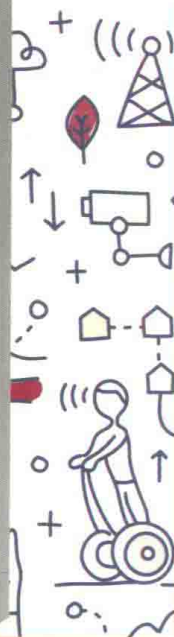
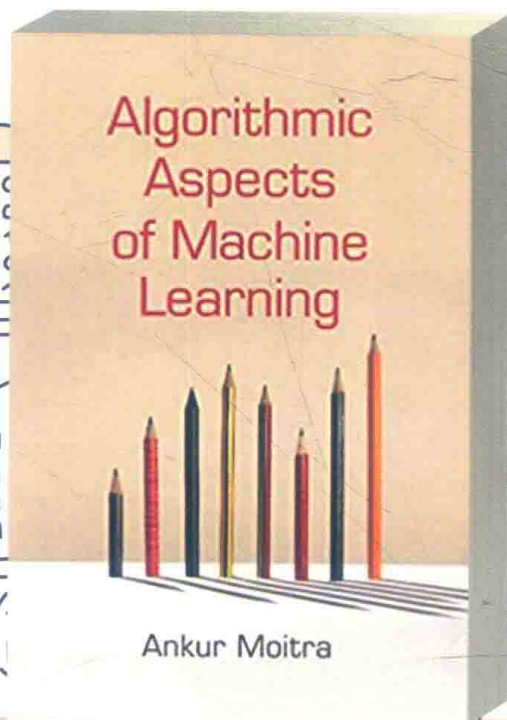
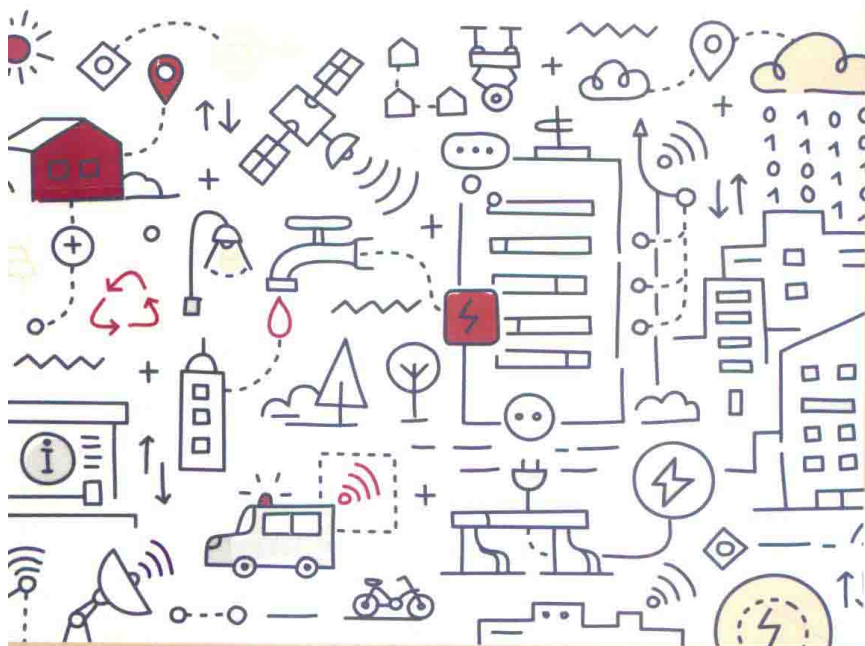


Algorithmic Aspects of Machine Learning

机器学习算法

[美] 安柯·莫特拉 (Ankur Moitra) 著
麻省理工学院

庄福振 赵朋朋 译
北京航空航天大学 苏州大学



Algorithmic Aspects of Machine Learning

机器学习算法

[美] 安柯·莫特拉 (Ankur Moitra) 著
麻省理工学院

庄福振 赵朋朋 译
北京航空航天大学 苏州大学



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

机器学习算法 / (美) 安柯·莫特拉 (Ankur Moitra) 著; 庄福振, 赵朋朋译. -- 北京: 机械工业出版社, 2021.4

(智能科学与技术丛书)

书名原文: Algorithmic Aspects of Machine Learning

ISBN 978-7-111-68048-2

I. ①机… II. ①安… ②庄… ③赵… III. ①机器学习-算法 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2021) 第 070722 号

本书版权登记号: 图字 01-2020-7584

This is a Simplified-Chinese edition of the following title published by Cambridge University Press:

Ankur Moitra, *Algorithmic Aspects of Machine Learning*, ISBN 978-1-107-18458-9.

© Ankur Moitra 2018

This Simplified-Chinese edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2021.

This Simplified-Chinese edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

Copies of this book sold without a Cambridge University Press sticker on the cover are unauthorized and illegal.

本书原版由剑桥大学出版社出版。

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售。

本书封面贴有 Cambridge University Press 防伪标签, 无标签者不得销售。

本书探索了理论计算机科学和机器学习这两个领域能够互相借鉴的知识。书中介绍了机器学习中的重要模型和主要问题, 并以一种容易理解的方式介绍了该领域的前沿研究成果以及现代算法工具, 包括矩量法、张量分解法和凸规划松弛法。本书共 8 章, 内容涵盖非负矩阵分解、主题模型、张量分解、稀疏恢复、稀疏编码、高斯混合模型和矩阵补全等。

本书适合理论计算机科学家、机器学习研究人员以及相关专业的学生阅读和学习。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 王春华 孙榕舒

责任校对: 殷虹

印刷: 三河市宏图印务有限公司

版次: 2021 年 5 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 12.5

书号: ISBN 978-7-111-68048-2

定价: 79.00 元

客服电话: (010) 88361066 88379833 68326294

投稿热线: (010) 88379604

华章网站: www.hzbook.com

读者信箱: hzsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

作者简介

安柯·莫特拉

(Ankur Moitra)

麻省理工学院洛克威尔国际数学系副教授。他是麻省理工学院计算机科学和人工智能实验室 (CSAIL) 的首席研究员，也是计算理论小组、机器学习和统计中心的核心成员。他的工作目标是开发在行为推理方面具有可证明的保证和基础的算法，来在理论计算机科学和机器学习之间架起一座桥梁。他是帕克德奖学金、斯隆奖学金、美国国家科学基金会 (NSF) 终身成就奖、NSF 计算与创新奖学金和赫兹奖学金的获得者。

译者简介



庄福振

北京航空航天大学研究员、博士生导师，于2011年7月在中国科学院研究生院获得博士学位。主要从事机器学习、数据挖掘，包括迁移学习、多任务学习、推荐系统以及大数据挖掘应用等方面的研究，相关成果已经在本领域顶级、重要国际期刊和国际会议上发表/录用论文100余篇。迁移学习的工作曾获得SDM2010和CIKM2010的最佳论文提名，2013年获得中国人工智能学会优秀博士学位论文奖。入选2015年微软亚洲研究院青年教师“铸星计划”，2017年入选中国科学院青年创新促进会。



赵朋朋

苏州大学教授、博士生导师，苏州工业园区科技领军人才，CCF高级会员，CCF人工智能与模式识别专委会委员，CCF大数据专家委员会通讯委员。近年来，主持国家自然科学基金项目2项、省市级项目5项，获苏州市科技进步奖2项，参加国家自然科学基金项目、教育部重点项目、江苏省重大科技支撑与自主创新项目等6项，主持多项企业/国防横向项目研发，包括推荐引擎、流式数据处理、大数据分析等。在相关学术会议和期刊上发表论文100余篇，其中包括50余篇TKDE、AAAI、IJCAI、WWW、ACM MM、ICDM等CCF-A/B类论文，被SCI、EI索引收录100余篇；申请国家发明专利30余项，其中20余项已获授权。担任AAAI、IJCAI、CIKM、DASFAA、PAKDD等权威国际会议程序委员会委员。主要研究方向包括推荐系统、数据挖掘、深度学习、大数据分析等。



华章图书

一本打开的书，一扇开启的门，
通向科学殿堂的阶梯，托起一流人才的基石。

随着人工智能在近十年的不断兴起，以及计算科学技术的发展进步，学术界/工业界对机器学习算法的研究逐渐深入，很多实际应用问题得以解决。机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。机器学习是人工智能的核心，主要研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断完善。

虽然机器学习的覆盖范围比较广，但本书主要集中在处理矩阵数据的矩阵运算算法方面，专业性比较强。相对于以往专注于机器学习理论以及机器学习算法实践方面的书籍，本书应该介于这两者之间，目的是使读者针对算法“知其然且知其所以然”。本书可以作为相关专业本科高年级学生或研究生的教材。

非常感谢本书作者的信任和机械工业出版社的委托，本书的翻译是由我和苏州大学的赵朋朋老师合作完成的。我们两个人的研究方向都是机器学习、数据挖掘，包括迁移学习、推荐系统等。还有一些研究生也参与了本书的翻译工作，他们是（排名不分先后）：郭庆宇、沈丹瑶、齐志远、罗安靖、赵静。出版社的编辑在排版和校对方面给予了我们极大的帮助。感谢他们为本书的出版付出的努力！

由于时间仓促以及一些翻译习惯上的差异，本书难免存在瑕疵，在此谨致歉意。若有发现，请及时反馈给我或出版社以进行修正，本人将不胜感激。

庄福振

2021年1月

前 言

Algorithmic Aspects of Machine Learning

本书基于麻省理工学院在 2013 年秋季、2015 年春季和 2017 年秋季开设的“Algorithmic Aspects of Machine Learning”课程。感谢学习本课程并使之成为一次美妙体验的所有学生。

译者序	4.5 练习	88
前言		
第 1 章 引言	第 5 章 稀疏恢复	91
1	5.1 介绍	91
第 2 章 非负矩阵分解	5.2 非相干性和不确定性	
5	原理	95
2.1 介绍	5.3 追踪算法	99
5	5.4 Prony 方法	102
2.2 代数算法	5.5 压缩感知	107
13	5.6 练习	113
2.3 稳定性和可分离性		
20	第 6 章 稀疏编码	115
2.4 主题模型	6.1 介绍	115
28	6.2 不完备情况	118
2.5 练习	6.3 梯度下降	124
34	6.4 过完备情况	130
第 3 章 张量分解：算法	6.5 练习	136
37		
3.1 旋转问题	第 7 章 高斯混合模型	137
37	7.1 介绍	137
3.2 张量入门	7.2 基于聚类的算法	142
39	7.3 密度估计的讨论	147
3.3 Jennrich 算法	7.4 无聚类算法	151
44	7.5 单变量算法	157
3.4 矩阵摄动界		
50		
3.5 练习		
59		
第 4 章 张量分解：应用		
61		
4.1 进化树和隐马尔可夫		
模型		61
4.2 社区发现		70
4.3 扩展到混合模型		75
4.4 独立成分分析		83

7.6 代数几何视图	162	8.2 核范数	173
7.7 练习	167	8.3 量子高尔夫	178
第 8 章 矩阵补全	169	参考文献	184
8.1 介绍	169	索引	191

·第1章·

引言

机器学习正在开始接管我们生活中方方面面的决策，包括：

- 1) 保证我们在自动驾驶汽车上的日常通勤安全。
- 2) 根据我们的症状和病史做出准确的诊断。
- 3) 定价和交易复杂证券。
- 4) 发现新科学，例如各种疾病的遗传基础。

但是令人惊讶的事实是，这些算法在没有任何可证明的行为保证下工作。当面临一个优化问题时，它们是否真的找到了最优解，或者只是一个不错的解？当它们假定一个概率模型时，是否能够从真实的后验分布中纳入新的证据和样本？机器学习在实践中非常有效，但这并不意味着我们理解它为什么如此有效。

如果你已经上过传统算法课程，通常会通过最坏情况分析来考虑算法。当你使用一个排序算法时，会根据其在最坏可能输入情况下执行的操作次数来衡量其运行时间。这是一种很方便的约束类型，因为这意味着你可以对你的算法所需的时间进行有意义的说明，而不必担心你通常提供给它的输入类型。

但是，机器学习算法(尤其是现代算法)的分析如此具有挑战性

的原因是，它们试图解决的问题类型在最坏输入情况下实际上是 NP-hard 问题。当你将寻找最适合你的数据的参数的问题转换为优化问题时，有时会像 NP-hard 问题一样很难找到一个好的解。当你假定一个概率模型并想使用它来进行推理时，有时也会遇到 NP-hard 的情况。

1

在本书中，我们将通过尝试为数据找到更现实的模型，来解决为机器学习提供可证明保证的问题。在许多应用中，我们可以根据问题的出现背景进行合理的假设，从而绕过这些最坏情况的障碍，并严格分析实际中使用的启发式方法，以及从根本上设计出解决机器学习中一些核心、重复性问题的新方法。

退一步讲，越过最坏情况分析的想法与理论计算机科学本身一样古老^{[95]⊖}。实际上，理解“典型”实例上算法的行为，有许多不同的含义，包括：

1) 输入的概率模型，甚至是结合最坏情况和平均情况分析的混合模型，如半随机模型^[38,71]或平滑分析^[39,130]。

2) 度量问题复杂度以及要求算法针对简单输入速度快的方式，例如参数化复杂度^[66]。

3) 稳定性概念，它试图阐明问题的哪些实例具有有意义的答案，以及哪些实例正是你实际想要解决的^[20,32]。

这绝不是主题或参考的详尽列表。无论如何，在本书中，我们将通过关于如何解决棘手问题的见解来处理机器学习问题。

最终，我们希望理论计算机科学和机器学习之间有很多可以互相借鉴的地方。理解为何用期望最大化或梯度下降等启发式方法处理非凸函数在实际中如此有效，对于理论计算机科学而言是一大挑

⊖ 毕竟，在真实输入上表现良好的启发式方法也是古老的（比现代机器学习早很多），因此需要解释它们。

战。但要在这些问题上取得进展，我们需要明白在机器学习背景下，哪种类型的模型和假设具有意义。另外，如果我们在这些困难的问题上取得进展，并探索出关于启发式方法为什么如此有效的新见解，就有希望对它们进行更好的设计。我们甚至可以希望找到全新的方法来解决机器学习的一些重要问题，尤其是通过利用算法工具包中的现代工具。

在本书中，我们将涵盖以下主题：

- 1) 非负矩阵分解。
- 2) 主题模型。
- 3) 张量分解。
- 4) 稀疏恢复。
- 5) 稀疏编码。
- 6) 高斯混合模型。
- 7) 矩阵补全。

2

希望随着该领域的发展和新发现的涌现，在后续的版本中能够加入更多的章节。

3

· 第 2 章 ·

非负矩阵分解

本章将探讨非负矩阵分解问题。首先，我们将对比非负矩阵分解问题与我们熟知的奇异值分解问题，这有利于我们更好地理解这一问题。在最坏的情况下，非负矩阵分解是 NP-hard 问题(说真的，你还期望什么呢?)，但是我们将做出特定领域的假设(称为可分离性)，这将使我们可以为这一问题中的一种重要的特殊情况提供可证明的算法。然后我们将运用该算法来学习主题模型的参数。这将是我们遇到的第一个针对计算难解型问题寻找解决思路的案例。

2.1 介绍

为了更好地理解非负矩阵分解问题背后的动机，以及为什么它在实际应用中有用，我们首先介绍奇异值分解(Singular Value Decomposition, SVD)，然后将两者进行比较。最终，这两种方法都将被应用到本节最后的文本分析问题中。

奇异值分解

奇异值分解是线性代数中最强大的工具之一，给定一个 $m \times n$

的矩阵 \mathbf{M} ，它的奇异值分解被写成

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

4 其中 \mathbf{U} 和 \mathbf{V} 是正交矩阵， $\mathbf{\Sigma}$ 是非负对角矩阵。我们也可以把上式写成

$$\mathbf{M} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

其中 \mathbf{u}_i 是 \mathbf{U} 的第 i 列， \mathbf{v}_i 是 \mathbf{V} 的第 i 列， σ_i 是 $\mathbf{\Sigma}$ 对角线上的第 i 个元素。在本章中，我们规定 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ 。在这种情况下， \mathbf{M} 的秩即为 r 。

在本书中，我们将会使用上述的分解形式，以及矩阵的特征分解(eigendecomposition)。如果 \mathbf{M} 是一个 $n \times n$ 的可对角化矩阵，它的特征分解可以写作

$$\mathbf{M} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$$

其中 \mathbf{D} 是对角矩阵。现在，有以下几点需要注意：

1) **存在性**：所有的矩阵都有奇异值分解，只有方阵才有特征分解。而且，不是所有的方阵都是可对角化的， \mathbf{M} 可对角化的充分条件是它所有的特征值都不相同。

2) **算法**：这两种矩阵分解方法都能被有效地计算。计算奇异值分解的最佳通用算法的时间复杂度是 $O(mn^2)$ ，如果 $m \geq n$ 。对于稀疏矩阵来说，还有更快的算法。计算矩阵特征分解的时间复杂度是 $O(n^3)$ ，该算法可以基于快速矩阵乘法进一步改善，尽管此类算法的稳定性和实用性尚不确定。

3) **唯一性**：奇异值分解是唯一的当且仅当矩阵的奇异值不相同。类似地，特征分解是唯一的当且仅当矩阵的特征值不相同。在某些情况下，我们只需要非零奇异值/特征值各不相同，因为可以忽略其他部分。