HZ BOOKS
华章教育

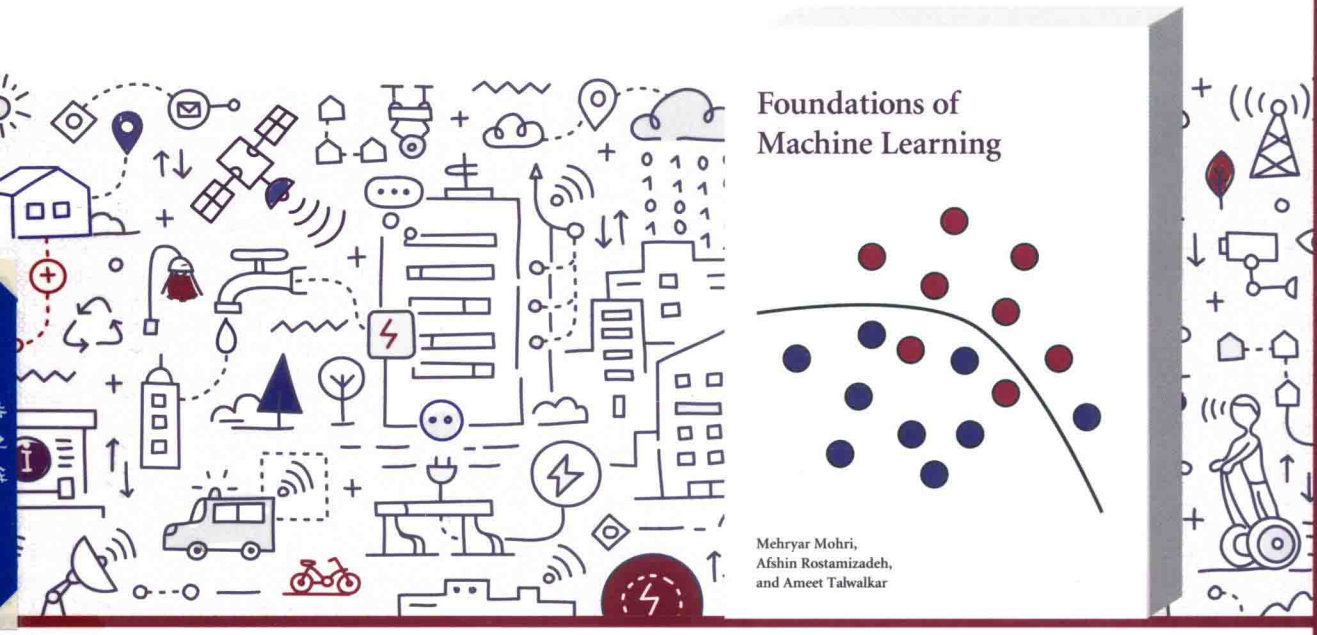
智能科学与技术丛书

Foundations of Machine Learning

机器学习基础

[美] 梅尔亚·莫里 (Mehryar Mohri) 纽约大学 / 谷歌研究院
阿夫欣·罗斯塔米扎达尔 (Afshin Rostamizadeh) 谷歌研究院
阿米特·塔尔沃卡尔 (Ameet Talwalkar) 卡内基·梅隆大学
◎ 著

张文生 中国科学院自动化研究所 ◎ 等译



机械工业出版社
China Machine Press

智能科学与技术丛书

Foundations of Machine Learning

机器学习基础

[美] 梅尔亚·莫里（Mehryar Mohri）
纽约大学 / 谷歌研究院

阿夫欣·罗斯塔米扎达尔（Afshin Rostamizadeh）
谷歌研究院

阿米特·塔尔沃卡尔（Ameet Talwalkar）
卡内基·梅隆大学

◎ 著

张文生
中国科学院自动化研究所

◎ 等译

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

机器学习基础 /（美）梅尔亚·莫里（Mehryar Mohri）等著；张文生等译 . —北京：机械工业出版社，2019.4
（智能科学与技术丛书）
书名原文：Foundations of Machine Learning

ISBN 978-7-111-62218-5

I. 机… II. ①梅… ②张… III. 机器学习 IV. TP181

中国版本图书馆 CIP 数据核字（2019）第 043636 号

本书版权登记号：图字 01-2013-6573

本书从概率近似正确（PAC）理论出发探讨机器学习的基础理论与典型算法，包括 PAC 学习框架、VC- 维、支持向量机、核方法、在线学习、多分类、排序、回归、降维、强化学习等丰富的内容。此外，附录部分简要回顾了与机器学习密切相关的概率论、凸优化、矩阵以及范数等必要的预备知识。

本书重在介绍典型算法的理论支撑并指出算法在实际应用中的关键点，注重理论细节与证明过程，可作为高等院校机器学习、统计学等课程的教材，或作为相关领域研究人员的参考读物。

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：朱秀英　　责任校对：殷　虹
印　　刷：中国电影出版社印刷厂　　版　　次：2019 年 5 月第 1 版第 1 次印刷
开　　本：185mm × 260mm　1/16　　印　　张：18.75
书　　号：ISBN 978-7-111-62218-5　　定　　价：99.00 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　88378833　　投稿热线：（010）88379604
购书热线：（010）68326294　　读者信箱：hzjsj@hzbook.com

纽约大学 Mehryar Mohri 教授是机器学习界的泰斗级人物，他与他的学生 Afshin Rostamizadeh 以及 Ameet Talwalkar 合著的《Foundations of Machine Learning》是机器学习领域一部具有里程碑意义的著作。包括哥伦比亚大学、北京大学在内的多所国内外顶级院校均有以该书为基础开设的研究生课程。

机器学习是人工智能研究领域中最活跃的分支之一，为信息科学领域解决实际学习问题提供了理论支撑与应用算法。机器学习又是一个多学科的交叉领域，涉及统计学、信息论、优化、博弈论、形式语言和自动机、应用心理学、生物学和神经生理学等。这种学科交叉融合带来的良性互动，无疑促进了包括机器学习在内的诸学科的发展与繁荣。

本书内容丰富，视野宽阔，深入浅出地介绍了目前机器学习的重要理论和关键算法。不同于常规的机器学习算法入门读物，本书试图从更高的视点和更深的层次来解读机器学习的理论基础，引入了对指导理论研究和实际应用都至关重要的概率近似正确(Probability Approximately Correct，PAC)学习理论。该理论旨在回答由机器学习得到的结果到底有多高的可信度与推广能力，从某种意义上来说，只有理解了这部分内容，才能对机器学习何时能学习以及为何能学习成功有更加深刻的理解。PAC 理论涉及的数学基础较多，而国内关于 PAC 的参考资料非常少，我们人工智能与机器学习研究团队为此进行了多方论证并多次召开专题讨论会。此外，本书还巧妙地从间隔(margin)角度衔接各个章节，对机器学习中的诸多方面进行了完美的统一。

本书主要面向人工智能、机器学习、模式识别、数据挖掘、计算机应用、生物信息学、数学和统计学等领域的研究生和相关领域的科技人员。出版中译本的目的是希望能为国内从事相关研究的广大学者和研究生提供一本全面、系统、权威的教科书和参考书。如果能做到这一点，译者将感到十分欣慰。

必须说明的是，本书的翻译是中国科学院自动化研究所人工智能与机器学习研究团队集体努力的结果，团队成员杨雪冰、孙正雅、郭肇禄、张志忠、唐永强、何泽文、张似衡、牛景昊、任泽林、李明达、张晨阳、吴雅婧、黄妍、杨萌林、李定、尹彦婷等参与了本书的翻译工作，杨雪冰老师参与了全书的审校与修正，他们付出了艰辛劳动，在此我深表感谢。感谢机械工业出版社华章公司编辑的大力协助，倘若没有他们的热情支持，本书的中译版难以如此迅速地与大家见面。另外，本书的翻译得到了国家自然科学基金委重点项目和面上项目(U1636220、61472423、61432008 等)的资助，特此感谢。

在翻译过程中，我们力求准确地反映原著内容，同时保留原著的风格。对于英文原版中的一些公式及表述错误，我们在翻译的过程中结合作者的课程讲稿进行了核校，并以译者注的形式指出和修正了部分错误。但由于译者水平有限，书中难免有不妥之处，恳请读者批评指正。

最后，谨把本书的中译版献给我的导师王珏研究员！王珏老师生前对机器学习理论、算法和应用非常关注，对机器学习中很多基础问题有着独到而深刻的理解，他启发并引领了我们研究团队对机器学习理论和算法的研究工作，使我们终身受益。

中国科学院自动化研究所

张文生

2019 年 3 月于北京

本书是关于机器学习的概述，适合作为该领域学生和研究人员的教科书。书中涵盖了机器学习领域的基本内容，并且提供了讨论及检验算法合理性所必需的理论基础和概念工具。不仅如此，本书还描述了应用相关算法时需要考虑的若干关键问题。

本书旨在介绍新的机器学习理论和概念，并且对于相对前沿的结果给出了简要的证明。总体而言，我们尽可能在证明的过程中选择简洁的方式。尽管如此，我们会讨论机器学习中出现的一些重要且复杂的主题，指出若干开放的研究问题。对于那些常常与其他主题合并或者未引起足够关注的主题，在本书中将单独成章以着重讨论，例如多分类、排序和回归。

尽管本书覆盖了机器学习中很多重要的主题，但是出于论述简洁的目的且因目前缺乏针对一些方法的坚实的理论保证，图模型和神经网络两个重要主题未能覆盖。

本书主要面向机器学习、统计和其他相关领域的学生和研究人员，适合作为研究生和高年级本科生课程的教科书，或者学术研讨会的参考文本。本书前三章为后续内容奠定理论基础，第 5 章亦引入了一些概念来完善理论，并被后面章节广泛使用，而其余各章大多自成体系。每章最后给出了一套练习题，并单独给出完整的解答。[⊖]

我们假定本书的读者熟悉线性代数、概率和算法分析的基本概念。但是，为了进一步辅助学习，我们在附录中简要回顾了线性代数和概率的相关知识，给出了凸优化的简介，并且提供了用于证明集中界的大量有用的工具。

据我们所知，没有一本教科书可以涵盖本书所介绍的全部内容。我们会要求每届机器学习专业的学生对学习本书的体会进行反馈。尽管针对不同的专业领域有一些很不错的机器学习参考书，但是这些书并不涉及对其他基本内容的一般性讨论。比如，关于核方法的书并不涉及对 boosting、排序、强化学习、学习自动机或者在线学习等主题的讨论。当然也存在更为一般的机器学习方面的书，与之截然不同的是，本书关注理论基础并重视证明。

书中所介绍的大部分材料来自机器学习研究生课程(机器学习基础)，该课程由本书第一作者在过去 7 年中在纽约大学库兰特数学科学研究所讲授。本书极大地受益于该课程的学生以及我们的朋友、同事和相关研究人员所提出的宝贵意见和建议，在此深表感激。

特别感谢 Corinna Cortes 和 Yishay Mansour 对于本书内容的设计和组织提出的许多

⊖ 关于本书的习题解答及其他教辅资源，请访问作者主页 cs. nyu. edu/faculty/mohri 查看和下载。——编辑注

重要建议，包括大量详细的注释。我们充分考虑了他们的建议，这对于改进全书帮助很大。此外，还要感谢 Yishay Mansour 用本书的最初版本进行教学，并向我们积极反馈。

我们还要感谢来自学术界和企业界研究实验室的同事和朋友所给予的讨论、建议和贡献，他们是：Cyril Allauzen、Stephen Boyd、Aldo Corbisiero、Spencer Greenberg、Lisa Hellerstein、Sanjiv Kumar、Ryan McDonald、Andres Muñoz Medina、Tyler Neylon、Peter Norvig、Fernando Pereira、Maria Pershina、Ashish Rastogi、Michael Riley、Umar Syed、Csaba Szepesvári、Eugene Weinstein 和 Jason Weston。

最后，我们还要感谢 MIT 出版社对本书所给予的帮助和支持。

⊖ 参考文献为网络资源，请访问华章网站 www.hzbook.com 下载。——编辑注

·第 1 章·

引　言

机器学习广义上可被定义为基于经验提升性能或者进行精准预测的计算方法。这里，**经验**(experience)指的是学习器可利用的过去的信息，这些信息通常以收集和分析的电子数据的形式存在。这样的数据表现为数字化的、带人工标注的训练集，或者表现为与环境交互产生的各类信息。无论在何种情形下，数据的质量和规模对于学习器能否预测成功都至关重要。

机器学习关注如何设计高效和准确的预测**算法**(algorithm)。与计算机科学其他领域类似，衡量算法质量的重要指标是时间和空间复杂度。但是，在机器学习中，我们另外需要**样本复杂度**(sample complexity)的概念来评估算法学习概念类所需的样本规模。更为一般地，算法的理论学习保证取决于所考虑概念类的复杂度和训练样本的规模。

由于机器学习算法[⊖]的成功取决于所采用的数据，因此机器学习本质上与数据分析和统计相关。更一般地，机器学习技术是一类将计算机科学中的基本概念与统计、概率和优化方面的思想相结合的数据驱动方法。

1.1　应用与问题

学习算法已经被成功部署于各种应用中，包括：

- 文本或文档分类，例如，垃圾邮件检测；
- 自然语言处理，例如，词法分析、词性标注、统计句法分析和命名实体识别；
- 语音识别、语音合成、说话人确认；
- 光学字符识别(OCR)；

⊖ 在本书讨论范畴内，如不加特殊说明，下文中机器学习算法、机器学习技术、机器学习模型以及机器学习保证等概念大多数情况下会简称为学习算法、学习技术、学习模型以及学习保证等。——译者注

1

- 计算生物学应用，例如，蛋白质功能、结构预测；
- 计算机视觉任务，例如，图像识别、人脸检测；
- 欺诈检测(信用卡、电话)和网络入侵；
- 游戏，例如，国际象棋和西洋双陆棋；
- 无人驾驶车辆控制(机器人、导航)；
- 医疗诊断；
- 推荐系统、搜索引擎、信息抽取系统。

上述列表并不全面，学习算法每天都在被用于新的应用中。而且，这些应用对应于各种学习问题，一些主要类型如下：

- **分类**(classification)：为每个事项指定类别。例如，文本分类为事项指定类别，诸如政治、商业、运动或者天气这样的类别；图像分类为事项指定类别，诸如风景、肖像或者动物这样的类别。这些任务中的类别个数通常相对较少，但是在一些困难的任务中类别个数可能很大，甚至是无限的，像在文字识别、文本分类或者语音识别中。
- **回归**(regression)：预测每个事项的实值。回归的例子包括预测股票价格或者经济变量的变化。在该问题中，错误预测的惩罚取决于真实值和预测值之间的差异大小，这与分类问题有所不同，分类问题中不同类别之间通常没有距离的概念。
- **排序**(ranking)：根据某种准则将事项进行排序。网页搜索是典型的排序例子，比如返回与搜索查询相关的网页。许多其他相似的排序问题出现在信息抽取或者自然语言处理系统的设计中。
- **聚类**(clustering)：将事项划分为同质区域。聚类通常用来分析大数据集合。例如，在社交网络分析中，聚类算法试图从大规模人群中识别出“社区”。
- **降维**(dimensionality reduction)或者**流形学习**(manifold learning)：将事项的原始表示转化为这些事项的低维表示，同时保持原始表示的若干性质。一个常见的例子就是在计算机视觉任务中预处理数字图像。

机器学习的实际目标主要在于精确预测未见事项和设计高效稳健的算法来产生这些预

2

测，甚至对大规模问题仍有不错的适用性。为此，大量的算法和理论问题应运而生。基本的问题包括：哪些概念类是可以被真正学习到的，以及什么条件下可以学习到这些概念？从计算的角度，这些概念被学习到的效果或程度如何？

1.2　定义与术语

我们将用垃圾邮件检测这个典型问题作为实例，借以说明一些基本的定义，并且描述在实际中如何使用和评估机器学习算法。垃圾邮件检测是通过学习将电子邮件信息自动分类为垃圾邮件或非垃圾邮件这两个类别。

- **样本**(example)：用于学习或评估的数据事项或实例。在垃圾邮件检测问题中，样本对应于我们用来学习和测试的电子邮件信息集合。
- **特征**(feature)：与样本关联的属性集合，通常表示为向量。在电子邮件消息情形下，相关的特征可能包括消息的长度、发件人的名字、标题的不同特性、消息正文包含的关键词，等等。
- **标签**(label)：分配给样本的数值或者类别。在分类问题中，样本被归入特定的类别，例如，在上述二分类问题中的垃圾和非垃圾类别；在回归问题中，事项被赋予实值标签。
- **训练样本**(training sample)：用于训练学习算法的样本。在垃圾邮件问题中，训练样本由电子邮件及其相应的标签组成。正如 1.4 节将描述的，针对不同的学习场景，训练样本是不同的。
- **验证样本**(validation sample)：用来调整学习算法参数的样本，这里的学习算法针对的是带标签的数据。学习算法通常具有一个或多个自由参数，验证样本被用来选择合适的模型参数值。
- **测试样本**(test sample)：用来评估学习算法性能的样本。测试样本与训练以及验证样本分开，在学习阶段是不可知的。在垃圾邮件问题中，测试样本由电子邮件样本组成，学习算法基于特征预测标签。通过比较这些预测的标签与测试样本的真实标签来衡量算法的性能。
- **损失函数**(loss function)：衡量预测标签和真实标签之间的差异或损失的函数。将所有标签集合记为 $\mathcal{Y}$，可能的预测集合记为 $\mathcal{Y}'$，损失函数为映射 L：$\mathcal{Y}\times\mathcal{Y}'\to\mathbb{R}_+$。在大多数情形下，$\mathcal{Y}'=\mathcal{Y}$，并且损失函数是有界的，但是这些条件并不总是成立的。常见的几种损失函数包括 0-1(或误分类)损失和平方损失，前者为定义在$\{-1, +1\}\times\{-1, +1\}$上的函数 $L(y, y')=1_{y'\neq y}$，后者为定义在 $I\times I$ 上的函数 $L(y, y')=(y'-y)^2$，其中，$I\subseteq\mathbb{R}$ 通常为有界区间。
- **假设集**(hypothesis set)：将特征(特征向量)映射到标签集合 $\mathcal{Y}$ 的函数集合。在我们的例子中，可能是将电子邮件特征映射到 $\mathcal{Y}=\{$垃圾，非垃圾$\}$的函数集合。更一般地，假设集中的假设可以是将特征映射到不同集合 $\mathcal{Y}'$ 的函数。在本例中，可以是将电子邮件特征向量映射到实数的线性函数，实数可以解释为**得分**(score，$\mathcal{Y}'=\mathbb{R}$)，得分越高说明越有可能是垃圾邮件。

现在我们来定义垃圾邮件的学习过程。给定带标签的样本集合，我们首先将数据随机划分为训练样本、验证样本和测试样本。样本的大小依不同的考虑而定。比如，用于验证的数据量取决于算法自由参数的个数。而且，当带标签的样本相对较少时，通常所选择的训练数据个数要比测试数据的多，因为学习性能直接依赖于训练样本。

其次，为每个样本关联与之相关的特征，这是设计机器学习算法的一个关键步骤。有用的特征可以有效地指导学习算法设计，相反，不好或者无信息的特征可能具有误导性。尽管这至关重要，但是，选择哪些特征在很大程度上还是交由使用者来决定。这个选择反

映了使用者对于学习任务的**先验知识**(prior knowledge)，对实际性能结果影响很大。

接下来，我们基于所选择的特征训练学习算法，训练过程中为自由参数固定不同的取值。根据这些参数的每种取值，算法可以从假设集合中得到不同的假设，我们通常从得到的假设中选择在验证样本上性能最佳的假设。最后，利用该假设预测测试样本的样例标签，通过比较预测标签和真实标签，我们利用损失函数评估算法的性能，损失函数是任务相关的，比如在垃圾邮件检测任务中会采用0-1损失。

因此，算法的性能是基于测试误差进行评估的，而不是在训练样本上的误差。学习算法可以是**一致的**(consistent)，即对训练数据可以完全无误地划分，但这种一致的算法可能在测试数据上的性能很差。对于由复杂决策平面定义的一致学习器，往往发生这样的情况，如图1-1所示，它们倾向于记住相对较少的训练样本而不是试图泛化得更好。这说明了记忆与泛化之间的主要区别，这也是高精度学习算法寻求的基本性质。一致学习器的理论保证将在本书第2章中详细讨论。

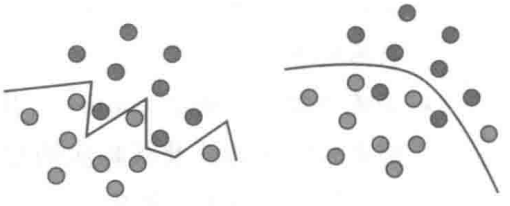


图1-1　左图的锯齿形曲线在蓝色和灰色训练样本上是一致的，但是这个复杂决策平面不可能很好地泛化到未见数据上。相反，右图的决策面更简单，虽然在训练样本上误分类几个点，但可能泛化得更好

1.3　交叉验证

实际上，可用的带标签样本数量很少，经常甚至少到无法留出验证样本，否则可能使得训练样本量不足。因此，不同于我们之前介绍的学习过程，一种被称为***n*-折交叉验证**(*n*-fold cross-validation)的方法得到了广泛采用，即利用带标签样本进行**模型选择**(model selection，选择算法的自由参数)和训练。

令$\boldsymbol{\theta}$表示算法自由参数向量。对于给定的$\boldsymbol{\theta}$值，该方法首先将包含m个带标签样本的样本集S随机划分为n组子样本，或称n折，其中第i折是样本规模为m_i的带标签样本$((x_{i1}, y_{i1}), \cdots, (x_{im_i}, y_{im_i}))$。于是，对于任何$i \in [1, n]$，学习算法在除了第$i$折之外的所有数据上进行训练，并生成假设$h_i$，$h_i$的性能在第$i$折上进行测试，如图1-2a所示。基于假设$h_i$的平均误差，称为**交叉验证误差**，对参数值$\boldsymbol{\theta}$进行评估。该误差用$\hat{R}_{\mathrm{CV}}(\boldsymbol{\theta})$表示，定义为

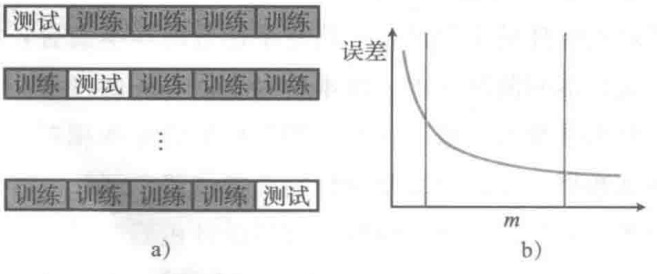


图1-2　*n*-折交叉验证。a)将训练数据划分为5折的图例。b)分类器预测误差随着训练样本个数变化的典型曲线图：误差随着训练点个数的增加而递减

$$\hat{R}_{\mathrm{CV}}(\boldsymbol{\theta})=\frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{1}{m_i}\sum_{j=1}^{m_i}L(h_i(x_{ij}),y_{ij})}_{h_i\text{在第}i\text{折上的误差}}$$

每折通常具有相同的大小，也就是对于所有 $i\in[1,\ n]$，有 $m_i=m/n$。那么，该如何选择 n？这需要从某种折中或权衡以及学习理论研究进展的角度做出合理的选择，目前我们在引言中还无法解决这个问题。在 n-折交叉验证中，每个训练样本集的规模为 $m-m/n=m(1-1/n)$，n 值越大(如图 1-2b 中右侧垂直的灰线所示)，则与 m(即全部样本大小)越接近，但是划分出的这些训练样本集很相似。因此，此时往往导致测试结果偏差小而方差大。相反，n 值越小，划分出的训练样本集之间的差异越大，同时每个训练样本集的规模(如图 1-2b 中左侧垂直的灰线所示)要明显小于 m。于是，此时往往导致测试结果方差小而偏差大。

5

在机器学习应用中，n 通常选择为 5 或 10。n-折交叉验证在模型选择中按如下方式使用。首先将全部标签数据划分为训练和测试样本，对于 $\boldsymbol{\theta}$ 的少量可能的取值，基于个数为 m 的训练样本计算 n-折交叉验证的误差 $\hat{R}_{\mathrm{CV}}(\boldsymbol{\theta})$。进而，可以得到使得 $\hat{R}_{\mathrm{CV}}(\boldsymbol{\theta})$ 取值最小的参数值 $\boldsymbol{\theta}_0$。根据 $\boldsymbol{\theta}_0$，便可如前一节中所述，在全部训练样本上(样本个数为 m)训练算法，并在测试样本上评估其性能。

n-折交叉验证的特殊情形就是当 $n=m$ 时，被称为**留一交叉验证**(leave-one-out cross-validation)，这是由于在每次迭代时，只有一个实例从训练样本中移出。之后将在第 4 章中介绍，平均留一误差为算法平均误差的近似无偏估计，可以被用来推导一些算法的简单理论保证。通常，留一误差的计算成本昂贵，这是由于需要在规模为 $m-1$ 的样本集上训练 n 次，不过对于某些算法，还是存在一些高效计算的方式来降低计算成本(见习题 10.9)。

除了模型选择，n-折交叉验证也常被用于性能评估。在这种情形下，给定参数向量 $\boldsymbol{\theta}$，全部带标签样本被随机划分为 n 折，其中训练和测试样本并无差别。用于评估的性能是全部样本上的 n-折交叉验证误差以及每折误差的标准差。

6

1.4 学习情境

我们接下来简要描述一下常见的机器学习情境。这些情境的区别在于可用训练数据的类型、到达顺序、获得训练数据的方法以及用来评估学习算法的测试数据。

- **监督学习**(supervised learning)：学习器获得标签样本作为训练数据，并对未见数据进行预测。这是与分类、回归和排序问题相关联的最常见的情境。在前面小节中讨论的垃圾邮件检测问题是监督学习的一个实例。
- **无监督学习**(unsupervised learning)：学习器只获得无标签训练数据，并对未见数据进行预测。由于标签样例在该情形下通常是不可获得的，所以定量地评估学习器性能是很困难的。聚类和维数约简是无监督学习问题的实例。

- **半监督学习**(semi-supervised learning)：学习器获得的训练样本由标签数据和无标签数据组成，并对未见数据进行预测。半监督学习在无标签数据容易获得而标签数据获得成本高的情境下是很常见的。应用中出现的很多类型问题，包括分类、回归或者排序任务，都可以被框定为半监督学习的实例。所希望的就是借助可用的无标签数据的分布，使得学习器取得比监督情境下更好的性能，分析其真正可实现的条件是当今很多理论和应用机器学习研究的主题。
- **直推学习**(transductive inference)：正如半监督情境，学习器获得标签训练样本以及无标签测试数据集合。但是，直推学习的目标是仅对特定测试数据预测标签。直推学习看似更为简单，且与各种现代应用中遇到的情境相吻合。然而，与半监督学习情境类似，该情境在何种假设下可以取得更好的性能仍在研究中，至今还没有彻底得到解决。
- **在线学习**(on-line learning)：与前面的情境相比，在线情境下学习需要多轮，同时训练和测试阶段混在一起。在每一轮，学习器获得一个无标签训练数据，对其做出预测之后，获得真实标签，并产生损失。在线情境下的目标是最小化所有轮的累积损失。与前面讨论的情境有所不同，在线学习中不做任何分布假设。事实上，在该情境中可能对抗式地选择实例及其标签。

7

- **强化学习**(reinforcement learning)：在强化学习中训练和测试阶段也混合在一起。为了收集信息，学习器主动地与环境进行交互，在一些情况下影响环境，并获得每个行动的即时奖赏。学习器的目标是经过一系列的行动以及与环境的交互最大化获得的奖赏。然而，环境不提供长期奖赏反馈，学习器必须在探索未知行动以获得更多信息与利用已收集信息之间进行选择，因此学习器面临着**探索还是利用**(exploration versus exploitation)的困境。
- **主动学习**(active learning)：学习器自适应地或者交互式地收集训练样本，通常以询问专家的方式请求新样本的标签。主动学习的目标是利用更少的带标签样本达到与标准监督学习可比较的性能。主动学习常被用在标签获得成本高的实际应用中，例如计算生物学应用。

实际应用中，还可能遇到许多其他需要折中考虑以及更为复杂的学习情境。

1.5　本书概览

本书介绍了若干基本的并且在数学上经过充分研究的算法。本书深入讨论了理论基础和实际应用，具体包括的内容如下：

- 概率近似正确(PAC)学习框架、有限假设集的学习保证；
- 无限假设集的学习保证、Rademacher 复杂度、VC-维；
- 支持向量机(SVM)、间隔理论；
- 核方法、正定对称核、表示定理、有理核；