

需要整本电子书，联系我QQ：3411317522；
此处是样章，取的完整版的前面几页，和最后
面几页；完整版是带书签的，样章没带书签；
另外需要其他书，也可以找我。

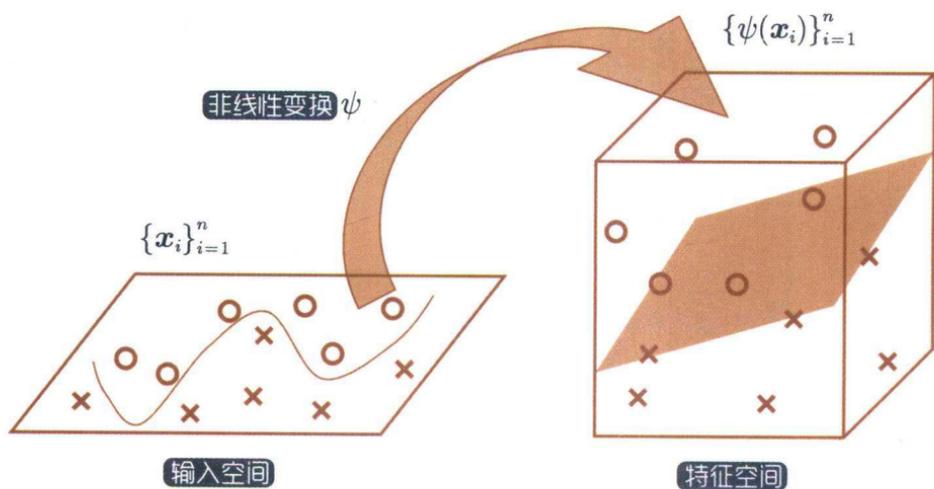
TURING

图灵程序
设计丛书

图解机器学习

[日] 杉山将 著

许永伟 译



187张图解轻松入门

提供可执行的MATLAB程序代码

覆盖机器学习中最经典、用途最广的算法



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

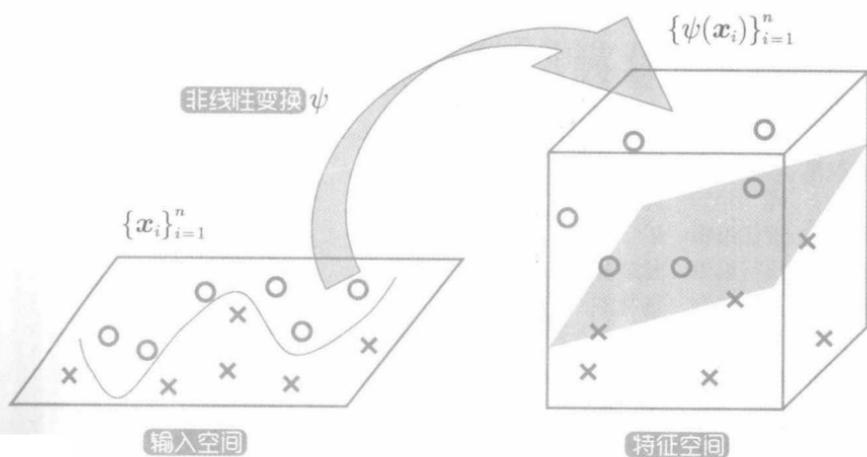
TURING

图灵程序
设计丛书

图解机器学习

[日] 杉山将 著

许永伟 译



人民邮电出版社

北京

图书在版编目 (C I P) 数据

图解机器学习 / (日) 杉山将著 ; 许永伟译. — 北京 : 人民邮电出版社, 2015. 4
(图灵程序设计丛书)
ISBN 978-7-115-38802-5

I. ①图… II. ①杉… ②许… III. ①机器学习—图解 IV. ①TP181-64

中国版本图书馆CIP数据核字(2015)第052954号

内 容 提 要

本书用丰富的图示,从最小二乘法出发,对基于最小二乘法实现的各种机器学习算法进行了详细的介绍。第 I 部分介绍了机器学习领域的概况;第 II 部分和第 III 部分分别介绍了各种有监督的回归算法和分类算法;第 IV 部分介绍了各种无监督学习算法;第 V 部分介绍了机器学习领域中的新兴算法。书中大部分算法都有相应的 MATLAB 程序源代码,可以用来进行简单的测试。

本书适合所有对机器学习有兴趣的初学者阅读。

-
- ◆ 著 [日] 杉山将
译 许永伟
责任编辑 乐馨
执行编辑 杜晓静
责任印制 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京天宇星印刷厂印刷
 - ◆ 开本: 880 × 1230 1/32
印张: 7.5
字数: 209 千字 2015 年 4 月第 1 版
印数: 1-4 000 册 2015 年 4 月北京第 1 次印刷
著作权合同登记号 图字: 01-2014-3345 号

定价: 49.00 元

读者服务热线: (010)51095186 转 600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京崇工商广字第 0021 号

站在巨人的肩上
Standing on Shoulders of Giants



iTuring.cn

版权声明

IRASUTO DE MANABU KIKAI GAKUSYUU

SAISYOUNIJYOUHOU NI YORU SHIKIBETSU MODERU GAKUSYUU O TYUUSHIN NI

© Masashi Sugiyama 2013

All rights reserved.

Original Japanese edition published by KODANSHA LTD.

Publication rights for Simplified Chinese character edition arranged with
KODANSHA LTD. through

KODANSHA BEIJING CULTURE LTD. Beijing, China.

本书由日本讲谈社授权人民邮电出版社发行简体字中文版，版权所有，未经书面同意，不得以任何方式做全面或局部翻印、仿制或转载。

译者序

机器学习领域是深不可测的吗？人工智能是天方夜谭吗？时至今日，机器学习研究的重要性与可行性已得到广泛承认，并在模式识别、通信、控制、金融、机器人、生物信息学等许多领域都有着广泛的应用。

如何自动归类筛选邮件和网页？如何向大家推荐你可能感兴趣的人？如何预测整体市场行情的好坏？如何从统计学的角度对照片进行归类？本书就介绍了这样一些算法。

如果想得到最通俗、简洁的讲解，本书最为合适。

如果想立即知道算法的性能，并期望有可运行的源代码，本书最为方便。

很多人都是看着日本的动画长大的。殊不知，大部分日本人都具有熟练的绘画能力。他们总可以把复杂、枯燥的事物用惟妙惟肖的漫画生动地表达出来。广告、网页、海报，甚至政府公告都图文并茂。市面上也有不计其数的“图解……”“图说……”一类的书籍。本书就是其中一例，这也是本书的最大特点。

杉山将博士今年赴任东京大学教授，他在机器学习领域颇有建树。他的研究室吸引了来自世界各地的机器学习研究者。本书承袭了日本特有的绘画特色，依靠作者丰富的机器学习经验，用最精简的文字，将原本复杂抽象的数学原理，用形象的漫画与数据图形进行了清晰的说明。作者也将最前沿和最核心的研究成果汇集到了本书之中。

本书的侧重点不在于机器学习原理的相关推导，而在于结论的分析和应用。读者朋友可以更快地掌握各种算法的特点和使用方法，提纲挈领地消化应用，而不必拘泥于算法的细节不能自拔。另外，本书

旁征博引，图文并茂，结构清晰，范例实用丰富，深入浅出地说明了机器学习中最典型和用途最广泛的算法。

本书内容覆盖面广，不但与市面上众多的机器学习书籍并无重复，更可与其互为补充。大部分算法都有简洁、现成的MATLAB源代码，读者朋友可以轻松地进行验证。以此为原型，再稍加修改扩充，即可做出为自己所用的项目代码。

机器学习领域日新月异，书中所涉及的概念和术语数目繁多，且有许多概念和术语目前尚无公认的中文译法。如果有不合读者朋友习惯的术语出现，请参考译者注，确认其原始词意。

本译稿得到了图灵公司编辑的悉心指导，她们为保证本书的质量做了大量的补译、校正及编辑工作，在此表示深深的谢意。

许永伟

2014年12月于东京

序

本书是关于机器学习的入门图书。说到“机器”，可能很多人都会想到机械表或车床等大型机器设备，但是机器学习里的“机器”指的是计算机。机器学习，是指让计算机具有人那样的学习、思考能力的技术的总称。近年来，随着计算机软硬件技术的发展，机器学习领域也得到了巨大的进步。本书就是介绍这一蓬勃发展中的机器学习算法的一本书。

在机器学习领域，借助高级的数学方法，各种新型算法层出不穷。因此对于初涉这一领域的研究人员、技术工作者和学生来说，要理解这些最前沿的技术往往有很多困难。然而大部分这些最新的机器学习算法，都是在最经典的算法——最小二乘法的基础上发展起来的。本书就是立足于这样的视点，对基于最小二乘法实现的各种机器学习算法做简单的介绍，并给出许多具体的实例。因此，只要理解了最小二乘法的基本原理，即可掌握能够处理中等数据规模的大多数高级算法。

本书由以下几部分构成。

第Ⅰ部分介绍了本书所涉及的机器学习领域的概况。首先在第1章，对监督学习、无监督学习和强化学习等基本概念进行了说明；第2章介绍了机器学习里需要使用到的各种各样的模型。

第Ⅱ部分介绍了与连续函数的近似问题相对应的各种回归算法。具体来说，首先在第3章引入了回归算法的基础，即最小二乘学习法；第4章介绍了能够避免过拟合问题的条件约束的最小二乘学习法。第5章介绍了通过把大部分参数置为0来大幅提高学习效率、计算精度的稀疏算法。第6章介绍了不易受到异常值影响的鲁棒学习法。

第Ⅲ部分介绍了各种分类算法。第7章介绍了回归问题中直接使用最小二乘学习法进行分类的算法。第8章引入了基于间隔最大化原

理的支持向量机分类器的算法，并且明确了支持向量机分类器和最小二乘学习法之间的关系，还介绍了把支持向量机分类器向鲁棒学习扩展的方法。第9章引入了把多个性能稍弱的分类器组合在一起生成高性能的分类器的集成学习法，介绍了Bagging和Boosting算法。第10章介绍了把各个模式以概率进行分类的Logistic回归的分类算法，以及最小二乘学习版的最小二乘概率分类器。第11章介绍了能够处理字符串那样的序列数据的模式分类的条件随机场。

第IV部分介绍了各种无监督学习算法。第12章介绍了除去数据中的异常值的方法。第13章介绍了把高维数据降到低维后进行学习的降维算法。第14章介绍了把数据集合化的聚类算法。

第V部分介绍了机器学习领域中的新兴算法。第15章介绍了把训练样本逐次输入的逐次学习算法。第16章介绍了在输入输出成对出现的训练样本集的基础上，灵活应用只有输入的训练样本集的半监督学习算法。第17章介绍了有监督的降维算法。第18章介绍了灵活应用其他学习任务的信息，来提高当前学习任务的学习精度的迁移学习法。第19章介绍了在多个学习任务之间实现信息共享，然后同时进行求解的高性能的多任务学习算法。

在第VI部分的第20章，主要论述了机器学习领域今后的发展。

如图1所示，第II部分、第III部分和第IV部分是相对独立的章节。但是第II部分的第5章和第6章，以及第III部分的第8章、第9章和第11章，包含稍有难度的数学内容，初学者在最开始的时候可以跳过这些内容。

对于书中的大部分算法，本书同时提供了能够进行简单的数值计算的MATLAB程序源代码。各个程序都浓缩在一页的范围内，读者朋友可以轻松录入，以对书中的各种学习算法进行简单的测试。另外，在各个程序行首添加如下代码：

```
rand('state', 0); randn('state', 0);
```

即可完全再现本书中介绍的所有实例的计算结果。

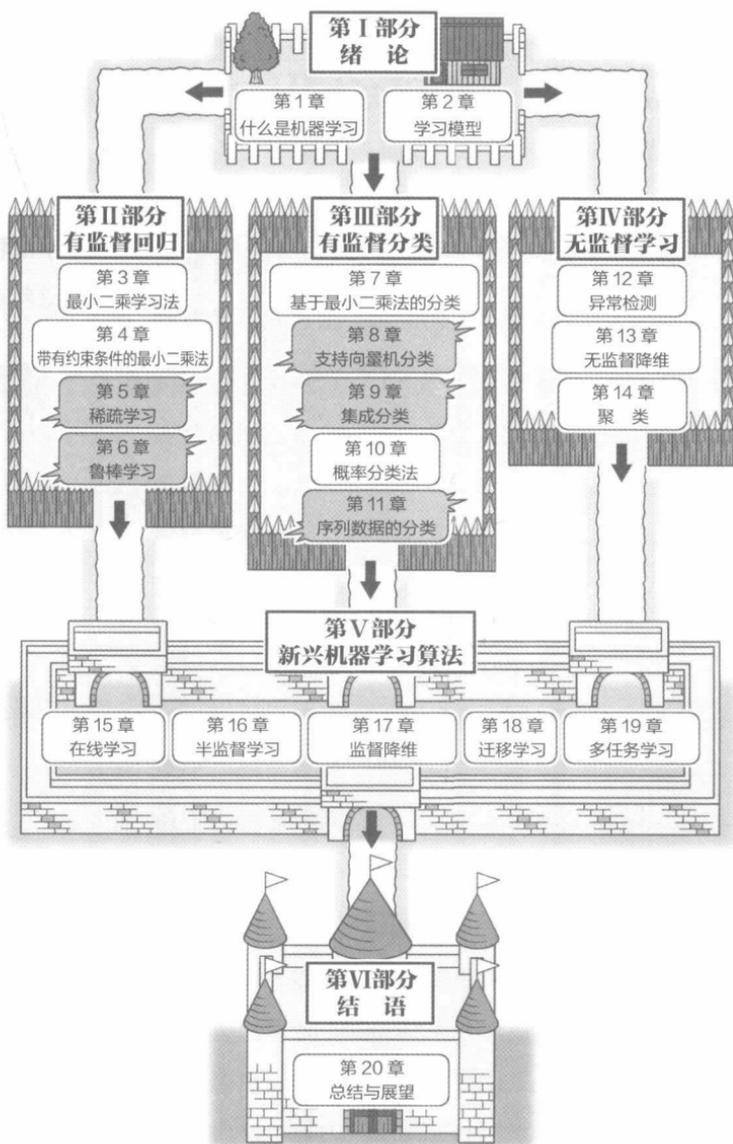


图1 本书的构成

最后，在本书执笔过程中，名古屋大学的金森敬文副教授、名古屋工业大学竹内一郎副教授、NTT Communications 科学技术研究所的山田诚博士、东京大学的鹿岛久嗣副教授、东京大学的武田朗子副教授、东京工业大学的山根一航先生、讲谈社的横山真吾先生、绘制插图的 Horiguchi Hiroshi 先生，给予了笔者巨大的支持和鼓励，在此一并表示真诚的感谢。

杉山将

2013年6月

目 录

第I部分 绪 论

| | | |
|-----|---------------|----|
| 第1章 | 什么是机器学习 | 2 |
| | 1.1 学习的种类 | 2 |
| | 1.2 机器学习任务的例子 | 4 |
| | 1.3 机器学习的方法 | 8 |
| 第2章 | 学习模型 | 12 |
| | 2.1 线性模型 | 12 |
| | 2.2 核模型 | 15 |
| | 2.3 层级模型 | 17 |

第II部分 有监督回归

| | | |
|-----|--------------------------|----|
| 第3章 | 最小二乘学习法 | 22 |
| | 3.1 最小二乘学习法 | 22 |
| | 3.2 最小二乘解的性质 | 25 |
| | 3.3 大规模数据的学习算法 | 27 |
| 第4章 | 带有约束条件的最小二乘法 | 31 |
| | 4.1 部分空间约束的最小二乘学习法 | 31 |
| | 4.2 l_2 约束的最小二乘学习法 | 33 |
| | 4.3 模型选择 | 37 |
| 第5章 | 稀疏学习 | 43 |
| | 5.1 l_1 约束的最小二乘学习法 | 43 |
| | 5.2 l_1 约束的最小二乘学习的求解方法 | 45 |
| | 5.3 通过稀疏学习进行特征选择 | 50 |

| | | |
|------------|------------------------------------|-----------|
| 5.4 | ℓ_p 约束的最小二乘学习法 | 51 |
| 5.5 | $\ell_1 + \ell_2$ 约束的最小二乘学习法 | 52 |
| 第6章 | 鲁棒学习 | 55 |
| 6.1 | ℓ_1 损失最小化学习 | 56 |
| 6.2 | Huber 损失最小化学习 | 58 |
| 6.3 | 图基损失最小化学习 | 63 |
| 6.4 | ℓ_1 约束的 Huber 损失最小化学习 | 65 |

第 III 部分 有监督分类

| | | |
|-------------|---------------------------|------------|
| 第7章 | 基于最小二乘法的分类 | 70 |
| 7.1 | 最小二乘分类 | 70 |
| 7.2 | 0/1 损失和间隔 | 73 |
| 7.3 | 多类别的情形 | 76 |
| 第8章 | 支持向量机分类 | 80 |
| 8.1 | 间隔最大化分类 | 80 |
| 8.2 | 支持向量机分类器的求解方法 | 83 |
| 8.3 | 稀疏性 | 86 |
| 8.4 | 使用核映射的非线性模型 | 88 |
| 8.5 | 使用 Hinge 损失最小化学习来解释 | 90 |
| 8.6 | 使用 Ramp 损失的鲁棒学习 | 93 |
| 第9章 | 集成分类 | 98 |
| 9.1 | 剪枝分类 | 98 |
| 9.2 | Bagging 学习法 | 101 |
| 9.3 | Boosting 学习法 | 105 |
| 第10章 | 概率分类法 | 112 |
| 10.1 | Logistic 回归 | 112 |
| 10.2 | 最小二乘概率分类 | 116 |

| | | |
|-------------|-------------------------------|------------|
| 第11章 | 序列数据的分类 | 121 |
| | 11.1 序列数据的模型化 | 122 |
| | 11.2 条件随机场模型的学习 | 125 |
| | 11.3 利用条件随机场模型对标签序列进行预测 | 128 |

第IV部分 无监督学习

| | | |
|-------------|-----------------------|------------|
| 第12章 | 异常检测 | 132 |
| | 12.1 局部异常因子 | 132 |
| | 12.2 支持向量机异常检测 | 135 |
| | 12.3 基于密度比的异常检测 | 137 |
| 第13章 | 无监督降维 | 143 |
| | 13.1 线性降维的原理 | 144 |
| | 13.2 主成分分析 | 146 |
| | 13.3 局部保持投影 | 148 |
| | 13.4 核函数主成分分析 | 152 |
| | 13.5 拉普拉斯特征映射 | 155 |
| 第14章 | 聚类 | 158 |
| | 14.1 K均值聚类 | 158 |
| | 14.2 核K均值聚类 | 160 |
| | 14.3 谱聚类 | 161 |
| | 14.4 调整参数的自动选取 | 163 |

第V部分 新兴机器学习算法

| | | |
|-------------|--------------------|------------|
| 第15章 | 在线学习 | 170 |
| | 15.1 被动攻击学习 | 170 |
| | 15.2 适应正则化学习 | 176 |

| | | |
|-------------|-------------------------------|------------|
| 第16章 | 半监督学习 | 181 |
| | 16.1 灵活应用输入数据的流形构造 | 182 |
| | 16.2 拉普拉斯正则化最小二乘学习的求解方法 | 183 |
| | 16.3 拉普拉斯正则化的解释 | 186 |
| 第17章 | 监督降维 | 188 |
| | 17.1 与分类问题相对应的判别分析 | 188 |
| | 17.2 充分降维 | 195 |
| 第18章 | 迁移学习 | 197 |
| | 18.1 协变量移位下的迁移学习 | 197 |
| | 18.2 类别平衡变化下的迁移学习 | 204 |
| 第19章 | 多任务学习 | 212 |
| | 19.1 使用最小二乘回归的多任务学习 | 212 |
| | 19.2 使用最小二乘概率分类器的多任务学习 | 215 |
| | 19.3 多次维输出函数的学习 | 216 |

第VI部分 结 语

| | | |
|-------------|--------------|------------|
| 第20章 | 总结与展望 | 222 |
| | 参考文献 | 225 |

第 I 部分 绪 论



什么是机器学习

近些年来，得益于互联网的普及，我们可以非常轻松地获取大量文本、音乐、图片、视频等各种各样的数据。机器学习，就是让计算机具有像人一样的学习能力的技术，是从堆积如山的数据（也称为大数据）中寻找出有用知识的数据挖掘技术。通过运用机器学习技术，从视频数据库中寻找出自己喜欢的视频资料，或者根据用户的购买记录向用户推荐其他相关产品等成为了现实（图 1.1）。本章将从宏观角度对什么是机器学习做相应的介绍，并对机器学习的基本概念进行说明。

1.1 学习的种类

计算机的学习，根据所处理的数据种类的不同，可以分为监督学习、无监督学习和强化学习等几种类型。

监督学习，是指有求知欲的学生从老师那里获取知识、信息，老师提供对错指示、告知最终答案的学习过程（图 1.2）。在机器学习里，学生对应于计算机，老师则对应于周围的环境。根据在学习过程中所获得的经验、技能，对没有学习过的问题也可以做出正确解答，使计



图 1.1 机器学习

计算机获得这种泛化能力，是监督学习的最终目标。监督学习，在手写文字识别、声音处理、图像处理、垃圾邮件分类与拦截、网页检索、基因诊断以及股票预测等各个方面，都有着广泛的应用。这一类机器学习的典型任务包括：预测数值型数据的回归、预测分类标签的分类、预测顺序的排序等。

无监督学习，是指在没有老师的情况下，学生自学的过程(图1.3)。在机器学习里，基本上都是计算机在互联网中自动收集信息，并从中获取有用信息。无监督学习不仅仅局限于解决像监督学习那样的有明确答案的问题，因此，它的学习目标可以不必十分明确。无监督学习在人造卫星故障诊断、视频分析、社交网站解析和声音信号解析等方面大显身

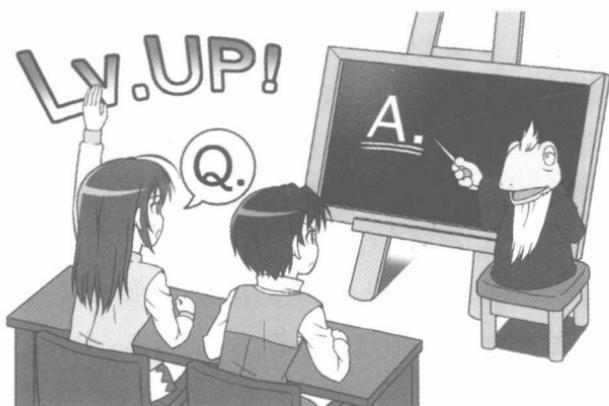


图1.2 监督学习



图1.3 无监督学习

手的同时，在数据可视化以及作为监督学习方法的前处理工具方面，也有广泛的应用。这一类机器学习的典型任务有聚类、异常检测等。

强化学习，与监督学习类似，也以使计算机获得对没有学习过的问题做出正确解答的泛化能力为目标，但是在学习过程中，不设置老师提示对错、告知最终答案的环节。然而，如果真的在学习过程中不能从周围环境中获得任何信息的话，强化学习就变成无监督学习了。强化学习，是指在没有老师提示的情况下，自己对预测的结果进行评估的方法。通过这样的自我评估，学生为了获得老师的最高嘉奖而不断地进行学习(图1.4)。婴幼儿往往会为了获得父母的表扬去做事情，因此，强化学习被认为是人类主要的学习模式之一。强化学习，在机器人的自动控制、计算机游戏中的人工智能、市场战略的最优化等方面均有广泛应用。在强化学习中经常会用到回归、分类、聚类和降维等各种各样的机器学习算法。

1.2 机器学习任务的例子

有关增强学习的详细解说，读者朋友可以参阅文献[5]。本节将对监督学习和无监督学习中典型的任务，例如回归、分类、异常检测、聚类和降维等做一一介绍。



图1.4 强化学习

回归，是指把实函数在样本点附近加以近似的有监督的函数近似问题^①(图1.5)。这里，我们来考虑一下以 d 次方的实数向量 \boldsymbol{x} 作为输入，实数值 y 作为输出的函数 $y=f(\boldsymbol{x})$ 的学习问题。在监督学习里，这里的真实函数关系 f 是未知的，作为训练集的输入输出样本 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ 是已知的。但是，一般情况下，在输出样本 y_i 的真实值 $f(\boldsymbol{x}_i)$ 中经常会观测到噪声。通过这样的设定，输入样本 \boldsymbol{x}_i 就是学生向老师请教的问题，输出样本 y_i 是老师对学生的解答，输出样本中包含的噪声则与老师的教学错误或学生的理解错误相对应。老师的知识(无论什么样的问题，都可以做出正确的解答)与真实的函数 f 相对应，使学生获得这个函数就是监督学习的最终目标。如果以 \hat{f} 来表示学生通过学习而获得的函数，那么学生对没有学习过的问题也可以做出正确解答的泛化能力的大小，就可以通过比较函数 f 和 \hat{f} 的相似性来进行分析。

分类，是指对于指定的模式进行识别的有监督的模型识别问题(图1.6)。在这里，以 d 次方的实数向量 \boldsymbol{x} 作为输入样本，而所有的输入样本，可以被划分为 c 个类别的问题来进行说明。作为训练集的输入输出样本 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ 是已知的。但是，分类问题中的输出样本 y_i ，并不是具体的实数，而是分别代表类别 $1, 2, \dots, c$ 。在这样的任务里，得到输出类别 $1, 2, \dots, c$ 的函数 $y=f(\boldsymbol{x})$ 的过程，就是机器学习的过程。因此，分类问题也可以像回归问题那样，被看作是函数近似问题。然而，在

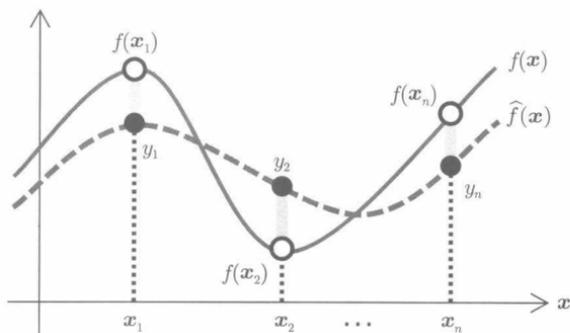


图1.5 回归

^① 回归是对一个或多个自变量和因变量之间的关系进行建模、求解的一种统计方法。——译者注

分类问题中，并不存在类别1比类别3更接近于类别2这样的说法。分类问题只是单纯地对样本应该属于哪一个类别进行预测，并根据预测准确与否来衡量泛化误差，这一点与回归是不同的。

异常检测，是指寻找输入样本 $\{\mathbf{x}_i\}_{i=1}^n$ 中所包含的异常数据的问题。在已知正常数据与异常数据的例子的情况下，其与有监督的分类问题是相同的。但是，一般情况下，在异常检测任务中，对于什么样的数据是异常的，什么样的数据是正常的，在事先是未知的。在这样的无监督的异常检测问题中，一般采用密度估计的方法，把靠近密度中心的数据作为正常的数据，把偏离密度中心的数据作为异常的数据(图1.7)。

聚类，与分类问题相同，也是模式识别问题，但是属于无监督学习的一种(图1.8)。即只给出输入样本 $\{\mathbf{x}_i\}_{i=1}^n$ ，然后判断各个样本分别属于1, 2, \dots , c 中的哪个簇^①。隶属于相同簇的样本之间具有相似的性质。

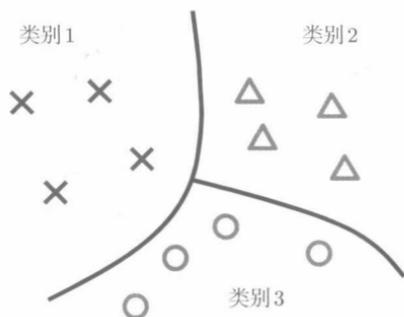


图1.6 分类

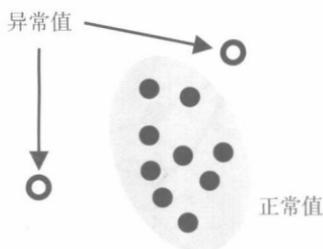


图1.7 异常检测

^① 聚类问题中经常以“簇”代替“类别”。——译者注

质，不同簇的样本之间具有不同的性质。在聚类问题中，如何准确地计算样本之间的相似度是很重要的课题。

降维，是指从高维度数据中提取关键信息，将其转换为易于计算的低维度问题进而求解的方法(图1.9)。具体来说，当输入样本 $\{\mathbf{x}_i\}_{i=1}^n$ 的维数 d 非常大的时候，可以把样本转换为较低维度的样本 $\{\mathbf{z}_i\}_{i=1}^n$ 。线性降维的情况下，可以使用横向量 \mathbf{T} 将其变换为 $\mathbf{z}_i = \mathbf{T}\mathbf{x}_i$ 。降维，根据数据种类的不同，可以分为监督学习和无监督学习两种。作为训练集的输入输出样本 $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ 是已知的时候，属于监督学习，可以把样本转换为较低维度的样本 $\{\mathbf{z}_i\}_{i=1}^n$ ，从而获得较高的泛化能力。与之相对，如果只有输入样本 $\{\mathbf{x}_i\}_{i=1}^n$ 是已知的话，就属于无监督学习，在转换为较低维度的样本 $\{\mathbf{z}_i\}_{i=1}^n$ 之后，应该保持原始输入样本 $\{\mathbf{x}_i\}_{i=1}^n$ 的数据分布性质，以及数据间的近邻关系不发生变化。

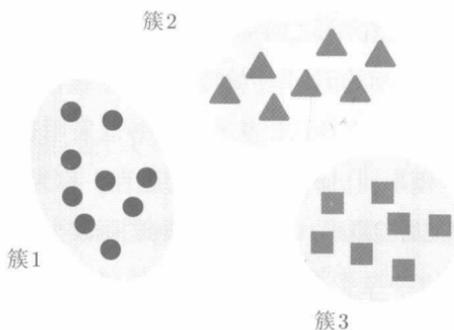


图 1.8 聚类分析

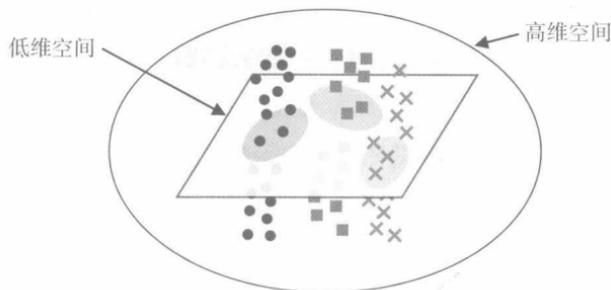


图 1.9 降维

1.3 机器学习的方法

机器学习有多种不同的流派。本节中，以对模式 \mathbf{x} 的类别 y 进行预测的分类问题为例，对机器学习中的主要流派，即产生式分类和判别式分类，以及频率派和贝叶斯派的基本方法加以介绍。

1.3.1 生成的分类和识别的分类

在已知模式 \mathbf{x} 的时候，如果能求得使分类类别 y 的条件概率 $p(y|\mathbf{x})$ 达到最大值的类别 \hat{y} 的话，就可以进行模式识别了。

$$\hat{y} = \underset{y}{\operatorname{argmax}} p(y|\mathbf{x})$$

在这里，“argmax”是取得最大值时的参数的意思。所以， $\max_y p(y|\mathbf{x})$ 是指当 y 取特定值时 $p(y|\mathbf{x})$ 的最大值，而 $\operatorname{argmax}_y p(y|\mathbf{x})$ 是指当 $p(y|\mathbf{x})$ 取最大值时对应的 y 的值(图 1.10)。在模式识别里，条件概率 $p(y|\mathbf{x})$ 通常也称为后验概率。上面的 \hat{y} 读作 y 翰特。在基于统计分析的机器学习中，预测结果一般以字母加符号 $\hat{\quad}$ 来表示，本书也采用这样的方法。应用训练集直接对后验概率 $p(y|\mathbf{x})$ 进行学习的过程，称为判别式分类。

另外，还可以把后验概率 $p(y|\mathbf{x})$ 表示为 y 的函数。

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} \propto p(\mathbf{x}, y)$$

通过上式，我们可以发现模式 \mathbf{x} 和类别 y 的联合概率 $p(\mathbf{x}, y)$ 与后验概率 $p(y|\mathbf{x})$ 是成比例的。正因为有这样的关系，我们可以通过使联合概率 $p(\mathbf{x}, y)$ 达到最大值的方法，来得到使后验概率 $p(y|\mathbf{x})$ 达到最大值的类别 \hat{y} 。

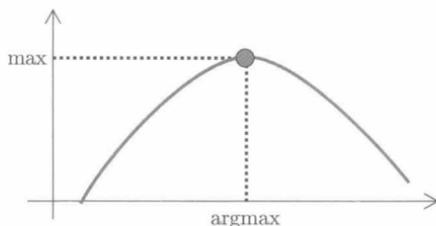


图 1.10 max 和 argmax

$$\hat{y} = \underset{y}{\operatorname{argmax}} p(\mathbf{x}, y)$$

在模式识别里，联合概率 $p(\mathbf{x}, y)$ 也称为数据生成概率，通过预测数据生成概率 $p(\mathbf{x}, y)$ 来进行模式识别的分类方法，称为生成的分类^[11]。

支持向量机分类器的发明者、著名的数学家弗拉基米尔·万普尼克^①在其著作^[15]中提到：

在实际问题中，信息往往是有限的。在解决一个感兴趣的问题时，不要把解决一个更一般的问题作为一个中间步骤。要试图得到所需要的答案，而不是更一般的答案。很可能你拥有足够的信息来很好地解决一个感兴趣的特定问题，但却没有足够的信息来解决一个一般性的问题。

为什么这么说呢？这是因为，即使手头的信息量不足以解决一般性问题，但对于解决特定问题，很可能是足够的。如果数据生成概率 $p(\mathbf{x}, y)$ 是已知的，

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}, y)}{\sum_y p(\mathbf{x}, y)}$$

那么，从上式就可以推出后验概率 $p(y|\mathbf{x})$ 。然而，如果后验概率 $p(y|\mathbf{x})$ 是已知的，却不能由此推导出数据生成概率 $p(\mathbf{x}, y)$ (图 1.11)。因此，比起计算后验概率 $p(y|\mathbf{x})$ ，可以说数据生成概率 $p(\mathbf{x}, y)$ 的计算是一般性(即求解更困难)的问题。进行模式识别时，只需计算出后验概率 $p(y|\mathbf{x})$ 就足够了。但在生成的分类中，则要计算数据生成概率 $p(\mathbf{x}, y)$ 这个一般性的问题。如果遵循上述的弗拉基米尔·万普尼克的原理，识别的分类就是比生成的分类更好的机器学习方法。

另一方面，在很多实际问题中，经常可以获得有关数据生成概率 $p(\mathbf{x}, y)$ 的一些先验知识。例如，在声音识别过程中，可以通过事先研究人类的喉咙构造或发声机理，获得很多有关数据生成概率 $p(\mathbf{x}, y)$ 的

① 全名 Vladimir Naumovich Vapnik，俄罗斯统计学家、数学家。——译者注

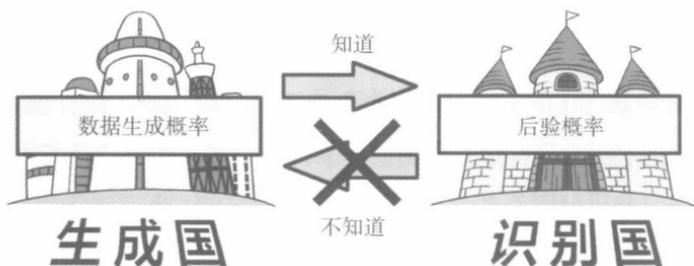


图1.11 万普尼克的理论

先验知识^[4]。像这样，在可以事先获得数据生成概率 $p(\mathbf{x}, y)$ 的先验知识的情况下，生成的分类就是比识别的分类更好的机器学习方法，即与上段论述是正好相反的。

1.3.2 统计概率和朴素贝叶斯

本小节中，我们以包含参数 θ 的模型 $q(\mathbf{x}, y; \theta)$ 为例，对计算数据生成概率 $p(\mathbf{x}, y)$ 的问题进行说明。

在统计概率的机器学习方法中，将模式 θ 作为决定论的变量，使用手头的训练样本 $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ 对模式 θ 进行学习。例如，在最大似然估计算法中，一般对生成训练集 \mathcal{D} 的最容易的方法所对应的模式 θ 进行学习。

$$\max_{\theta} \prod_{i=1}^n q(\mathbf{x}_i, y_i; \theta)$$

在统计概率方法中，如何由训练集 \mathcal{D} 得到高精度的模式 θ 是主要的研究课题。

与之相对，在朴素贝叶斯方法中，将模式 θ 作为概率变量，对其先验概率 $p(\theta)$ 加以考虑，计算与训练集 \mathcal{D} 相对应的后验概率 $p(\theta | \mathcal{D})$ 。通过运用贝叶斯定理，就可以使用先验概率 $p(\theta)$ 来求解后验概率 $p(\theta | \mathcal{D})$ ，如下所示：

$$p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta)p(\theta)}{p(\mathcal{D})} = \frac{\prod_{i=1}^n q(\mathbf{x}_i, y_i | \theta)p(\theta)}{\int \prod_{i=1}^n q(\mathbf{x}_i, y_i | \theta)p(\theta) d\theta}$$

方法也是现在的一个研究热点。例如，使用学习得到的回归器或分类器只对输入数据自动计算输出的自主学习，充分利用互联网将输出数据的给定工作分配给不特定多数人的群众外包方法，近年来也受到了越来越多的关注。

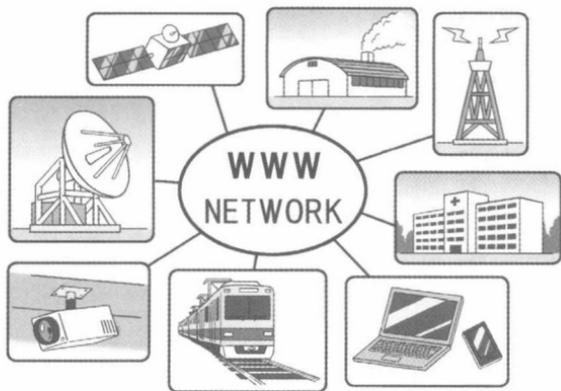


图 20.1 大数据。近些年来，随着互联网或多重传感器技术的应用，使得大量获取多种数据成为了可能

在第IV部分中，介绍了异常检测、降维、聚类等无监督学习方法。这些无监督学习算法不仅可以单独用来从数据中获得相应知识，还可以作为有监督学习算法的前处理方法，使得最终的学习精度得以提升。实际上已经有研究指出，对于2.3节中介绍的人工神经网络，利用无监督学习方法首先进行事先学习，可以使得后续监督学习的最终精度得以大幅提升^[6]。在大数据时代，这些无监督学习算法的重要性会日益提高。

本书中同时也列举了很多使用最小二乘法或梯度法等进行机器学习的简单实例。这些算法对中小规模的数据有很好的学习效果，但是对于稍大规模的数据，则需要花费大量的计算、处理时间。因此，如何基于高度最优化理论开发出计算效率高的机器学习算法，是大数据时代迫切需要解决的问题。除了第15章中介绍的在线学习和使用多个计算机进行并行处理的学习算法，在数据获取保存的模型、计算机或互联网的硬件系统设计等各个方面，有待进行的研究还有很多很多。

另外，从互联网或公共场所设置的多种多样的传感器中获取的数据，有时会包含个人的脸部数据或行动轨迹等信息。像这样，如何在保护个人隐私的前提下使得机器学习真正服务于现代社会、提高大众的生活品质，也是机器学习研究领域的一个重要课题。

参考文献

- [1] 赤穂昭太郎. カーネル多変量解析—非線形データ解析の新しい展開. 岩波書店, 2008.
- [2] 赤池弘次, 甘利俊一, 北川源四郎, 樺島祥介, 下平英寿. 赤池情報量規準 AIC—モデリング・予測・知識発見. 共立出版, 2007.
- [3] 麻生英樹, 津田宏治, 村田昇. パターン認識と学習の統計学—新しい概念と手法. 岩波書店, 2003.
- [4] 古井貞熙. 人と対話するコンピュータを創っています—音声認識の最前線. 角川学芸出版, 2009.
- [5] 八谷大岳, 杉山将. 強くなるロボティック・ゲームプレイヤーの作り方—実践で学ぶ強化学習. 毎日コミュニケーションズ, 2008.
- [6] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507, 2006.
- [7] 金森敬文, 畑埜晃平, 渡辺治. ブースティング—学習アルゴリズムの設計技法. 森北出版, 2006.
- [8] 鹿島久嗣. カーネル法による構造データマイニング. *情報処理*, Vol.46, No.1, pp. 27–33, 2005.
- [9] 川人光男. 脳の計算理論. 産業図書, 1996.
- [10] Christopher M. Bishop. *Pattern Recognition And Machine Learning*. Springer-Verlag New York Inc., 2006.
- [11] 杉山将. 統計的機械学習—生成モデルに基づくパターン認識. オーム社, 2009.
- [12] M. Sugiyama and M. Kawanabe. *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. The MIT Press, 2012.
- [13] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 2012.
- [14] 富岡亮太, 鈴木大慈, 杉山将. スパース正則化およびマルチカーネル学

習のための最適化アルゴリズムと画像認識への応用. 画像ラボ, Vol.21,
No.4, pp.5-11, 2010.

[15] 瓦普尼克(著), 张学工(译). 统计学习理论的本质. 清华大学出版社, 2000.

[16] 渡辺澄夫. 代数幾何と学習理論. 森北出版, 2006.



图解机器学习

- **专业实用:** 东京大学教授、机器学习权威专家执笔，浓缩机器学习的关键知识点。
- **图文并茂:** 187张图示帮助理解，详略得当，为读懂大部头开路。
- **角度新颖:** 基于最小二乘法讲解各种有监督学习的回归和分类算法，以及无监督学习算法。
- **实战导向:** 配有可执行的MATLAB程序代码，边学习边实践。

图灵社区: iTuring.cn
热线: (010)51095186转600

分类建议 计算机 / 编程语言与程序设计

人民邮电出版社网址: www.ptpress.com.cn



ISBN 978-7-115-38802-5



9 787115 388025 >

ISBN 978-7-115-38802-5

定价: 49.00元