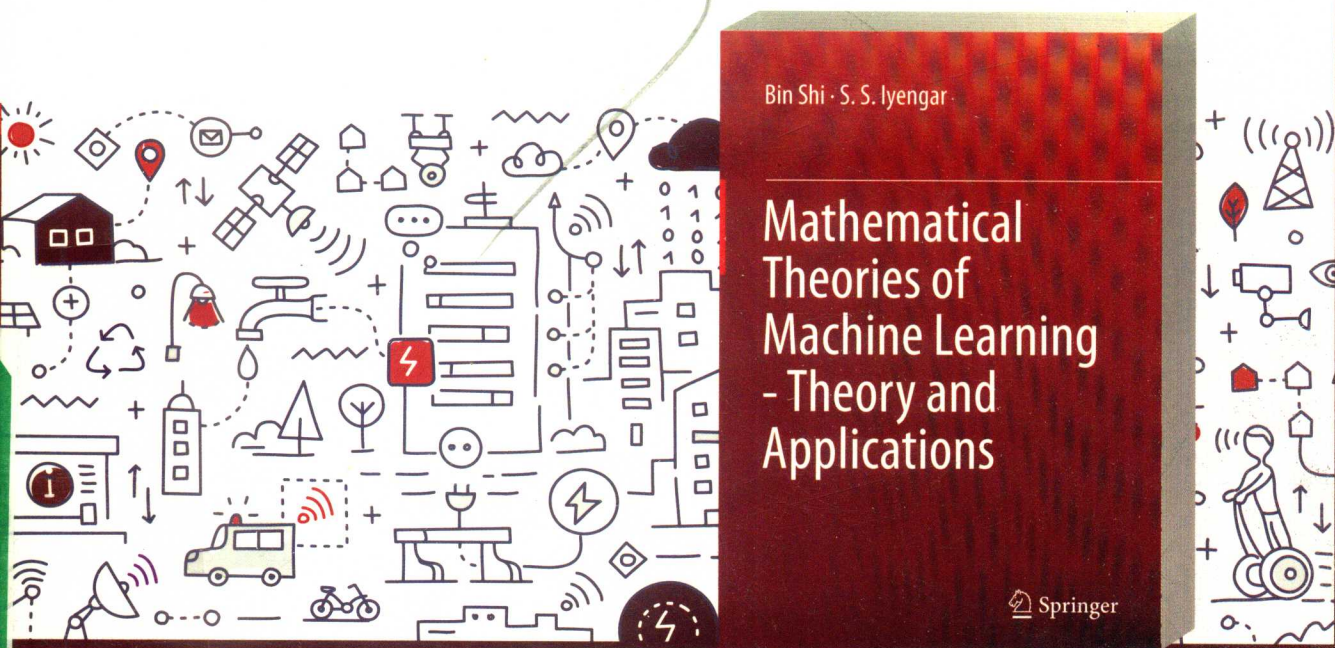


# Mathematical Theories of Machine Learning - Theory and Applications

## 机器学习的数学理论

[中] 史斌 (Bin Shi)  
[美] S.S. 艾扬格 (S.S.Iyengar) © 著

李飞 © 等译





## 图书在版编目 ( CIP ) 数据

机器学习的数学理论 / 史斌, (美) S. S. 艾扬格 (S. S. Iyengar) 著; 李飞等译. —北京: 机械工业出版社, 2020.8

(智能科学与技术丛书)

书名原文: Mathematical Theories of Machine Learning—Theory and Applications

ISBN 978-7-111-66136-8

I. 机… II. ①史… ②S… ③李… III. 机器学习—数学理论 IV. TP181

中国版本图书馆 CIP 数据核字 (2020) 第 128849 号

本书版权登记号: 图字 01-2020-1327

First published in English under the title

Mathematical Theories of Machine Learning—Theory and Applications

by Bin Shi and S. S. Iyengar

Copyright © Springer Nature Switzerland AG 2020

This edition has been translated and published under licence from

Springer Nature Switzerland AG

All Rights Reserved

本书中文简体字版由 Springer 授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

本书重点研究机器学习的数学理论。第一部分探讨了在非凸优化问题中, 选择梯度下降步长来避免严格鞍点的最优性和自适应性。第二部分提出了在非凸优化中寻找局部极小值的算法, 并利用牛顿第二定律在一定程度上得到无摩擦的全局极小值。第三部分研究了含有噪声和缺失数据的子空间聚类问题, 这是一个由随机高斯噪声的实际应用数据和含有均匀缺失项的不完全数据激发的问题; 还提出了一种新的具有粘性网正则化的 VAR 模型及其等价贝叶斯模型, 该模型既考虑了稳定的稀疏性, 又考虑了群体选择。

本书可作为本科生或研究生的入门教材。对于希望进一步加强对机器学习的理解的教授、行业专家和独立研究人员来说, 本书也是极佳的选择。

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: 柯敬贤

责任校对: 李秋荣

印刷: 三河市宏图印务有限公司

版次: 2020 年 8 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 10.5

书号: ISBN 978-7-111-66136-8

定价: 69.00 元

客服电话: (010) 88361066 88379833 68326294

投稿热线: (010) 88379604

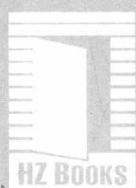
华章网站: www.hzbook.com

读者信箱: hzjsj@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



## 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。

随着科技的发展和经济的进步，人工智能技术的发展所带来的效益愈发凸显，越来越多的企业和学者将目光转向人工智能相关技术。机器学习算法作为最基础的人工智能算法之一，对人工智能技术的发展起到了关键作用。研究机器学习的数学基础具有非凡的意义，能够为深度学习等人工智能技术提供源源不断的前进动力，因此该领域的研究一直是非常有吸引力的。

作者史斌博士在机器学习的数学基础算法方面有着很深的造诣，其研究成果也得到了很多学者的引用和认可。本书总结了史斌博士的部分研究成果，定能推动机器学习技术的发展。但是本书专业性较强，需要读者具有一定的数学基础。另外，在阅读本书之前读者最好具备一定的机器学习算法基础。当然，机器学习初学者也可以通过本书来快速了解机器学习相关算法中非常核心的研究领域。

本书的翻译工作由海军航空大学和 92212 部队的机器学习研究者共同完成。由于工作繁忙，所以对翻译工作进行了分工，其中第 1~3 章由赵文飞博士完成，第 4 章由王希彬博士完成，第 5 章由刘涛博士完成，第 6 章由刘伟博士完成，第 7 章由甄伟完成，第 8~11 章由李飞博士完成。全书由李飞和赵文飞进行了翻译风格上的统一。

由于本书专业性强，个别专业术语的翻译可能无法完全体现作者的原意，衷心希望各位读者批评指正。

译者

2020 年 4 月

## 序言

Mathematical Theories of Machine Learning-Theory and Applications

本书将对机器学习领域产生重大影响。目前已经有一些书讨论了不同类别的机器学习技术，而本书深入研究的是机器学习算法的数学基础。这是很有必要的，因为从业者和学者都必须有一种方法来衡量大量算法应用的有效性。

本书的主要贡献之一是讨论了凸约束-稀疏子空间聚类(CoCoSSC)。一些机器学习方法的优劣取决于最速下降方法的收敛性，当目标函数为非凸目标(或凸约束目标)时，CoCoSSC方法设计的梯度下降方法具有更快的收敛性。

有许多应用将受益于这一基础工作，应用于网络安全的机器学习就是这样一种应用。在实际应用中，其目标是减少网络分析师无法承受的数据量。具体而言，有一些例子表明，基于最速梯度下降的逻辑回归分类器有助于在数据库里将相关的网络主题与非网络主题分离开来。另一个类似的应用是识别被利用的恶意软件，该恶意软件是大型漏洞数据库的子集。

此外，人工智能有可能给许多行业带来革命性的改变，例如无人驾驶汽车、金融、国家安全、医药和电子商务等应用领域。本书将深入挖掘以上应用中所蕴含的凸约束优化技术的数学原理，该原理同样适用于作为机器学习算法基础的最速下降优化。

戴维·R. 马丁内斯  
波士顿，马萨诸塞州

非常感谢李涛教授对机器学习的贡献，以及他在本书内容编写的早期阶段给予的指导和建议。也特别感谢来自卡内基梅隆大学、加州大学圣塔芭芭拉分校、迈阿密大学以及南加州大学的所有合作者，他们关于本书给出了宝贵的意见和建议。正是这些意见和建议，激励着我们成功地完成了本书。更重要的是，要感谢 Yining Wang、Jason Lee、Simon S. Du、Yuxiang Wang、Yudong Tau 和 Wentao Wang 对本书的评论和贡献，其中许多研究成果已提交各种会议和期刊发表。所以，感谢这些出版商的支持。

还要对佛罗里达大学的所有其他合作教师(特别是 Kianoosh G. Boroojeni 博士)以及 Sanjeev Kaushik Ramani 等研究生表示衷心的感谢，本书顺利出版离不开他们的帮助和支持。感谢 IBM Yorktown Heights、Xerox、美国国家科学基金会和其他机构为开展这项研究提供的资金支持。本书的大部分内容都来自史斌的研究工作。

## 前言

Mathematical Theories of Machine Learning-Theory and Applications

机器学习是一种核心的、变革性的方式，通过它，我们可以重新思考我们正在做的一切。我们正在深思熟虑地将它应用到所有的产品中，无论是搜索、广告、YouTube 还是游戏。虽然刚刚起步，但你会看到我们如何系统地思考将机器学习应用到所有这些领域。

——Sundar Pichai, 谷歌首席执行官

机器学习及相关技术是最有趣的研究课题之一，它有可能改变世界的发展方向。然而，在目前的研究现状中，机器学习的研究还没有一个坚实的理论框架，不能为分析提供基础，也不能为实验运行提供指导。本书试图确定并解决在现代机器学习、人工智能、深度神经网络等方面具有重大研究兴趣的各个领域中存在的问题，这些技术可以完成非凡的任务，但是如何使用它们高度依赖的基本概念仍然是一个谜。梯度下降法是一种广泛应用于神经网络训练的方法。当使用梯度下降法时，无论是收敛到局部最小值还是全局最小值，都存在的一个挑战是缺乏关于该算法何时收敛的指导性准则。本书试图解决这个关键问题。本书为读者提供了新的理论框架，可以用于收敛性分析。

本书也代表了作者和合作者在机器学习领域数学方面的重大贡献。在整本书中，我们确保读者能够很好地理解和感受梯度下降技术的理论框架，以及在神经网络训练中使用这些理论框架的方法。为了强调这一点，书中使用了我们最近的一些研究成果，以及其他研究人员正在探索的综合成果。当阅读本书的各个章节时，读者会接触到各种非常重要的应用，比如子空间聚类和时间序列分析。本书力求达到理论与应用的平衡，因此，书中会同时给出理论以及相关应用。我们希望在机器学习领域为读者提供正确的工具，使阅读更加精彩，同时对读者产生巨大的影响。

与诸如 Goodfellow、Bengio 和 Courville 的《深度学习》等现有书籍相比，本书更深入地定义和展示了梯度下降领域的最新研究成果，使之成为学生和专业人士更为全面的工具。此外，本书还将这些概念与诸如子空间聚类和时间序列数据之类的应用联系起来，使其成为该领域中更好的选择。

本书的目标读者涵盖从事机器学习的所有人，无论是学生、教授、行业专家，还是独立研究人员。编撰本书的目的是为日常研究活动提供一本方便的手册。

本书分为几个独立的部分，以便读者首先接触到机器学习、神经网络、优化、梯度下降法等基本概念。在接下来的内容中，读者可以学习和理解选择梯度下降步长的最优性和自适应性，从而避开非凸优化问题中的严格鞍点。当所有鞍点都是严格的时，我们首先给出寻找局部最小值的梯度下降法的一个最大允许固定步长，它是梯度利普希茨常数( $1/L$ )的 2 倍。虽然在最坏情况下步长大于  $2/L$  的梯度下降法发散，但是对于严格的鞍形非凸优化问题，我们同样得到了梯度下降法的最优步长。其中一个重要的结果是只要梯度下降的诱导映射是局部微分同胚的，就可以确保算法收敛到严格鞍点的勒贝格测度为 0，而以前的研究工作都要求这个映射是全局微分同胚的。其次，我们还考虑了步长的自适应选择，证明如果每次迭代的步长与局部梯度利普希茨常数的倒数成正比，梯度下降法不会收敛到任何严格鞍点。据我们所知，这是第一个揭示变步长梯度下降法也可以避开鞍点的研究成果，应用动力系统理论中 Hartman 积映射定理的推广可以证明这一点。

本书还定义和阐述了用于在非凸优化方案中寻找局部最小值的算法，从而帮助我们获得在某种程度上符合无摩擦牛顿第二定律的全局最小值。基于辛欧拉算法，以运动中可观察和可控制的速度为关键观测量，模拟了无摩擦的牛顿第二定律，并从解析解的直观分析出发，对该算法的高速收敛性进行了理论分析。最后，给出了高维强凸、非强凸和非凸函数的实验结果。本书还描述了一些离散算法，这些算法将用于测试速度或动能的可观测性和可控性，以及人工耗散能量。

此后又研究了含有噪声和缺失数据的问题子空间聚类，这是一个很有实际应用价值的问题。考虑到应用中具有随机高斯噪声和具有一致缺失项的不完整数据，我们的主要贡献是 CoCoSSC——一种受 CoCoLasso 启发的新颖的噪声子空间聚类方法。值得注意的是，CoCoSSC 在将输入数据传递到 Lasso SSC 算法之前，使用了一种基于半正定规划的预处理步骤来“去偏”和“去噪”，这使得它更加稳定，并且是一个  $L_1$  标准化的自回归模型。我们从理论上证明了即使有  $1 - \Omega(n^{-2/5})$  比例的数据缺失，同时又被信噪比(SnR)为  $n^{-1/4}$  的加性高斯噪声干扰，CoCoSSC 仍能正常工作。与已知的只能处理恒定比例的数据丢失和  $n^{-1/6}$  的高斯噪声信噪比的算法相比，CoCoSSC 算法的效率有了显著的改善。与现有的粒子学习方法相比，我们的方法改进了粒子学习的样本完全推理策略。对合成的和实际的时间序列数据的大量实证研究，表明了该方法的有效性和高效率，同时有效的数值计算结果也证明了我们提出的算法的有效性和高效率。

史斌，加州大学伯克利分校

S. S. 艾扬格，迈阿密大学

**史斌(Bin Shi)博士** 目前是加州大学伯克利分校的博士后研究员。他的研究重点是机器学习理论，特别是机器学习中的优化理论。史斌博士 2006 年毕业于中国海洋大学应用数学专业，获理学学士学位；2008 年至 2011 年师从复旦大学袁小平教授学习现代常微分方程理论，并接受严格的数学训练；2011 年获复旦大学数学专业和麻省大学达特茅斯分校理论物理专业理学双硕士学位。他的研究兴趣集中在统计机器学习和优化，以及一些理论计算机科学，他的研究成果已发表在 NIPS OPT-2017 研讨会和 *INFORMS Journal on Optimization* (机器学习特刊) 上。

**S. S. 艾扬格(S. S. Iyengar)博士** 是迈阿密佛罗里达国际大学杰出的大学教授、杰出的 Ryder 教授和计算与信息科学学院院长，是分布式传感器网络/传感器融合、机器人技术计算领域以及高性能计算领域的先驱。

他曾是印度科学理工学院(IISC)班加罗尔分校的 Satish Dhawan 教授，以及泰米尔纳德邦 Kalpakkam IGCAR 的 Homi Bhabha 教授，还曾是巴黎大学、清华大学、KAIST 等的客座教授。

他发表研究论文 600 余篇，在 MIT 出版社、John Wiley & Sons 出版社、Prentice Hall 出版社、CRC 出版社、Springer Verlag 出版社等出版 22 部专著，这些出版物已在世界各地的重点大学使用。他拥有许多专利，其中一些专利还出现在得克萨斯州达拉斯市举办的世界最佳技术论坛上。他的研究出版物涉及高效算法、并行计算、传感器网络和机器人的设计与分析。在过去的 40 年里，他指导了 55 名博士生、100 名硕士生和许多本科生，这些学生现在遍布世界各地，有的是重点大学的教师，有的是国家实验室/工业领域的科学家或工程师。他的许多本科生仍在从事他的研究项目。最近，艾扬格博士获得了 Times Network 媒体集团评选的 2017 年度非居民印度人奖，这

是一个为全球印度领导人设立的著名奖项。

艾扬格博士是欧洲科学院成员，IEEE、ACM、AAAS、美国国家发明家科学院(NAI)、美国设计与工艺学会(SPDS)、美国工程师学会(FIE)、美国医学与生物工程学会(AIMBE)的高级或资深会员。由于对传感器融合算法和并行算法的贡献，他获得了班加罗尔印度科学研究所的杰出校友奖和IEEE 计算机协会技术成就奖。他还在喷气推进实验室获得了IBM 杰出教师奖和NASA 夏季奖学金。他是2010年得克萨斯州奥斯汀市跨学科学学习与高级研究学院的研究员。

他获得了各种国内和国际奖项，包括Times Network 媒体集团评选的2017年度非居民印度人奖、2013年美国国家发明家科学院院士奖、2013年伦敦上议院的NRI 圣雄甘地·普拉瓦西奖章，以及国际敏捷制造协会(ISAM)授予的终身成就奖，以表彰他在教学、研究和管理领域的杰出成就以及对印度理工学院(BHU)在工程和计算机科学领域做出的毕生贡献。2012年，他和Nulogix 荣获2012年佛罗里达创新-产业奖(i2i)。因在传感器网络、计算机视觉和图像处理领域的研究，他获得了厦门大学颁发的杰出研究奖。他与他的研究小组的里程碑式的贡献，包括在分布式传感器网络中开发用于监视和目标定位的网格覆盖与Brooks-Iyengar 融合算法。他获得了富布赖特杰出研究奖，以及2019年IEEE 智能和安全信息学研究领导奖；在第25届国际IEEE 高性能计算会议(2019年)上，因其对分布式传感器网络的贡献而获得终身成就奖，该奖由Infosys 的联合创始人Narayana Murthy 博士颁发；获得佛罗里达州青光眼装置创新技术工业创新奖、LSU Rainmaker 奖，以及杰出研究硕士奖。他还被授予荣誉理工科博士学位。他在世界上许多公司和大学的顾问委员会任职，还曾在许多国家科学委员会任职，如美国国立卫生研究院生物信息学国家医学图书馆、国家科学基金会评审小组、美国宇航局空间科学、国土安全部、海军安全办公室等。他对美国海军研究实验室的贡献是一项开拓性工作的核心，该项工作旨在为科学技术发展图像分析，并扩大美国海军研究实验室的目标。

他的研究成果可以在多家公司和多个国家实验室中看到，如雷神公司、Telcordia 公司、摩托罗拉公司、美国海军、DARPA 和其他美国机构。他在 DARPA 与 BBN、剑桥、马萨诸塞、MURI、PSU / ARL、杜克大学、威斯康星大学、加州大学洛杉矶分校(UCLA)、康奈尔大学和 LSU 的研究人员项目演示中做出了重要贡献。他也是 *International Journal of Distributed Sensor Networks* 的创刊编辑。他曾是多家期刊的编委会成员，也是多所大学的博士委员会成员，包括卡内基梅隆大学(CMU)、杜克大学和世界各地的许多其他大学。他目前是 *ACM Computing Surveys* 等期刊的编辑。

他还是 FIU 发现实验室的创始主任。他的研究成果被广泛引用。他的基础工作已经转化为独特的技术。在长达 40 年的职业生涯中，艾扬格博士以一种独特的方式致力于运用数学形态学来定量地理解计算过程，并将其应用于许多领域。

# 目 录

Mathematical Theories of Machine Learning-Theory and Applications

译者序  
序言  
致谢  
前言  
作者简介

## 第一部分 引言

第 1 章 绪论 .....	3
1.1 神经网络 .....	4
1.2 深度学习 .....	7
1.3 梯度下降法 .....	8
1.4 小结 .....	10
1.5 本书结构 .....	10
第 2 章 通用数学框架 .....	13
2.1 机器学习与计算统计学 .....	13
2.2 小结 .....	17
第 3 章 优化理论简述 .....	19
3.1 机器学习所需的优化理论 .....	20
3.2 在线算法：机器学习的顺序更新 .....	30
3.3 小结 .....	32

第 4 章 改进的 CoCoSSC 方法 .....	35
4.1 问题描述 .....	35
4.2 梯度加速下降法 .....	37
4.3 CoCoSSC 方法 .....	37
4.4 在线时变粘性网算法 .....	39
4.5 小结 .....	40
第 5 章 关键术语 .....	41
5.1 一些定义 .....	41
5.2 小结 .....	44
第 6 章 关于非凸规划几何的相关研究 .....	45
6.1 多元时间序列数据集 .....	48
6.2 粒子学习 .....	50
6.3 在气候变化中的应用 .....	50
6.4 小结 .....	51

## 第二部分 机器学习的数学框架：理论部分

第 7 章 收敛到最小值的梯度下降法：最优和自适应的步长规则 .....	55
7.1 引言 .....	55
7.2 符号与预备知识 .....	59
7.3 最大允许步长 .....	62
7.4 自适应步长规则 .....	65
7.5 定理 7.1 的证明 .....	66
7.6 定理 7.2 的证明 .....	68
7.7 辅助定理 .....	72
7.8 技术证明 .....	72
7.9 小结 .....	75

第 8 章 基于优化的守恒定律方法 .....	77
8.1 准备：直观的解析演示 .....	77
8.2 辛方法与算法 .....	79
8.3 局部高速收敛现象的渐近分析 .....	85
8.4 实验演示 .....	93
8.5 小结与展望 .....	102

### 第三部分 机器学习的数学框架：应用部分

第 9 章 含有噪声和缺失观测值的稀疏子空间聚类的样本 复杂度的改进 .....	107
9.1 CoCoSSC 算法的主要结果 .....	107
9.2 证明 .....	111
9.3 数值结果 .....	115
9.4 技术细节 .....	118
9.5 小结 .....	123
第 10 章 多元时间序列中稳定和分组因果关系的在线发现 .....	125
10.1 问题表述 .....	125
10.2 粘性网正则化 .....	126
10.3 在线推理 .....	130
10.4 实验验证 .....	136
10.5 小结与展望 .....	143
第 11 章 后记 .....	145
参考文献 .....	147

| 第一部分 |

Mathematical Theories of Machine Learning-Theory and Applications

# 引 言