片特征作为迁移案例，检查相同表示和假设是否仍成立。

若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“计算图深度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $J_{ij} = \partial y_i / \partial x_j$ ； $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在公开科研论文的标题与摘要元数据中，当计算图深度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“雅可比、海森与自动微分”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成

稳定能力。工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用虚构游戏角色的状态日志建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构机器人在二维网格中的轨迹作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“计算图深度”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $J_{ij} = \partial y_i / \partial x_j$ ； $H_{ij} = \partial^2 f / \partial x_i \partial x_j$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么工厂流程图记录每个零件如何影响成品，反向模式从最终误差倒查每个旋钮。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在虚构游戏角色的状态日志中，当计算图深度变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“计算图深度”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

255. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
256. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
257. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

258. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

259. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

260. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

261. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择公开科研论文的标题与摘要元数据作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择公开道路网络与虚构出行任务，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

262. **基础题：** 用自己的话解释“雅可比、海森与自动微分”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

263. **应用题：** 在校园图书馆的公开借阅类别中设计一个只改变“计算图深度”的实验，列出基线、三档设置、指标和失败样例。

264. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/jacobian-hessian-autodiff.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S014] Stephen Boyd, Lieven Vandenberghe, 《Convex Optimization》, [访问 web.stanford.edu](https://web.stanford.edu))
- [S015] PyTorch, 《Automatic Differentiation with torch.autograd》, [访问 pytorch.org](https://pytorch.org))
- [S025] PyTorch, 《PyTorch Tutorials》, [访问 pytorch.org](https://pytorch.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 23 章 梯度下降、动量与 Adam

Gradient optimizers

项目	内容
学习单元	月 6 · 微积分、优化和数值计算
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 22 章 雅可比、海森与自动微分](#)

结论先行： 比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。**版本边界：** 本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“梯度下降、动量与 Adam”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 jacobian-hessian-autodiff。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。类比的¹价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 23-1 梯度下降、动量与 Adam：单变量实验（教学模拟）

项目	内容
只改变量	学习率，范围 0.0001–1，步长 0.0001
实验目标	只改变“学习率”，观察结果、代价和风险如何一起变化，理解比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把学习率分别设为 0.0001、0.03 和 1，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性

项目	内容
	能或现实世界因果效果。

实验目标：只改变“学习率”，观察结果、代价和风险如何一起变化，理解比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。

静态替代说明：静态替代：把学习率分别设为 0.0001、0.03 和 1，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\theta_{t+1} = \theta_t - \eta \cdot g_t$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 23-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
    return (objective(x + h) - objective(x - h)) / (2 * h)

x = 0.0
```

```
for step in range(20):
    x -= 0.08 * finite_difference(x)
    print({"x": round(x, 4), "loss": round(objective(x), 4)})
```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“梯度下降、动量与 Adam”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“学习率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失

败样例。对于 $\theta_{t+1} = \theta_t - \eta \cdot g_t$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当学习率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“梯度下降、动量与Adam”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“学习率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、

符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\theta_{t+1} = \theta_t - \eta \cdot g_t$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当学习率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“梯度下降、动量与 Adam”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接

受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“学习率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\theta_{t+1} = \theta_t - \eta \cdot g_t$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当学习率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“梯度下降、动量与 Adam”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成电商目录但不含真实用户行为建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把虚构游戏角色的状态日志作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“学习率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\theta_{t+1} = \theta_t - \eta \cdot g_t$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成电商目录但不含真实用户行为中，当学习率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“梯度下降、动量与Adam”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式。。接着逐项标记观察到的事实、由公式推导

的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用公开天文观测的缩小样例建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开科研论文的标题与摘要元数据作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“学习率”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\theta_{t+1} = \theta_t - \eta \cdot g_t$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么下山时只看脚下坡度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在公开天文观测的缩小样例中，当学习率变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“学习率”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

265. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
266. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
267. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。

268. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。

269. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。

270. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。

271. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择合成电商目录但不含真实用户行为作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择合成的植物叶片特征，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

272. **基础题：** 用自己的话解释“梯度下降、动量与 Adam”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

273. **应用题：** 在不含个人信息的教学模拟天气记录中设计一个只改变“学习率”的实验，列出基线、三档设置、指标和失败样例。

274. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/gradient-optimizers.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S014] Stephen Boyd, Lieven Vandenberghe, 《Convex Optimization》, [访问 web.stanford.edu](https://web.stanford.edu))
- [S015] PyTorch, 《Automatic Differentiation with torch.autograd》, [访问 pytorch.org](https://pytorch.org))
- [S025] PyTorch, 《PyTorch Tutorials》, [访问 pytorch.org](https://pytorch.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

第 24 章 数值稳定、条件数与非凸地形

Numerical stability and non-convexity

项目	内容
学习单元	月 6 · 微积分、优化和数值计算
建议时间	8 小时
评审状态	draft · 核实 2026-08-06
练习与实验	3 道练习 · 1 个互动实验

先修关系：[第 23 章 梯度下降、动量与 Adam](#)

结论先行：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。**版本边界：**本章属于 2026.08 Beta，评审状态为 draft。它是可运行、可核查的教学稿，不等于专家签字后的正式版。

1. 为什么重要，现在能做什么

学完本章，你应该能把“数值稳定、条件数与非凸地形”从一个名词变成可操作的分析框架。你能解释它在完整 AI 系统中的位置，写出最小数学关系，运行一个 CPU 可承受的示例，设计只改变一个变量的实验，识别至少三种失败原因，并把方法迁移到新的公开或虚构场景。真正的能力不是复述答案，而是面对结果不理想时知道先检查哪里、需要补什么证据、何时应该拒绝继续自动化。

本章聚焦：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。这件事重要，因为 AI 系统的错误通常不是“模型突然不聪明”，而是问题定义、数据表示、目标函数、评估或部署条件之间不一致。把这些环节分开，才能定位故障，也才能对他人解释为什么一个结果值得或不值得相信。

2. 先修检查与生活类比

先用一分钟自测：你能否说明输入和输出分别是什么？能否区分规则、算法、模型与系统？能否读懂变量、函数和简单图表？如果其中一项还不稳，先回到前置章节 gradient-optimizers。不要跳过地基；后续章节默认你会保存代码、记录环境、保护隐私，并区分教学模拟与真实测量。

生活类比：用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。类比的价值是帮助你先获得方向感，但它会省略重要细节。本章后面会用数学、代码和反例修正直觉。请在笔记里写下类比对对应关系，并额外写一条“类比失效处”；这一步能防止把好记的故事误当成严谨证据。

3. 互动实验：只改变一个变量

实验 24-1 数值稳定、条件数与非凸地形：单变量实验（教学模拟）

项目	内容
只改变量	浮点位宽，范围 4–64bit，步长 4
实验目标	只改变“浮点位宽”，观察结果、代价和风险如何一起变化，理解理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。
键盘操作	聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。
静态替代	静态替代：把浮点位宽分别设为 4bit、32bit 和 64bit，手工比较三组教学模拟输出。
结果边界	页面数值由公开公式生成，只用于教学，不代表真实模型性

项目	内容
	能或现实世界因果效果。

实验目标：只改变“浮点位宽”，观察结果、代价和风险如何一起变化，理解理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。

静态替代说明：静态替代：把浮点位宽分别设为 4bit、32bit 和 64bit，手工比较三组教学模拟输出。

键盘操作：聚焦滑块后使用方向键微调，Page Up/Page Down 大步调整，Home/End 到达边界。

风险提示：页面数值由公开公式生成，只用于教学，不代表真实模型性能或现实世界因果效果。

4. 直觉到数学

核心关系： $\log \sum \exp(x_i) = m + \log \sum \exp(x_i - m)$ 。先不要急着推公式。逐个给符号贴标签：它表示现实中的什么对象、是标量还是向量、单位是什么、允许取什么范围、由谁测量。然后用最小数字手算一次，再交给代码。若手算、代码和图形不一致，优先检查单位、索引、符号方向和边界，而不是马上换模型。

数学的作用不是让内容显得困难，而是把模糊说法变成可检查关系。公式必须和任务定义相互约束：如果目标变化，损失和指标也可能需要变化；如果数据尺度变化，数值稳定性和优化速度也可能变化；如果输出被用于真实决策，还必须把错误代价、拒答和人工复核写进系统。

5. 最小实现

下面代码保留固定种子和“教学模拟”标签。它的目的不是刷新榜单，而是让你看到输入、状态、计算和输出怎样连接。真实项目要补充数据许可、环境文件、独立测试、异常处理和测量记录。

代码 24-1 python 最小实现

```
import math

def objective(x: float) -> float:
    return (x - 2.0) ** 2 + 0.2 * math.sin(3 * x)

def finite_difference(x: float, h: float = 1e-5) -> float:
```

```

    return (objective(x + h) - objective(x - h)) / (2 * h)
x = 0.0
for step in range(20):
    x -= 0.08 * finite_difference(x)
print({"x": round(x, 4), "loss": round(objective(x), 4)})

```

运行后先检查输出类型、范围和极端值，再改变一个参数。不要一次改很多行；每次改变都记录“我预期什么、实际发生什么、可能原因是什么”。如果缺少 PyTorch，可先完成浏览器互动实验和纯 Python 主线；GPU 只属于后续研究复现轨道。

6. 五轮系统化理解

1. 问题边界

先把要解决的任务写成输入、输出、约束、失败代价与不做什么。对“数值稳定、条件数与非凸地形”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“浮点位宽”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\log \sum \exp(x) = m + \log \sum \exp(x_i - m)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当浮点位宽变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

3. 机制与因果链

沿着数据进入、内部计算、输出形成和外部行动逐步追踪，而不是只记名词。对“数值稳定、条件数与非凸地形”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“浮点位宽”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\log \sum \exp(x_i) = m + \log \sum \exp(x_i - m)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当浮点位宽变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. 评估与不确定性

把平均指标拆成切片、置信区间、校准、错误代价和最坏情况。对“数值稳定、条件数与非凸地形”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规

则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“浮点位宽”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\log \sum \exp(x) = m + \log \sum \exp(x_i - m)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它们替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当浮点位宽变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

9. 工程与复现

固定版本、环境、种子和测量记录，使别人能从原始输入重建结论。对“数值稳定、条件数与非凸地形”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能

力。用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用校园图书馆的公开借阅类别建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把公开天文观测的缩小样例作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“浮点位宽”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\log \sum \exp(x_i) = m + \log \sum \exp(x_i - m)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在校园图书馆的公开借阅类别中，当浮点位宽变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

11. 迁移与研究

把本章方法搬到新情境，用对照和消融判断哪些部分可迁移、哪些需要重做。对“数值稳定、条件数与非凸地形”而言，第一步不是背下定义，而是把核心问题写成一句能被反驳、能被测量的话：理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。。接着逐项标记观察到的事实、由公式推导的结果、来自文献的结论和仍然不确定的部分。这样做能防止把界面效果、单个案例或流畅文字误当成稳定能力。用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。这个类比只帮助建立直觉，不能代替形式定义；当类比与公式冲突时，以经过核实的定义、实验和边界为准。

先用合成音频波形与环境声标签建立最小案例。只保留一个输入、一个可解释的中间量和一个输出，写清数据是“教学模拟”还是“实际测量”。然后建立一个不使用本章复杂方法的基线，例如固定规则、均值、随机策略或人工清单。基线不是为了显得简单，而是为了回答：如果什么都不优化，最低可接受表现是什么？再把合成电商目录但不含真实用户行为作为迁移案例，检查相同表示和假设是否仍成立。若必须改变数据定义、损失或评价标准，就把改变记录为新的实验条件，而不是悄悄调整。

本层实验只调节“浮点位宽”。先预测它变小时结果、成本和风险会怎样，再移动滑块核对教学模拟曲线。至少记录低、中、高三个点，不能只保留最好看的点。若曲线与预测不符，依次排查单位、边界、符号方向、初始化、随机性和实现。若曲线符合预测，也不能直接声称现实有效：页面只验证公式在简化条件下的行为，真实结论仍需要代表性数据、独立测试和部署观察。

常见错误是一次同时改变数据、模型和阈值，最后无法知道改进来自哪里。正确做法是建立实验矩阵：行表示唯一改变的变量，列记录假设、设置、种子、运行环境、主要指标、切片指标、资源成本和失败样例。对于 $\text{logsumexp}(x) = m + \log \sum \exp(x_i - m)$ ，逐个说明符号的对象、形状、单位和取值范围；把零值、极大值、空输入、重复输入和相反方向代入，检查是否符合直觉。任何无法解释的异常都应保留在失败日志，而不是删除。

完成本层的标准不是“看懂了”，而是你能不看正文讲出三件事：本章方法解决什么问题；它依赖哪些假设；怎样设计一个能让它失败的测试。再向同学解释为什么用很短的尺量长距离会累积误差；问

题本身敏感时，再精密的算法也需报告不确定性。只是入口，并给出一个类比失效的地方。如果解释中出现“肯定”“完全”“智能”“理解”等宽泛词，就把它替换成可观察行为和适用条件。最后写一句研究问题：在合成音频波形与环境声标签中，当浮点位宽变化时，哪个机制最可能先成为瓶颈，怎样用对照实验验证？

7. PyTorch 实现路线

当本章涉及学习模型时，统一采用 Python 与 PyTorch：先把输入变成带明确形状的张量，再定义 `torch.nn.Module`，选择与任务一致的损失，使用 `optimizer.zero_grad()`、`loss.backward()` 和 `optimizer.step()` 完成更新。评估阶段使用 `model.eval()` 与 `torch.inference_mode()`，并保存 `state_dict`、预处理配置、词表或标签表。若本章属于规则、搜索、数据库或治理主题，则 PyTorch 不是强行套用的答案；应保留可解释的 Python 基线，在研究线中再比较神经方法。

形状检查要写在代码里：批次轴、特征轴、序列轴、通道轴不能靠猜。训练曲线至少同时记录训练损失、验证指标和资源成本；只保存最好一次会产生选择偏差。出现 NaN 时先检查输入范围、除零、对数和指数，再检查学习率、梯度范数与混合精度。出现训练好而测试差时先排查泄漏、重复样本、切分方式和分布偏移。

8. 实验设计与结果记录

为本章建立一张实验表。固定问题定义、数据版本、基线、硬件、依赖和随机种子，只改变“浮点位宽”。每个设置至少运行三次教学模拟；真实研究应根据方差和成本决定更充分的重复次数。结果栏必须标记为“实际测量”“教学模拟”或“文献引用”之一，三者不能混写。

推荐记录字段：实验 ID、日期、假设、唯一改变、数据哈希、代码提交、Python/PyTorch 版本、设备、种子、主要指标、切片指标、时间、内存、失败样例、是否符合预期、下一步。若采用外部结果，记录作者、标题、DOI 或 URL、访问日期、实验环境和原文限制；不要把论文中的一个数字搬到不同数据与硬件上直接比较。

9. 常见失败与排查顺序

- 275. **问题定义错位。** 指标提高但现实任务没有改善。重新写输入、决策和失败代价。
- 276. **数据泄漏或重复。** 训练与测试共享样本、时间或主体信息。按真实部署边界重新切分。
- 277. **表示不一致。** 训练和推理使用不同预处理、词表、单位或类别顺序。把预处理随模型版本化。
- 278. **只看平均值。** 少数群体、极端场景或长尾失败被平均掩盖。增加切片、区间和最坏情况。
- 279. **一次改变太多。** 无法归因改进来源。回到单变量实验和消融。
- 280. **把流畅当正确。** 生成结果自然但证据、逻辑或计算错误。增加引用核对、约束和拒答。
- 281. **部署条件变化。** 延迟、内存、网络、传感器或用户行为不同。建立监控、降级与回滚。

排查时按“数据→目标→实现→优化→评估→部署→治理”顺序逐层缩小范围。先构造最小可复现失败，再看日志和中间量；不要通过删除失败样本让图表变好。涉及人身、医疗、法律、财务、教育机会或其他高风险决策时，本章练习不能替代专业判断，也不能直接上线。

10. 迁移任务

工程线：把本章最小实现包装成一个可重复运行的命令，加入输入校验、测试、环境文件、日志、模型或规则版本、CPU 资源上限和 README。选择校园图书馆的公开借阅类别作为公开或合成案例，提供一个正常输入、一个边界输入和一个应拒绝的输入。

研究线：选择虚构游戏角色的状态日志，提出一个可证伪问题。复现简单基线，再改变一个机制；至少做一次消融、一次分布变化测试和一次失败类别分析。报告效应大小与不确定性，不只报告最好分数。若结果没有改善，也要解释它排除了什么假设。

11. 三级练习

- 282. **基础题：** 用自己的话解释“数值稳定、条件数与非凸地形”，说明它的输入、输出和一个关键假设；再指出生活类比的一个局限。

283. **应用题：** 在公开道路网络与虚构出行任务中设计一个只改变“浮点位宽”的实验，列出基线、三档设置、指标和失败样例。

284. **挑战题：** 设计一个会让本章方法失效的分布变化或对抗条件，给出检测、拒答、人工复核和回滚方案。

独立答案位于 `content/ai-book/answers/numerical-stability.md`，网页中默认折叠，建议先完成再核对。

12. 本章能力检查

- 我能在两分钟内说清本章解决的问题与不解决的问题。
- 我能解释核心公式中每个符号的对象、形状和单位。
- 我能运行最小代码，并预测改变一个变量的方向性影响。
- 我能区分教学模拟、实际测量和文献引用。
- 我能给出至少三种失败原因，并按顺序排查。
- 我能说明何时必须停止自动化并交给给人。
- 我能提出一个可证伪研究问题和最低成本基线。

13. 来源与核实

本章来源于下列一手框架、官方文档、开放教材或研究论文的结构化核对；正文为原创解释，不复制原图、习题答案或大段原文。时效内容在每个版本重新核实。

- [S009] NumPy Developers, 《NumPy Documentation》, [访问 numpy.org](https://numpy.org))
- [S014] Stephen Boyd, Lieven Vandenberghe, 《Convex Optimization》, [访问 web.stanford.edu](https://web.stanford.edu))
- [S015] PyTorch, 《Automatic Differentiation with torch.autograd》, [访问 pytorch.org](https://pytorch.org))
- [S025] PyTorch, 《PyTorch Tutorials》, [访问 pytorch.org](https://pytorch.org))

核实日期：2026-08-06。评审状态：`draft`。只有技术评审、事实核查和专家签字完成后，状态才能按 `draft` → `technical_review` → `fact_checked` → `stable` 前进。

月 6 阶段项目：训练一条可视化的微型回归曲线

项目结论：从损失函数、梯度、学习率到稳定性完整记录训练，比较至少三种设置。当前状态为 draft。

交付物

- 问题与边界说明
- 可运行最小实现
- 实验记录与失败分析
- 来源与许可清单
- 风险、隐私与人工闸门
- 复现 README

验收方法

- 285. 在 CPU 或浏览器主线运行；若使用 GPU，另提供缩小数据冒烟测试。
- 286. 保存环境、依赖、随机种子、数据许可、测量记录和失败样例。
- 287. 逐条标记实际测量、教学模拟和文献引用，写清适用边界。
- 288. 不收集个人信息；高风险输出必须有人审阅、可拒答、可回滚。

独立练习答案

先完成题目、保留自己的推理和错误，再查看参考答案。答案给出最低证据结构，不是唯一正确表述。

第 1 章答案 · AI 不是魔法：任务、模型与系统

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，区分任务、算法、训练模型、产品系统与社会情境，避免把聊天界面等同于全部 AI。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比

“把 AI 系统想成学校食堂：菜谱像算法，厨师训练像优化，菜品像输出，卫生制度和反馈渠道属于系统。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成的植物叶片特征，先用固定规则或简单统计量作为基线。把“人工复核比例”设置为 0%、50%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 2 章答案 · 从符号主义到基础模型：范式如何更替

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，用问题、表示、搜索、学习和规模化五条线理解 AI 史，而不是背年份。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“像城市交通从纸质地图、固定信号灯到实时导航演进，每一代解决旧瓶颈，也带来新依赖。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择虚构游戏角色的状态日志，先用固定规则或简单统计量作为基线。把“资源投入”设置为 0%、50%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上

一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 3 章答案 · 先问证据：怎样判断一个 AI 说法

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，区分演示、基准、受控实验、真实部署和因果结论，学会追问样本、基线与不确定性。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“像判断一双跑鞋是否更快：一次试跑不是证据，需要相同跑道、足够次数和清楚的计时方法。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开天文观测的缩小样例，先用固定规则或简单统计量作为基线。把“样本量”设置为 10 个、100 个和 1000 个，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 4 章答案 · 价值不是附录：伦理、劳动与公共影响

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，把受影响者、利益、伤害、申诉和替代方案放进设计起点。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“像设计校规：不能只问执行是否高效，还要问谁被误判、能否申诉、规则是否透明。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成音频波形与环境声标签，先用固定规则或简单统计量作为基线。把“风险容忍度”设置为 0%、35%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 5 章答案·变量、类型与表达式：让计算机保存意义

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解值、名称、类型、表达式和状态，避免把变量当成数学里的固定未知数。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“变量像贴在收纳盒上的标签，标签可以换到另一个盒子，但盒内物品的类型会影响能做什么。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择校园能源使用的匿名汇总数据，先用固定规则或简单统计量作为基线。把“可转换数据比例”设置为 0%、50%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。

页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 6 章答案·控制流、函数与测试：把大问题切小

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，用分支表达规则、用循环处理重复、用函数隔离责任、用测试固定预期。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“像组织运动会：流程表决定先后，岗位说明隔离责任，彩排就是测试。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开古诗文本与版权已明确的材料，先用固定规则或简单统计量作为基线。把“分支阈值”设置为 0、0.5 和 1，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果

必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 7 章答案·列表、字典、文件与模块：组织一批数据

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，按访问方式选择列表、集合、字典，理解序列化、文本编码与模块边界。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“书架适合按顺序浏览，通讯录适合按名字查找；结构应由使用方式决定。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择虚构机器人在二维网格中的轨迹，先用固定规则或简单统计量作为基线。把“数据条目”设置为 10 条、100 条和 1000 条，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。

页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 8 章答案·算法、复杂度与调试：找到慢和错的原因

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，从输入规模分析时间和空间增长，用最小复现、断言和日志定位故障。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“排队逐个问人是线性查找，按学号二分翻名册更快，但名册必须先排序。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开科研论文的标题与摘要元数据，先用固定规则或简单统计量作为基线。把“输入规模”设置为 $10n$ 、 $100n$ 和 $1000n$ ，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。

页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 9 章答案 · 命令行与 Git：给每次改变留下可回溯证据

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解工作区、暂存区、提交、分支和合并，学会小步提交与非破坏式恢复。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“Git 像实验室的带时间戳实验记录本，提交不是云备份，而是一组可解释的变化。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成电商目录但不含真实用户行为，先用固定规则或简单统计量作为基线。把“单次提交改动”设置为 1 项、12 项和 100 项，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 10 章答案·数据生命周期：采集、清洗、字典与许可

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，从问题定义到删除策略管理数据，记录来源、字段含义、许可、缺失与偏差。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“食材从采购到储存都有标签和保质期，数据也需要来路、用途和退出机制。”能提供直觉，但它省略了数学

约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择校园图书馆的公开借阅类别，先用固定规则或简单统计量作为基线。把“缺失比例”设置为 0%、10%和 80%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 11 章答案 · SQL 与关系模型：让问题变成可核对查询

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解表、键、连接、聚合和事务，用查询结果回答明确问题。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“学校里学生表和课程表通

过学号、课程号连接，不必在每张表重复全部信息。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择不含个人信息的教学模拟天气记录，先用固定规则或简单统计量作为基线。把“匹配比例”设置为 0%、30%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 12 章答案·环境、随机种子与实验记录

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，区分可重复、可复现与可推广，固定依赖、种子、数据版本和测量方法。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“照菜谱重做

仍需相同原料、火力和计量工具；只有种子相同远远不够。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开道路网络与虚构出行任务，先用固定规则或简单统计量作为基线。把“重复运行次数”设置为 1 次、5 次和 20 次，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 13 章答案 · 向量：把方向、特征和相似度放进坐标

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解向量的长度、方向、基、坐标和距离，连接到样本特征表示；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“地图上的位移箭头

和一名学生的多科成绩都可写成有顺序的数。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成的植物叶片特征，先用固定规则或简单统计量作为基线。把“向量夹角”设置为 0° 、 45° 和 180° ，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 14 章答案·矩阵：一次描述许多线性关系

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，把矩阵理解为线性变换和批量方程，而不只是数字表格。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“同一块透明投影片会按统一

规则移动所有点，矩阵就是这套规则的紧凑记录。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择虚构游戏角色的状态日志，先用固定规则或简单统计量作为基线。把“缩放系数”设置为 $0.2\times$ 、 $1\times$ 和 $3\times$ ，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 15 章答案 · 投影、特征方向与 PCA

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解正交投影、特征向量、奇异值和主成分如何提取主要变化方向。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“手电把三维物体投

到墙面；选对墙面能保留更多形状信息。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开天文观测的缩小样例，先用固定规则或简单统计量作为基线。把“保留主成分”设置为 1 个、2 个和 10 个，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 16 章答案·张量、广播与数值线性代数

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，连接维度、形状、索引、广播和批处理，理解避免显式求逆等数值习惯。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“张量像按班

级、学生、科目叠成的多层表格；形状说明每一轴代表什么。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成音频波形与环境声标签，先用固定规则或简单统计量作为基线。把“批量大小”设置为 1、16 和 128，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 17 章答案 · 概率与随机变量：描述不知道的部分

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，区分频率、主观不确定性、事件、条件概率、期望和方差。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“天气预报的降雨概率不是

天空下了百分之三十的雨，而是对不确定事件的量化。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择校园能源使用的匿名汇总数据，先用固定规则或简单统计量作为基线。把“事件概率”设置为 0、0.5 和 1，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 18 章答案 · 分布与贝叶斯更新

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，认识常用分布、似然、先验和后验，理解证据如何更新信念。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“侦探先有多个假设，新线

索会改变各假设权重，但不会凭空保证某个假设正确。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开古诗文本与版权已明确的材料，先用固定规则或简单统计量作为基线。把“先验权重”设置为 0%、50%和 100%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 19 章答案 · 估计、区间与假设检验

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，区分参数、估计量、抽样误差、置信区间、效应量和 p 值。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“量一张桌子多次会有微小

差异；区间表达测量的不确定程度，而非隐藏误差。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择虚构机器人在二维网格中的轨迹，先用固定规则或简单统计量作为基线。把“置信水平”设置为 80%、95% 和 99%，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 20 章答案·熵、交叉熵、KL 与互信息

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，把信息量连接到不确定性压缩、分类损失和表示学习。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“猜一枚公平硬币需要更多信息，

猜几乎总是正面的硬币需要更少惊讶。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开科研论文的标题与摘要元数据，先用固定规则或简单统计量作为基线。把“分布尖锐度”设置为 0.1、1 和 5，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 21 章答案·导数、梯度与链式法则

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，从变化率理解导数，从最陡方向理解梯度，用链式法则拆解复合计算。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“走山路时坡

度告诉你向前一步高度怎样变；梯度是多方向坡度组成的箭头。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择合成电商目录但不含真实用户行为，先用固定规则或简单统计量作为基线。把“横坐标”设置为 $-5x$ 、 $1x$ 和 $5x$ ，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0分：只复述名词。
- 1分：说明了机制，但没有实验或边界。
- 2分：有基线、单变量、指标和失败分析。
- 3分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 22 章答案 · 雅可比、海森与自动微分

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解多输入多输出导数结构、曲率和计算图上的前向/反向传播。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“工厂流程图记录每个

零件如何影响成品，反向模式从最终误差倒查每个旋钮。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择校园图书馆的公开借阅类别，先用固定规则或简单统计量作为基线。把“计算图深度”设置为 1 层、12 层和 100 层，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 23 章答案 · 梯度下降、动量与 Adam

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，比较批量梯度、随机梯度、动量、自适应步长的假设和失败方式；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“下山时只看脚下坡

度会左右摇摆；动量记住最近方向，自适应方法调节各方向步幅。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择不含个人信息的教学模拟天气记录，先用固定规则或简单统计量作为基线。把“学习率”设置为 0.0001、0.03 和 1，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

第 24 章答案 · 数值稳定、条件数与非凸地形

答案用于自检，不是唯一表述。先保留自己的推理，再比较遗漏。

基础题参考

合格答案应包含三部分：第一，理解浮点误差、溢出、下溢、病态问题、局部结构和诊断方法。；第二，输入、内部机制和输出之间的关系；第三，至少一个适用假设。生活类比“用很短的尺量长距离会累积误差；问题本身敏感时，再精密的算法也需报告不确定性。”能提供直觉，但它省略了数学约束、数据偏差和部署风险，因此不能作为实验证据。若只写概念定义、没有边界或反例，答案仍不完整。

应用题参考

选择公开道路网络与虚构出行任务，先用固定规则或简单统计量作为基线。把“浮点位宽”设置为 4bit、32bit 和 64bit，其余条件保持一致。记录主要指标、资源成本、至少一个切片和失败样例。页面结果必须写“教学模拟”；如果自己运行真实代码，则保存环境、数据版本、种子和测量记录后才能写“实际测量”。

挑战题参考

可以主动改变输入分布、加入噪声、制造缺失、交换标签含义或改变部署资源。检测层使用输入校验、漂移或分布外分数；决策层设置置信阈值和拒答；流程层对高风险输出要求人工复核；运维层保留上一版本、变更记录和回滚触发条件。关键不是列工具名，而是说明每道控制对应哪个失败路径、由谁负责、怎样验证控制有效。

自评量规

- 0 分：只复述名词。
- 1 分：说明了机制，但没有实验或边界。
- 2 分：有基线、单变量、指标和失败分析。
- 3 分：进一步包含不确定性、迁移、人工闸门和可复现记录。

本卷来源台账

本卷共引用 23 条来源。正文为原创解释；不复制受版权保护的图表、习题答案或大段原文。访问日期均为当前 Beta 版核实记录。

[S001] AI competency framework for students

项目	内容
作者/机构	UNESCO
来源等级	国际组织一手框架
发布日期	2024
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 unesco.org

[S002] Five Big Ideas in AI

项目	内容
作者/机构	AI4K12 Initiative
来源等级	教育框架一手资料

项目	内容
发布日期	2020
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 ai4k12.org

[S003] Computer Science Curricula 2023

项目	内容
作者/机构	ACM, IEEE-CS, AAAI
来源等级	专业学会课程框架
发布日期	2024
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 csed.acm.org

[S004] Artificial Intelligence Risk Management Framework (AI RMF 1.0)

项目	内容
作者/机构	NIST
来源等级	标准机构一手资料
发布日期	2023
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 nist.gov

[S005] AI Index Report 2026

项目	内容
作者/机构	Stanford HAI
来源等级	研究机构年度报告
发布日期	2026
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名

项目	内容
DOI/URL	访问 hai.stanford.edu

[S006] Python 3 Documentation

项目	内容
作者/机构	Python Software Foundation
来源等级	官方文档
发布日期	持续更新
访问日期	2026-08-06
许可	PSF License
DOI/URL	访问 docs.python.org

[S007] Pro Git

项目	内容
作者/机构	Scott Chacon, Ben Straub
来源等级	官方托管开放教材
发布日期	持续更新

项目	内容
访问日期	2026-08-06
许可	CC BY-NC-SA 3.0
DOI/URL	访问 git-scm.com

[S008] SQLite Documentation

项目	内容
作者/机构	SQLite Project
来源等级	官方文档
发布日期	持续更新
访问日期	2026-08-06
许可	Public domain / blessing
DOI/URL	访问 sqlite.org

[S009] NumPy Documentation

项目	内容
作者/机构	NumPy Developers

项目	内容
来源等级	官方文档
发布日期	持续更新
访问日期	2026-08-06
许可	BSD-3-Clause
DOI/URL	访问 numpy.org

[S010] Dive into Deep Learning

项目	内容
作者/机构	Aston Zhang et al.
来源等级	开放教材
发布日期	持续更新
访问日期	2026-08-06
许可	CC BY-SA 4.0
DOI/URL	访问 d2l.ai

[S011] 18.06 Linear Algebra

项目	内容
作者/机构	MIT OpenCourseWare
来源等级	大学公开课一手资料
发布日期	2010
访问日期	2026-08-06
许可	CC BY-NC-SA
DOI/URL	访问 ocw.mit.edu

[S012] e-Handbook of Statistical Methods

项目	内容
作者/机构	NIST/SEMATECH
来源等级	标准机构参考手册
发布日期	持续更新
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 itl.nist.gov

[S013] A Mathematical Theory of Communication

项目	内容
作者/机构	Claude E. Shannon
来源等级	同行评审奠基论文
发布日期	1948
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 doi.org

[S014] Convex Optimization

项目	内容
作者/机构	Stephen Boyd, Lieven Vandenberghe
来源等级	大学开放教材
发布日期	2004
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名

项目	内容
DOI/URL	访问 web.stanford.edu

[S015] Automatic Differentiation with torch.autograd

项目	内容
作者/机构	PyTorch
来源等级	官方教程
发布日期	持续更新
访问日期	2026-08-06
许可	BSD-3-Clause
DOI/URL	访问 pytorch.org

[S019] Probabilistic Graphical Models

项目	内容
作者/机构	Daphne Koller, Nir Friedman
来源等级	经典教材出版信息
发布日期	2009

项目	内容
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 mitpress.mit.edu

[S020] scikit-learn User Guide

项目	内容
作者/机构	scikit-learn developers
来源等级	官方文档
发布日期	持续更新
访问日期	2026-08-06
许可	BSD-3-Clause
DOI/URL	访问 scikit-learn.org

[S021] Machine Learning Crash Course

项目	内容
作者/机构	Google

项目	内容
来源等级	官方公开课程
发布日期	持续更新
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 developers.google.com

[S025] PyTorch Tutorials

项目	内容
作者/机构	PyTorch
来源等级	官方教程
发布日期	持续更新
访问日期	2026-08-06
许可	BSD-3-Clause
DOI/URL	访问 pytorch.org

[S072] Datasheets for Datasets

项目	内容
作者/机构	Timnit Gebru et al.
来源等级	同行评审论文预印本
发布日期	2018
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 arxiv.org

[S080] Artifact Review and Badging

项目	内容
作者/机构	ACM
来源等级	专业学会复现规范
发布日期	持续更新
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名

项目	内容
	名
DOI/URL	访问 acm.org

[S082] Colaboratory FAQ

项目	内容
作者/机构	Google
来源等级	官方文档
发布日期	持续更新
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 research.google.com

[S086] Recommendation on the Ethics of Artificial Intelligence

项目	内容
作者/机构	UNESCO
来源等级	国际组织规范性文件

项目	内容
发布日期	2021
访问日期	2026-08-06
许可	以来源页许可为准；本书仅作原创解释、短引用与明确署名
DOI/URL	访问 unesco.org

全书术语表

共 81 个核心术语。定义用于建立共同语言，不能替代章节中的数学、实验和边界分析。

术语	英文	小白解释
人工智能	Artificial Intelligence	让机器执行通常需要感知、推理、学习、规划或生成等能力的研究与工程领域。
算法	Algorithm	把输入转换为输出的一组明确步骤；算法本身不等于训练后的模型。
模型	Model	从数据或规则中得到、用于预测或决策的数学结构及其参数。
参数	Parameter	训练过程中被优化的数值，例如神经网络权重。
超参数	Hyperparameter	由开发者或搜索过程设定、不是直接从单次梯度更新学得s的配置。
数据集	Dataset	带有明确来源、字段、许可和用途的一组样本。

术语	英文	小白解释
特征	Feature	模型使用的输入表示。
标签	Label	监督学习中希望模型预测的目标值。
训练	Training	用目标函数和优化方法调整模型参数的过程。
推理	Inference	用已经得到的模型对新输入计算输出。
损失函数	Loss function	把预测与目标之间的差异变成可优化数值的函数。
目标函数	Objective	训练或决策希望最大化或最小化的量，可能包含多个损失和约束。
梯度	Gradient	函数对多个变量的一阶变化率组成的向量。
泛化	Generalization	模型在未参与训练、但与目标场景相关的数据上保持能力。
过拟合	Overfitting	模型过度适应训练数据细节，导致新数据表现下降。

术语	英文	小白解释
欠拟合	Underfitting	模型或训练不足以捕捉任务中的重要结构。
正则化	Regularization	通过惩罚、噪声或结构限制降低过拟合风险的方法。
基线	Baseline	用于判断新方法是否真正改进的简单或既有方法。
消融实验	Ablation	移除或替换系统一部分，观察它对结果的贡献。
随机种子	Random seed	初始化伪随机数生成器的值；固定种子不是完整可复现性的替代品。
张量	Tensor	带有一个或多个轴的数值数组。
嵌入	Embedding	把离散对象映射到连续向量空间的表示。
注意力	Attention	根据查询与键的匹配程度加权汇总值的机制。
Transformer	—	以注意力、前馈层、残差和归一化为核心的序列模型架构。

术语	英文	小白解释
词元	Token	文本经过词元器切分后的基本单位，可是字符、子词或其他片段。
上下文窗口	Context window	模型一次计算中能够直接处理的词元范围。
基础模型	Foundation model	在广泛数据上训练、可适配多种下游任务的大规模模型。
语言模型	Language model	为词元序列分配概率或预测后续词元的模型。
检索增强生成	RAG	在生成前检索外部资料，并把证据提供给生成模型的系统模式。
Agent	—	围绕目标执行观察、计划、工具调用、验证和停止循环的系统。
提示注入	Prompt injection	不可信输入试图改变模型或 Agent 原定指令和权限的攻击。
强化学习	Reinforcement learning	智能体通过与环境交互、基于

术语	英文	小白解释
		奖励学习策略的方法。
策略	Policy	在给定状态下选择动作的规则或概率分布。
价值函数	Value function	从状态或状态动作对出发的预期累计回报。
马尔可夫决策过程	MDP	用状态、动作、转移、奖励和折扣描述序列决策的形式框架。
卷积	Convolution	用共享局部核在输入位置上滑动计算的运算。
计算机视觉	Computer vision	让机器从图像或视频中提取、生成和使用信息的领域。
自然语言处理	NLP	让计算机处理、理解、检索或生成人类语言的领域。
自动语音识别	ASR	把语音波形转换为文字或其他符号序列。
扩散模型	Diffusion model	通过学习逐步去噪过程生成样本的一类模型。

术语	英文	小白解释
生成对抗网络	GAN	通过生成器与判别器对抗训练学习生成分布的架构。
图神经网络	GNN	在图结构上通过邻域聚合或消息传递学习表示的模型。
校准	Calibration	预测概率与事件长期发生频率之间的一致程度。
分布偏移	Distribution shift	训练、评估与部署数据分布不一致的现象。
分布外	OOD	与训练或预期部署分布明显不同的输入。
因果推断	Causal inference	在明确假设下估计干预对结果影响的方法。
公平性	Fairness	围绕不同群体或个体的机会、错误、待遇和影响制定并审查的规范目标。
鲁棒性	Robustness	系统面对扰动、噪声、故障或环境变化仍保持可接受行为的能力。
可解释性	Explainability	为模型行为提供可理解说明的

术语	英文	小白解释
		方法和性质；解释清楚不等于预测正确。
差分隐私	Differential privacy	通过限制单条记录对输出分布的影响提供可量化隐私保证的框架。
联邦学习	Federated learning	在数据留在多个设备或组织的情况下协同训练的范式。
模型卡	Model card	记录模型用途、评估、限制和 risk 的结构化文档。
数据表	Datasheet	记录数据来源、构建、用途、限制和维护方式的结构化文档。
MLOps	—	管理机器学习数据、训练、部署、监控、版本和事故响应的工程实践。
延迟	Latency	一次请求从进入到完成所需的时间。
吞吐	Throughput	单位时间内系统能处理的请求或词元数量。

术语	英文	小白解释
量化	Quantization	用较低精度数值表示权重或激活以降低资源成本的方法。
蒸馏	Distillation	让较小模型学习较大模型输出或内部结构的方法。
并行训练	Distributed training	把训练计算、数据或参数分布到多个设备或进程。
检查点	Checkpoint	保存训练状态以便恢复、评估或部署的版本化工件。
可复现	Reproducible	他人依据充分工件和说明能够得到相容结果的性质。
同行评审	Peer review	由相关领域同行对方法、证据和表述进行独立审查。
置信区间	Confidence interval	按指定重复抽样程序构造、具有长期覆盖率解释的区间。
p 值	p-value	在零假设及检验假设成立时，观察到至少同样极端统计量的概率。
效应量	Effect size	差异或关联实际大小的度量。

术语	英文	小白解释
熵	Entropy	概率分布不确定性的度量。
交叉熵	Cross entropy	用一个分布编码另一个分布时的平均信息量，也是常见分类损失。
KL 散度	KL divergence	衡量一个概率分布相对另一个分布差异的非对称量。
贝叶斯更新	Bayesian update	用似然把先验信念更新为后验信念。
知识图谱	Knowledge graph	用实体、关系和属性构成的结构化知识网络。
本体	Ontology	对领域概念、关系和约束的显式规范。
搜索启发式	Heuristic	估计距离或代价、用于决定优先探索方向的函数。
约束满足	CSP	为变量选择值，使所有约束同时满足的问题。
SLAM	—	机器人同时定位并建立环境地图的问题。

术语	英文	小白解释
控制	Control	根据目标与反馈选择输入，使动态系统按期望演化。
仿真到现实	Sim-to-real	把在仿真中得到的方法迁移到真实系统的过程及差距问题。
世界模型	World model	学习环境状态和动作如何随时间演化的内部模型。
多模态	Multimodal	联合处理文本、图像、音频、视频或动作等多种信息形式。
AI for Science	—	使用 AI 辅助科学建模、模拟、实验设计、分析和发现的方向。
人机协同	Human-in-the-loop	在系统关键环节保留明确人工检查、判断或控制权的设计。
风险	Risk	事件发生可能性与后果的组合，必须结合具体情境评估。

版权、版本与勘误

2026.08 Beta · 第一卷：计算与数学地基 · 评审状态 draft。

Copyright © 2026 NEWVAR LAB. All rights reserved.

全部原创正文、实验定义和代码保留全部版权。第三方名称、论文与链接仅作明确署名的教学引用，其权利归原权利人。

本书仍处于 Beta：可能存在表达、事实、代码或排版问题。请通过 NEWVAR 版本页查看核实日期、评审状态与更正记录。

任何时效性结论都应在使用前回到一手来源重新核实；医疗、法律、财务、教育机会和安全关键场景必须由合格专业人员判断。

学习网站：[访问 newvarlab.com](https://newvarlab.com)）。