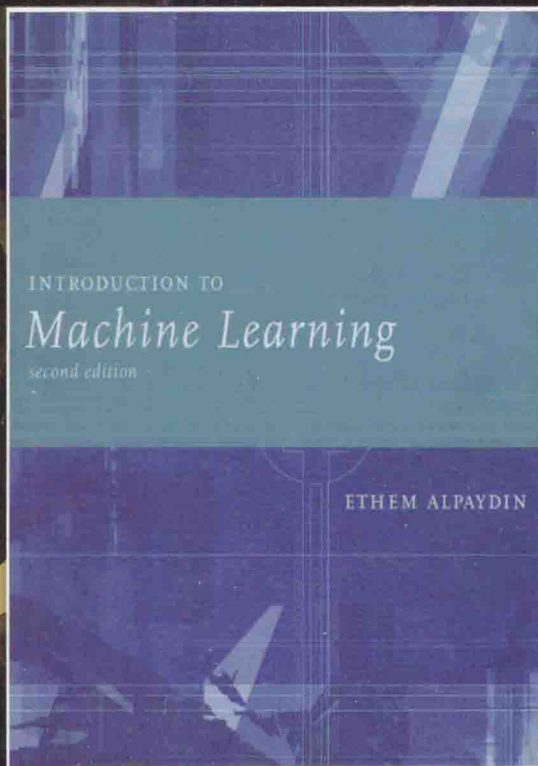


机器学习导论

(土耳其) Ethem Alpaydin 著 范明 咎红英 牛常勇 译

Introduction to Machine Learning
Second Edition



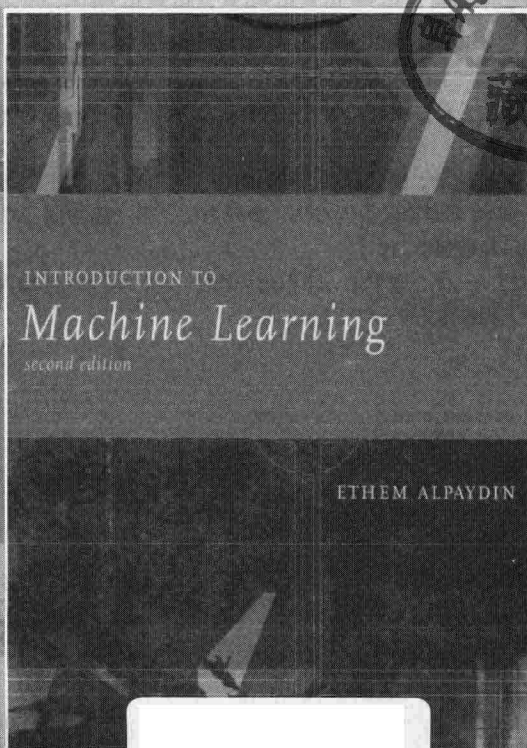
计 算 机 科 学 丛

原书第2版

机器学习导论

(土耳其) Ethem Alpaydin 著 范明 管红英 牛常勇 译

Introduction to Machine Learning
Second Edition



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

机器学习导论 (原书第 2 版) / (土耳其) 阿培丁 (Alpaydin, E.) 著; 范明等译. —北京: 机械工业出版社, 2014.3

(计算机科学丛书)

书名原文: Introduction to Machine Learning, Second Edition

ISBN 978-7-111-45377-2

I. 机… II. ①阿… ②范… III. 机器学习—研究 IV. TP181

中国版本图书馆 CIP 数据核字 (2013) 第 321864 号

本书版权登记号: 图字: 01-2013-1430

Ethem Alpaydin: Introduction to Machine Learning, Second Edition (ISBN 978-0-262-01243-0).

Original English language edition copyright © 2010 by Massachusetts Institute of Technology.

Simplified Chinese Translation Copyright © 2014 by China Machine Press.

Simplified Chinese translation rights arranged with MIT Press through Bardon-Chinese Media Agency.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission, in writing, from the publisher.

All rights reserved..

本书中文简体字版由 MIT Press 通过 Bardon-Chinese Media Agency 授权机械工业出版社在中华人民共和国境内独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书讨论了机器学习在统计学、模式识别、神经网络、人工智能、信号处理等不同领域的应用, 其中涵盖了监督学习、贝叶斯决策理论、参数方法、多元方法、多层感知器、局部模型、隐马尔可夫模型、分类算法评估和比较以及增强学习。

本书可供完成计算机程序设计、概率论、微积分和线性代数课程的高年级本科生和研究生使用, 也可供对机器学习感兴趣的工程技术人员参考。

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: 肖晓慧

印 刷: 藁城市京瑞印刷有限公司

版 次: 2014 年 4 月第 1 版第 1 次印刷

开 本: 185mm × 260mm 1/16

印 张: 22

书 号: ISBN 978-7-111-45377-2

定 价: 59.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

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

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：www.hzbook.com

电子邮件：hzsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章教育

华章科技图书出版中心

中文版序

Introduction to Machine Learning, Second Edition

机器学习是计算机科学发展最快的领域之一。每年，我们都看到新的应用，并且学习算法的理论也发展很快。

我十分高兴地看到我的书的第2版用中文出版，并就为翻译所做出的努力，感谢范明教授。在此之前，范教授翻译了本书第1版以及一些统计学习和数据挖掘的名著。

我希望我的书的中文读者发现本书是有益的，并且就像我乐于写它一样乐于阅读它。

Ethem Alpaydin

于伊斯坦布尔博阿齐奇大学

2013年12月

Preface of the Chinese Edition

Machine learning is one of the fastest developing fields in computer science. Every year, we are seeing new applications, and the theory of learning algorithms is also developing very fast.

It gives me great pleasure to see the second edition of my book printed in Chinese, and for the effort in doing the translation, I would like to thank Professor Fan who previously have translated the first edition as well as several other well-known texts on statistics and data mining.

I hope that the Chinese readers of my book will find it beneficial and enjoy reading it as much as I enjoyed writing it.

Ethem Alpaydin

Boğazici University, Istanbul

December 2013

自从有计算机以来，人们就希望计算机能够学习。然而，机器学习真正取得实质性进展，能够成功地解决一些实际问题，并最终成为一个学科分支还是近 20 余年的事。

对于许多问题，我们的前人和先行者已经知道如何求解。例如，欧几里得告诉我们可以用辗转相除法求两个整数的最大公约数；Dijkstra 告诉我们如何有效地求两点之间的最短路径；Hoare 向我们展示了怎样将杂乱无章的对象快速排序……对于这些问题，我们清楚地知道求解步骤。因此，让计算机求解这些问题只需要设计算法和数据结构、进行编程，而不需要让计算机学习。

还有一些问题，人们可以轻而易举地做好，但是却无法解释清楚我们是如何做的。例如，尽管桌子千差万别、用途各异，但是我们一眼就能看出某个物体是否是桌子；尽管不同的人的手写阿拉伯数字大小不一、笔画粗细不同，但是我们还是可以轻易识别一个数字是不是 8；尽管声音时大时小，有时可能还有点沙哑，但是我们还是可以不用费力气地听出熟人的声音。诸如此类的例子不胜枚举。对于这些问题，我们不知道求解步骤。因此，让计算机来做这些事就需要让计算机学习。

我们知道桌子不是木材和各种材料的随机堆砌，手写数字不是像素的随机分布，熟人的声音也不是各种声波的随机混合。现实世界总是有规律的。机器学习正是从已知实例中自动发现规律，建立对未知实例的预测模型；根据经验不断提高，不断改进预测性能。

这是一本全面论述机器学习这一主题的教科书，适合作为高等院校计算机相关专业高年级本科生和研究生机器学习入门课程的教材。该书涵盖了监督学习、贝叶斯决策理论、参数方法、多元方法、维度归约、聚类、非参数方法、决策树、线性判别式、多层感知器、局部模型、隐马尔可夫模型、分类算法评估和比较、组合多学习器以及增强学习。作者对来自统计学、模式识别、神经网络、人工智能、信号处理、控制和数据挖掘等不同领域的机器学习问题和学习方法进行了统一论述。

第 2 版从 16 章扩展到 19 章，除增加 3 章外，对许多章节都进行了改写和扩充，以便反映机器学习的新进展。尤其是，核方法、贝叶斯估计和图模型这三个在第 1 版只用几节介绍的主题都各自扩充成了一整章，进行了更深入的讨论。此外，第 2 版还更加全面地介绍了机器学习实验的设计与分析，这在同类书籍中是独具特色的。

第 2 版由范明翻译。咎红英（第 1~6 章）和牛常勇（第 15、17 和 18 章）参加了第 1 版的翻译。原书作者 Ethem Alpaydin 为第 2 版的中文版重新写了序。

译文中的错误和不当之处，敬请读者朋友指正。意见和建议请发往 mfan@zzu.edu.cn，我们不胜感激。

范 明

2013 冬于郑州大学

机器学习使用实例数据或过去的经验训练计算机，以优化性能标准。当人们不能直接编写计算机程序解决给定的问题，而是需要借助于实例数据或经验时，就需要学习。一种需要学习的情况是人们没有专门技术，或者不能解释他们的专门技术。以语音识别，即将声学语音信号转换成 ASCII 文本为例。看上去我们可以毫无困难地做这件事，但是我们却不能解释我们是如何做的。由于年龄、性别或口音的差异，不同的人读相同的词发音却不同。在机器学习中，这个问题的解决方法是从不同的人那里收集大量发音样本，并学习将它们映射到词。

另一种需要学习的情况是要解决的问题随时间变化或依赖于特定的环境。我们希望有一个能够自动适应环境的通用系统，而不是为每个特定的环境编写一个不同的程序。以计算机网络上的包传递为例。最大化服务质量的、从源地到目的地的路径随网络流量的改变而改变。学习路由程序能够通过监视网络流量自动调整到最佳路径。另一个例子是智能用户界面，它能够自动适应用户的生物特征，即用户的口音、笔迹、工作习惯等。

机器学习在各个领域都有许多成功的应用：已经有了识别语音和笔迹的商用系统。零售商分析他们过去的销售数据，了解顾客行为，以便改善顾客关系管理。金融机构分析过去的交易，以便预测顾客的信用风险。机器人学习优化它们的行为，以便使用最少的资源来完成任务。在生物信息学方面，使用计算机不仅可以分析海量数据，而且还可以提取知识。这些只是我们（即你和我）将在本书讨论的应用的一部分。我们只能想象一下可使用机器学习实现的未来应用：可以在不同的路况、不同的天气条件下自己行驶的汽车，可以实时翻译外语的电话，可以在新环境（例如另一个星球的表面）航行的自动化机器人。机器学习的确是一个令人激动的研究领域！

本书讨论的许多方法都源于各种领域：统计学、模式识别、神经网络、人工智能、信号处理、控制和数据挖掘。过去，这些不同领域的研究遵循不同的途径，侧重点也不同。本书旨在把它们组合在一起，给出问题的统一处理并提供它们的解。

本书是一本入门教材，用于高年级本科生和研究生的机器学习课程，以及在业界工作、对这些方法的应用感兴趣的工程技术人员。预备知识是计算机程序设计、概率论、微积分和线性代数方面的课程。本书的目标是充分解释所有的学习算法，使得从本书给出的方程到计算机程序只是一小步。为了使这一任务更容易完成，对于某些情况，我们给出了算法的伪代码。

适当选取一些章节，本书可用作一学期的课程。再额外讨论一些研究论文的话，本书也可以用作两学期的课程，这时每章后的参考文献将很有用。

本书网页为 <http://www.cmpe.boun.edu.tr/~ethem/i2ml/>，我将在那里提供一些与本书有关的信息，如勘误表。我真诚地欢迎你将反馈意见发到我的邮箱：alpaydin@boun.edu.tr。

我非常喜欢写这本书，希望你能喜欢读它。

获得好想法的途径是与有才干的人一起工作，与他们一起工作也是一种乐趣。Boğaziçi 大学计算机工程系是一个极好的工作场所，在我写这本书时，我的同事们为我提供了我所需的所有支持。我也要感谢我过去和现在的学生，在他们身上，我实际检验了现在写进这本书中的内容。

在写本书时，我得到了土耳其科学院青年科学家奖励计划的资助（EATÜBA-GEBİP/2001-1-1）。

我特别感谢 Michael Jordan。对于他多年来的支持和最近对本书的支持，我深表感谢。他针对本书大体组织和第 1 章所给出的建议在内容和形式上都大大改进了本书。Taner Bilgiç、Vladimir Cherkassky、Tom Dietterich、Fikret Gürgen、Olca Taner Yildiz 和 MIT 出版社的未留名审稿人也部分阅读了本书，并提供了非常宝贵的反馈。我希望他们在注意到我采纳了他们的建议但却没有特别致谢时，能够体会到我的感激之情。当然，书中的错误和不足应当由我个人负责。

我的父母信任我，我感谢他们永恒的爱和支持。无论我何时需要，Sema Oktuğ 总在身边，我将永远感激她的友谊。我还要感谢 Hakan Ünlü，在过去的几年中，我们无数次讨论了与生活、宇宙和万事万物相关的众多主题。

本书使用 Chris Manning 准备的 LATEX 宏排版，对此我很感谢他。我要感谢 MIT 出版社的编辑们，以及 Bob Prior、Valerie Geary、Kathleen Caruso、Sharon Deacon Warne、Erica Schultz 和 Emily Gutheinz，感谢他们在本书完成期间的不断支持和帮助。

自从第1版2004年问世以来，机器学习已经取得重要进展。首先，在应用领域迅速扩展。现在，互联网相关的技术，如搜索引擎、推荐系统、垃圾邮件过滤和入侵检测系统都在常规地使用机器学习。在生物信息学和计算生物学领域，由数据学习的方法被越来越广泛地使用。在自然语言处理应用（例如，机器翻译）中，我们看到越来越快地从编程的专家系统过渡到从实例文本的大型语料库中自动地学习。在机器人、医疗诊断、语音和图像识别、生物测定学、财经等领域，我们看到了本书讨论的机器学习方法越来越多的应用。这些应用有时是在模式识别的名义下，有时乔装为数据挖掘或披着某种其他外衣。

其次，理论上取得了重要进步。尤其是，核函数的思想和使用核函数的核机器使得我们可以更好地表示问题，并且与使用梯度下降训练的、具有S形隐藏单元的多层感知器相比，核机器的凸优化工作前进了一大步。贝叶斯方法通过选定适当的先验分布将专家的知识添加到数据隐含的知识上。图模型允许使用相互关联的节点的网络表示变量之间的依赖，并且有效地推断算法使得我们可以查询该网络。因此，有必要将核方法、贝叶斯估计和图模型这三个在第1版只用几节介绍的主题用更长的篇幅处理，扩充为新的三章。

该领域另一个极其重要的事情是认识到需要更好地设计机器学习实验。从使用单个检验集到交叉验证，再到配对 t 检验，我们已经走了很长的路。这就是为什么我要在第2版重写统计检验这一章，使之包含机器学习实验的设计与分析。要点是，检验不應該是在所有运行都完成之后再做的单独步骤（尽管介绍机器学习实验的新一章被安排在本书最后），应该提前设计实验的整个过程、定义相关因素、确定合适的实验过程。这时并且只有这时才做实验并分析实验结果。

长期以来，人们尤其是科学界的年长者相信，要想使机器像我们一样有智能，即要想使人工智能成为现实，无论就一般而言还是特殊地就计算机科学而言，我们当前的知识都是不够的。大多数人都认为，我们需要新的技术、新型材料、新型计算装置或新的程序设计技术，并且即便如此，我们也只能以有限的方式“模拟”人类智能的某些方面，而不可能完全实现人类的智能。

我相信我们很快将证明他们是错误的。我们最早在国际象棋中看到了这种证明，并且现在我们正在各个领域看到这种证明。有了足够的内存和计算能力，我们可以使用相对简单的算法来完成任务；这里的技巧是学习，或者从实例数据中学习，或者使用增强学习通过试错学习。看来，（例如，机器翻译）使用监督学习并且更多的是非监督学习算法很快将会成为可能。对于许多其他领域，例如使用增强学习的机器人无人导航也是如此。我相信在人工智能的许多领域这种态势都将继续，而关键是学习。只要我们为机器提供足够的数据（不必是监督的）和计算能力，如果机器可以自己学习，我们则不需要提出新的算法。

感谢世界各地使用第 1 版的所有教师和学生。感谢给我发送评论和勘误或以任何方式提供反馈的人。请继续给我发这样的邮件。我的 email 地址是 alpaydin@boun.edu.tr。

第 2 版还在网站上提供了更多的支持。本书的网站地址是 <http://www.cmpe.boun.edu.tr/~ethem/i2ml>。

感谢过去和现在跟我做论文的学生 Mehmet Gönen、Esmâ Kihç、Murat Semerci、M. Aydın Ulaş 和 Olcay Taner Yıldız，感谢在过去的几年选修 CmpE 544、CmpE 545、CmpE 591 和 CmpE 58E 课程的学生。检查你的知识的最佳方法是讲述它。

第 2 版再次与 MIT 出版社合作是一件令人愉快的事。感谢 Bob Prior、Ada Brunstein、Erin K. Shoudy、Kathleen Caruso 和 Marcy Ross 的帮助和支持。

符号表

Introduction to Machine Learning, Second Edition

x	标量值
\mathbf{x}	向量
X	矩阵
\mathbf{x}^\top	转置
X^{-1}	逆矩阵
X	随机变量
$P(X)$	概率质量函数, X 是离散的
$p(X)$	概率密度函数, X 是连续的
$P(X Y)$	给定 Y , X 的条件概率
$E[X]$	随机变量 X 的期望值
$\text{Var}(X)$	X 的方差
$\text{Cov}(X, Y)$	X 和 Y 的协方差
$\text{Corr}(X, Y)$	X 和 Y 的相关性
μ	均值
σ^2	方差
Σ	协方差矩阵
m	均值的估计
s^2	方差的估计
S	协方差矩阵的估计
$\mathcal{N}(\mu, \sigma^2)$	一元正态分布, 均值为 μ , 方差为 σ^2
\mathcal{Z}	单位正态分布: $\mathcal{N}(0, 1)$
$\mathcal{N}_d(\mu, \Sigma)$	d -变量正态分布, 均值向量为 μ , 协方差矩阵为 Σ
x	输入
d	输入数 (输入的维度)
y	输出
r	要求的输出
K	输出数(类)
N	训练实例数
z	隐藏的值, 内蕴维, 潜在因子
k	隐藏维数, 潜在因子数
C_i	类 i
X	训练样本

$\{x^t\}_{t=1}^N$	x 的集合, 上标 t 遍取 1 到 N
$\{x^t, r^t\}_t$	上标为 t 的输入和期望输出的有序对的集合
$g(x \theta)$	x 的函数, 其定义依赖于参数集 θ
$\arg \max_{\theta} g(x \theta)$	参数 θ , g 关于它取最大值
$\arg \min_{\theta} g(x \theta)$	参数 θ , g 关于它取最小值
$E(\theta \mathcal{X})$	样本 \mathcal{X} 上具有参数 θ 的误差函数
$l(\theta \mathcal{X})$	样本 \mathcal{X} 上具有参数 θ 的似然函数
$\mathcal{L}(\theta \mathcal{X})$	样本 \mathcal{X} 上具有参数 θ 的对数似然函数
$1(c)$	如果 c 为真, 则值为 1, 否则为 0
$\#\{c\}$	c 为真的元素数目
δ_{ij}	Kronecker δ : 如果 $i=j$, 取 1, 否则取 0

目 录

Introduction to Machine Learning, Second Edition

出版者的话	
中文版序	
译者序	
前言	
致谢	
关于第2版	
符号表	
第1章 绪论 1	
1.1 什么是机器学习..... 1	
1.2 机器学习的应用实例 3	
1.2.1 学习关联性..... 3	
1.2.2 分类 3	
1.2.3 回归 6	
1.2.4 非监督学习..... 7	
1.2.5 增强学习 8	
1.3 注释 8	
1.4 相关资源..... 10	
1.5 习题 11	
1.6 参考文献..... 12	
第2章 监督学习 13	
2.1 由实例学习类 13	
2.2 VC维 15	
2.3 概率逼近正确学习 16	
2.4 噪声 17	
2.5 学习多类..... 18	
2.6 回归 19	
2.7 模型选择与泛化 21	
2.8 监督机器学习算法的维 23	
2.9 注释 24	
2.10 习题 25	
2.11 参考文献 25	
第3章 贝叶斯决策定理 27	
3.1 引言 27	
3.2 分类 28	
3.3 损失与风险 29	
3.4 判别式函数 31	
3.5 效用理论..... 31	
3.6 关联规则..... 32	
3.7 注释 33	
3.8 习题 33	
3.9 参考文献..... 34	
第4章 参数方法 35	
4.1 引言 35	
4.2 最大似然估计 35	
4.2.1 伯努利密度 36	
4.2.2 多项密度 36	
4.2.3 高斯(正态)密度 37	
4.3 评价估计: 偏倚和方差 37	
4.4 贝叶斯估计 38	
4.5 参数分类..... 40	
4.6 回归 43	
4.7 调整模型的复杂度: 偏倚/方差 两难选择..... 45	
4.8 模型选择过程 47	
4.9 注释 50	
4.10 习题 50	
4.11 参考文献 51	
第5章 多元方法 52	
5.1 多元数据..... 52	
5.2 参数估计..... 52	
5.3 缺失值估计 53	
5.4 多元正态分布 54	
5.5 多元分类..... 56	
5.6 调整复杂度 59	
5.7 离散特征..... 61	
5.8 多元回归..... 62	
5.9 注释 63	
5.10 习题 63	
5.11 参考文献 64	

第 6 章 维度归约	65	8.8 注释	110
6.1 引言	65	8.9 习题	111
6.2 子集选择	65	8.10 参考文献	112
6.3 主成分分析	67	第 9 章 决策树	113
6.4 因子分析	71	9.1 引言	113
6.5 多维定标	75	9.2 单变量树	114
6.6 线性判别分析	77	9.2.1 分类树	114
6.7 等距特征映射	80	9.2.2 回归树	118
6.8 局部线性嵌入	81	9.3 剪枝	119
6.9 注释	83	9.4 由决策树提取规则	120
6.10 习题	84	9.5 由数据学习规则	121
6.11 参考文献	85	9.6 多变量树	124
第 7 章 聚类	86	9.7 注释	125
7.1 引言	86	9.8 习题	126
7.2 混合密度	86	9.9 参考文献	127
7.3 k -均值聚类	87	第 10 章 线性判别式	129
7.4 期望最大化算法	90	10.1 引言	129
7.5 潜在变量混合模型	93	10.2 推广线性模型	130
7.6 聚类后的监督学习	94	10.3 线性判别式的几何意义	131
7.7 层次聚类	95	10.3.1 两类问题	131
7.8 选择簇个数	96	10.3.2 多类问题	132
7.9 注释	96	10.4 逐对分离	132
7.10 习题	97	10.5 参数判别式的进一步讨论	133
7.11 参考文献	97	10.6 梯度下降	135
第 8 章 非参数方法	99	10.7 逻辑斯谛判别式	135
8.1 引言	99	10.7.1 两类问题	135
8.2 非参数密度估计	99	10.7.2 多类问题	138
8.2.1 直方图估计	100	10.8 回归判别式	141
8.2.2 核估计	101	10.9 注释	142
8.2.3 k -最近邻估计	102	10.10 习题	143
8.3 到多元数据的推广	103	10.11 参考文献	143
8.4 非参数分类	104	第 11 章 多层感知器	144
8.5 精简的最近邻	105	11.1 引言	144
8.6 非参数回归: 光滑模型	106	11.1.1 理解人脑	144
8.6.1 移动均值光滑	106	11.1.2 神经网络作为并行	
8.6.2 核光滑	108	处理的典范	145
8.6.3 移动线光滑	108	11.2 感知器	146
8.7 如何选择光滑参数	109	11.3 训练感知器	148

11.4	学习布尔函数	150	12.11	习题	190
11.5	多层感知器	151	12.12	参考文献	190
11.6	作为普适近似的 MLP	153	第 13 章	核机器	192
11.7	后向传播算法	154	13.1	引言	192
11.7.1	非线性回归	154	13.2	最佳分离超平面	193
11.7.2	两类判别式	157	13.3	不可分情况: 软边缘超平面	195
11.7.3	多类判别式	158	13.4	ν -SVM	197
11.7.4	多个隐藏层	158	13.5	核技巧	198
11.8	训练过程	158	13.6	向量核	199
11.8.1	改善收敛性	158	13.7	定义核	200
11.8.2	过分训练	159	13.8	多核学习	201
11.8.3	构造网络	161	13.9	多类核机器	202
11.8.4	线索	162	13.10	用于回归的核机器	203
11.9	调整网络规模	163	13.11	一类核机器	206
11.10	学习的贝叶斯观点	164	13.12	核维度归约	208
11.11	维度归约	165	13.13	注释	209
11.12	学习时间	167	13.14	习题	209
11.12.1	时间延迟神经网络	167	13.15	参考文献	210
11.12.2	递归网络	168	第 14 章	贝叶斯估计	212
11.13	注释	169	14.1	引言	212
11.14	习题	170	14.2	分布参数的估计	213
11.15	参考文献	170	14.2.1	离散变量	213
第 12 章	局部模型	173	14.2.2	连续变量	215
12.1	引言	173	14.3	函数参数的贝叶斯估计	216
12.2	竞争学习	173	14.3.1	回归	216
12.2.1	在线 k -均值	173	14.3.2	基函数或核函数的使用	218
12.2.2	自适应共鸣理论	176	14.3.3	贝叶斯分类	219
12.2.3	自组织映射	177	14.4	高斯过程	221
12.3	径向基函数	178	14.5	注释	223
12.4	结合基于规则的知识	182	14.6	习题	224
12.5	规范化基函数	182	14.7	参考文献	224
12.6	竞争的基函数	184	第 15 章	隐马尔可夫模型	225
12.7	学习向量量化	186	15.1	引言	225
12.8	混合专家模型	186	15.2	离散马尔可夫过程	225
12.8.1	协同专家模型	188	15.3	隐马尔可夫模型	227
12.8.2	竞争专家模型	188	15.4	HMM 的三个基本问题	229
12.9	层次混合专家模型	189	15.5	估值问题	229
12.10	注释	189	15.6	寻找状态序列	231

15.7	学习模型参数	233	17.12	注释	270
15.8	连续观测	235	17.13	习题	271
15.9	带输入的 HMM	236	17.14	参考文献	272
15.10	HMM 中的模型选择	236	第 18 章	增强学习	275
15.11	注释	237	18.1	引言	275
15.12	习题	238	18.2	单状态情况: K 臂赌博机问题	276
15.13	参考文献	239	18.3	增强学习基础	277
第 16 章	图方法	240	18.4	基于模型的学习	278
16.1	引言	240	18.4.1	价值迭代	279
16.2	条件独立的典型情况	241	18.4.2	策略迭代	279
16.3	图模型实例	245	18.5	时间差分学习	280
16.3.1	朴素贝叶斯分类	245	18.5.1	探索策略	280
16.3.2	隐马尔可夫模型	246	18.5.2	确定性奖励和动作	280
16.3.3	线性回归	248	18.5.3	非确定性奖励和动作	282
16.4	d -分离	248	18.5.4	资格迹	283
16.5	信念传播	249	18.6	推广	285
16.5.1	链	249	18.7	部分可观测状态	286
16.5.2	树	250	18.7.1	场景	286
16.5.3	多树	251	18.7.2	例子: 老虎问题	287
16.5.4	结树	252	18.8	注释	290
16.6	无向图: 马尔可夫随机场	253	18.9	习题	291
16.7	学习图模型的结构	254	18.10	参考文献	292
16.8	影响图	255	第 19 章	机器学习实验的 设计与分析	294
16.9	注释	255	19.1	引言	294
16.10	习题	256	19.2	因素、响应和实验策略	296
16.11	参考文献	256	19.3	响应面设计	297
第 17 章	组合多学习器	258	19.4	随机化、重复和阻止	298
17.1	基本原理	258	19.5	机器学习实验指南	298
17.2	产生有差异的学习器	258	19.6	交叉验证和再抽样方法	300
17.3	模型组合方案	260	19.6.1	K -折交叉验证	300
17.4	投票法	261	19.6.2	5×2 交叉验证	301
17.5	纠错输出码	263	19.6.3	自助法	302
17.6	装袋	265	19.7	度量分类器的性能	302
17.7	提升	265	19.8	区间估计	304
17.8	重温混合专家模型	267	19.9	假设检验	307
17.9	层叠泛化	268	19.10	评估分类算法的性能	308
17.10	调整系综	268	19.10.1	二项检验	308
17.11	级联	269			

绪论

1.1 什么是机器学习

为了在计算机上解决问题，我们需要算法。算法是指令的序列，它把输入变换成输出。例如，我们可以为排序设计一个算法，输入是数的集合，而输出是它们的有序列表。对于相同的任务可能存在不同的算法，而我们感兴趣的是如何找到需要的指令或内存最少，或者二者都最少的最有效算法。

然而，对于某些任务，我们没有算法；例如，我们没有将垃圾邮件与正常邮件分开的算法。我们知道输入是邮件文档，最简单的情况是一份字符文件。我还知道输出应该是指出消息是否为垃圾邮件的“是”或“否”，但是我们不知道如何把这种输入变换成输出。所谓的垃圾邮件随时间而变，因人而异。

我们缺乏的是知识，作为补偿我们有数据。我们可以很容易地编辑数以千计的实例消息，其中一些我们知道是垃圾邮件，而我们要做到的是希望从中“学习”垃圾邮件的结构。换言之，我们希望计算机(机器)自动地为这一任务提取算法。不需要学习如何将数排序，因为我们已经有这样的算法；但是，对于许多应用而言，我们确实没有算法，而是有实例数据。

随着计算机技术的发展，我们现在已经拥有存储和处理海量数据以及通过计算机网络从远程站点访问数据的能力。目前大多数的数据存取设备都是数字设备，记录的数据也很可靠。以一家连锁超市为例，它拥有遍布全国各地的数百家分店，并且在为数百万顾客提供数千种商品的零售服务。销售点的终端设备记录每笔交易的详细资料，包括日期、顾客识别码、购买商品和数量、消费总额等。这是典型的每日几个 G 字节的数据。连锁超市希望能够预测某种产品可能的顾客。对于这一任务，算法同样并非是显然的；它随时间而变，因地域而异。只有分析这些数据，并且将它转换为可以利用的信息时，这些存储的数据才能变得有用，例如做预测。

我们并不确切地知道哪些人倾向于购买这种口味的冰淇淋，或者这位作家的下一本书是什么，也不知道谁喜欢看这部新电影、访问这座城市，或点击这一链接。我们不能确切地知道哪些人比较倾向于购买哪种特定的商品，也不知道应该向喜欢读海明威作品的人推荐哪位作者。如果我们知道，我们就不需要任何数据分析；我们只管供货并记录下编码就可以了。但是，正因为我们不知道，所以才只能收集数据，并期望从数据中提取这些问题或相似问题的答案。

我们确信存在某种过程，可以解释我们所观测到的数据。尽管我们不清楚数据产生过程(例如顾客行为)的细节，但是，我们知道数据产生不是完全随机的。人们并不是去超市随机购买商品。当人们买啤酒时，也会买薯片；夏天买冰淇淋，而冬天则为 Glühwein[⊖]买香

⊖ Glühwein 是一种温热、有点甜味、加香料的葡萄酒。圣诞节期间，在欧洲很受欢迎。——译者注