

前 言

这是一本面向中文读者的机器学习教科书, 为了使尽可能多的读者通过本书对机器学习有所了解, 作者试图尽可能少地使用数学知识. 然而, 少量的概率、统计、代数、优化、逻辑知识似乎不可避免. 因此, 本书更适合大学三年级以上的理工科本科生和研究生, 以及具有类似背景的对机器学习感兴趣的人士. 为方便读者, 本书附录给出了一些相关数学基础知识简介.

全书共 16 章, 大体上可分为 3 个部分: 第 1 部分包括第 1~3 章, 介绍机器学习基础知识; 第 2 部分包括第 4~10 章, 介绍一些经典而常用的机器学习方法; 第 3 部分包括第 11~16 章, 介绍一些进阶知识. 前 3 章之外的后续各章均相对独立, 读者可根据自己的兴趣和时间情况选择使用. 根据课时情况, 一个学期的本科生课程可考虑讲授前 9 章或前 10 章; 研究生课程则不妨使用全书.

书中除第 1 章外, 每章都给出了十道习题. 有的习题是帮助读者巩固本章学习, 有的是为了引导读者扩展相关知识. 一学期的一般课程可使用这些习题, 再辅以两到三个针对具体数据集的大作业. 带星号的习题则有相当难度, 有些并无现成答案, 谨供富有进取心的读者启发思考.

本书在内容上尽可能涵盖机器学习基础知识的各方面, 但作为机器学习入门读物且因授课时间的考虑, 很多重要、前沿的材料未能覆盖, 即便覆盖到的部分也仅是管中窥豹, 更多的内容留待读者在进阶课程中学习. 为便于有兴趣的读者进一步钻研探索, 本书每章均介绍了一些阅读材料, 谨供读者参考.

笔者以为, 对学科相关的重要人物和事件有一定了解, 将会增进读者对该学科的认识. 本书在每章最后都写了一个与该章内容相关的小故事, 希望有助于读者增广见闻, 并且在紧张的学习过程中稍微放松调剂一下.

书中不可避免地涉及大量外国人名, 若全部译为中文, 则读者在日后进一步阅读文献时或许会对不少人名产生陌生感, 不利于进一步学习. 因此, 本书仅对一般读者耳熟能详的名字如“图灵”等加以直接使用, 对故事中的一些主要人物给出了译名, 其他则保持外文名.

机器学习发展极迅速, 目前已成为一个广袤的学科, 罕有人士能对其众多分支领域均有精深理解. 笔者自认才疏学浅, 仅略知皮毛, 更兼时间和精力所限, 书中错谬之处在所难免, 若蒙读者诸君不吝告知, 将不胜感激.

周志华

2015 年 6 月

目 录

第 1 章 绪论	1
1.1 引言	1
1.2 基本术语	2
1.3 假设空间	4
1.4 归纳偏好	6
1.5 发展历程	10
1.6 应用现状	13
1.7 阅读材料	16
习题	19
参考文献	20
休息一会儿	22
第 2 章 模型评估与选择	23
2.1 经验误差与过拟合	23
2.2 评估方法	24
2.3 性能度量	28
2.4 比较检验	37
2.5 偏差与方差	44
2.6 阅读材料	46
习题	48
参考文献	49
休息一会儿	51
第 3 章 线性模型	53
3.1 基本形式	53
3.2 线性回归	53
3.3 对数几率回归	57
3.4 线性判别分析	60
3.5 多分类学习	63

3.6 类别不平衡问题	66
3.7 阅读材料	67
习题	69
参考文献	70
休息一会儿	72
第 4 章 决策树	73
4.1 基本流程	73
4.2 划分选择	75
4.3 剪枝处理	79
4.4 连续与缺失值	83
4.5 多变量决策树	88
4.6 阅读材料	92
习题	93
参考文献	94
休息一会儿	95
第 5 章 神经网络	97
5.1 神经元模型	97
5.2 感知机与多层网络	98
5.3 误差逆传播算法	101
5.4 全局最小与局部极小	106
5.5 其他常见神经网络	108
5.6 深度学习	113
5.7 阅读材料	115
习题	116
参考文献	117
休息一会儿	120
第 6 章 支持向量机	121
6.1 间隔与支持向量	121
6.2 对偶问题	123
6.3 核函数	126
6.4 软间隔与正则化	129
6.5 支持向量回归	133

6.6 核方法	137
6.7 阅读材料	139
习题	141
参考文献	142
休息一会儿	145
第 7 章 贝叶斯分类器	147
7.1 贝叶斯决策论	147
7.2 极大似然估计	149
7.3 朴素贝叶斯分类器	150
7.4 半朴素贝叶斯分类器	154
7.5 贝叶斯网	156
7.6 EM算法	162
7.7 阅读材料	164
习题	166
参考文献	167
休息一会儿	169
第 8 章 集成学习	171
8.1 个体与集成	171
8.2 Boosting	173
8.3 Bagging与随机森林	178
8.4 结合策略	181
8.5 多样性	185
8.6 阅读材料	190
习题	192
参考文献	193
休息一会儿	196
第 9 章 聚 类	197
9.1 聚类任务	197
9.2 性能度量	197
9.3 距离计算	199
9.4 原型聚类	202
9.5 密度聚类	211

9.6 层次聚类	214
9.7 阅读材料	217
习题	220
参考文献	221
休息一会儿	224
第 10 章 降维与度量学习	225
10.1 k 近邻学习	225
10.2 低维嵌入	226
10.3 主成分分析	229
10.4 核化线性降维	232
10.5 流形学习	234
10.6 度量学习	237
10.7 阅读材料	240
习题	242
参考文献	243
休息一会儿	246
第 11 章 特征选择与稀疏学习	247
11.1 子集搜索与评价	247
11.2 过滤式选择	249
11.3 包裹式选择	250
11.4 嵌入式选择与 L_1 正则化	252
11.5 稀疏表示与字典学习	254
11.6 压缩感知	257
11.7 阅读材料	260
习题	262
参考文献	263
休息一会儿	266
第 12 章 计算学习理论	267
12.1 基础知识	267
12.2 PAC学习	268
12.3 有限假设空间	270
12.4 VC维	273

12.5 Rademacher复杂度	279
12.6 稳定性	284
12.7 阅读材料	287
习题	289
参考文献	290
休息一会儿	292
第 13 章 半监督学习	293
13.1 未标记样本	293
13.2 生成式方法	295
13.3 半监督SVM	298
13.4 图半监督学习	300
13.5 基于分歧的方法	304
13.6 半监督聚类	307
13.7 阅读材料	311
习题	313
参考文献	314
休息一会儿	317
第 14 章 概率图模型	319
14.1 隐马尔可夫模型	319
14.2 马尔可夫随机场	322
14.3 条件随机场	325
14.4 学习与推断	328
14.5 近似推断	331
14.6 话题模型	337
14.7 阅读材料	339
习题	341
参考文献	342
休息一会儿	345
第 15 章 规则学习	347
15.1 基本概念	347
15.2 序贯覆盖	349
15.3 剪枝优化	352

15.4 一阶规则学习	354
15.5 归纳逻辑程序设计	357
15.6 阅读材料	363
习题	365
参考文献	366
休息一会儿	369
第 16 章 强化学习	371
16.1 任务与奖赏	371
16.2 K -摇臂赌博机	373
16.3 有模型学习	377
16.4 免模型学习	382
16.5 值函数近似	388
16.6 模仿学习	390
16.7 阅读材料	393
习题	394
参考文献	395
休息一会儿	397
附录	399
A 矩阵	399
B 优化	403
C 概率分布	409