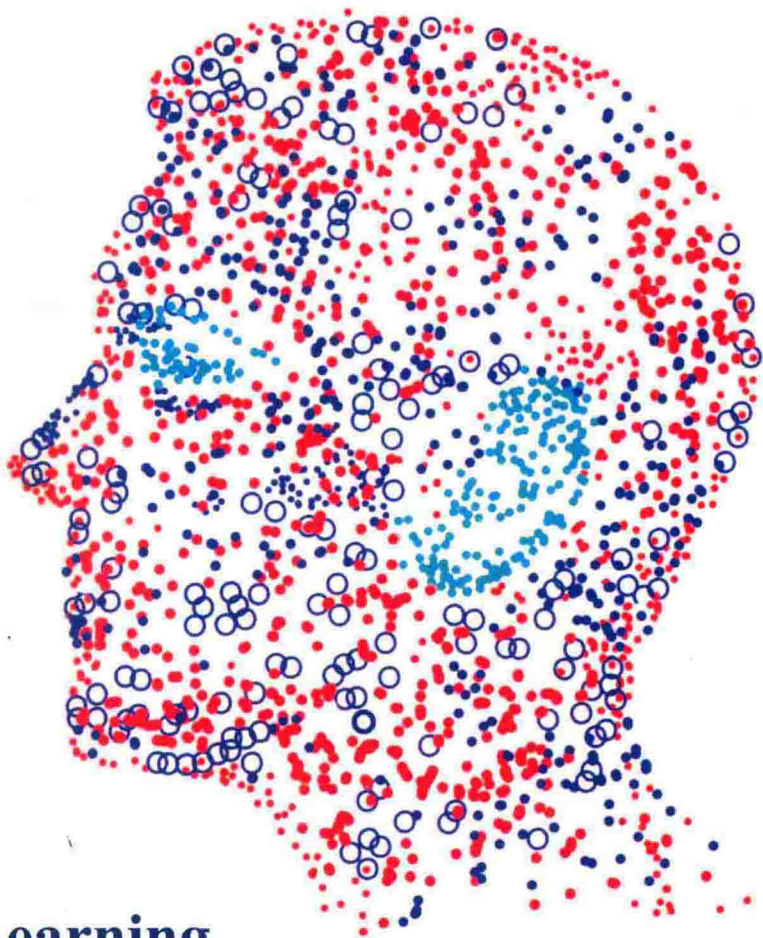


机器学习

[英] Peter Flach 著 段菲 译



Machine Learning

The Art and Science of Algorithms
that Make Sense of Data



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

机器学习

[英] Peter Flach 著 段菲 译



Machine Learning

The Art and Science of Algorithms
that Make Sense of Data

人民邮电出版社
北京

图书在版编目 (C I P) 数据

机器学习 / (英) 弗拉赫 (Flach, P.) 著 ; 段菲译.

— 北京 : 人民邮电出版社, 2016. 1

(图灵程序设计丛书)

ISBN 978-7-115-40577-7

I. ①机… II. ①弗… ②段… III. ①机器学习

IV. ①TP181

中国版本图书馆CIP数据核字(2015)第233670号

内 容 提 要

本书是最全面的机器学习教材之一。书中首先介绍了机器学习的构成要素(任务、模型、特征)和机器学习任务,接着详细分析了逻辑模型(树模型、规则模型)、几何模型(线性模型和基于距离的模型)和概率模型,然后讨论了特征、模型的集成,以及被机器学习研究者称为“实验”的方法。作者不仅使用了已有术语,还引入了一些新的概念,同时提供了大量精选的示例和插图解说。

本书适合具备机器学习理论基础的理工科、信息技术类学生,以及数学、自动化、计算机科学等专业人员阅读。

◆ 著 [英] Peter Flach

译 段 菲

责任编辑 岳新欣

执行编辑 戴 玮 王春青

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: 787 × 1092 1/16

印张: 18.25 彩插 10

字数: 515千字 2016年1月第1版

印数: 1-4 000册 2016年1月北京第1次印刷

著作权合同登记号 图字: 01-2013-3888号

定价: 79.00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

站在巨人的肩上
Standing on Shoulders of Giants



iTuring.cn

版权声明

Machine Learning: The Art and Science of Algorithms that Make Sense of Data first edition (978-1-107-42222-3) by Peter Flach first published by Cambridge University Press 2012.

All rights reserved.

This simplified Chinese edition for the People's Republic of China is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press & Posts & Telecom Press 2016.

This book is in copyright. No reproduction of any part may take place without the written permission of Cambridge University Press and Posts & Telecom Press.

This edition is for sale in the People's Republic of China (excluding Hong Kong SAR, Macao SAR and Taiwan Province) only.

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。

谨以此书献给 Hessel Flach (1923—2006)。

推荐序

人工智能、大数据分析、机器人等领域在近年来日益引人注目，而机器学习则是其中一类非常重要的理论和工具。Peter Flach教授的这部著作可作为机器学习的入门图书，帮助广大迫切希望了解和掌握机器学习的同学和工程师奠定良好的基础。

本书各章节的选题恰到好处，不但对经典机器学习框架中的模型做了非常系统的梳理和分类，涵盖了机器学习基础知识的主要部分，如不同的学习模型、特征、集成学习，而且还对机器学习实验，尤其是机器学习算法的评价(ROC分析)给予了特别的关注，这是十分难能可贵的(在一定程度上弥补了同类教科书的空白)。只要理解了上述内容，相信读者便掌握了机器学习的基本要素，同时有能力进一步就一些更专门、更前沿的主题，如在线学习、主动学习、强化学习、深度学习等，进行更为深入的学习和探索。从这个意义上讲，作者对本书的导论性著作的定位已充分地达成了。

对于如何帮助读者充分理解书中的知识点，作者也予以了充分考虑。书中配有相当数量的图解和实例，旨在突出机器学习理论的直观性。这些内容的数学门槛不高，对工程师和工科学生无疑会有很大的帮助。在各章节中，作者还对每种方法的历史影响进行了介绍，相信会十分有助于提升读者的学习兴趣。作者作为在该领域耕耘多年的研究人员，拥有非常丰富的实践经验，在不少章节中都详尽地分享了实践经验，比如特征和实验这两章。相信这些宝贵的经验会为读者朋友们具体实践机器学习理论带来更好的可操作性。

最后要提到译者段菲的翻译，通篇用语规范、表达准确，总体说来是非常不错的翻译版本。作为计算机视觉领域的优秀研究人员，他本人也在使用机器学习方面有着丰富的研究和实践经验，相信这些都为这本书的翻译质量带来不少保证。

张益民

英特尔(中国)研究院首席研究员

2015年11月10日于北京

序

本书的写作始于2008年夏，当时我就职的布里斯托大学授予我为期一年的研究经费。出于两点考虑，我决定着手编写一部著作，对机器学习进行一般性介绍：一是这样一部著作所涵盖的知识点会对市面上的许多专业书籍起到补充作用；二是通过编写这部著作，我能够获取一些新知识，所谓教学相长嘛。

任何试图编写一部机器学习导论性著作的人都会面临这样一个挑战：如何在公平对待机器学习领域无与伦比的丰富内容时，还能保证其内在原理的统一性。如果过于强调这门学科多样性，则可能使该书最终成为一本“菜谱”式图书而丧失统一性；若过于专注自己感兴趣的领域，则可能会错失许多其他有趣的方向和应用。经过反复斟酌，我确定了本书的基本宗旨，即统一性和多样性“两手都要抓，两手都要硬”。具体表现是通过将任务和特征分开处理（每个机器学习方法都有这个东西，但大家往往不会过多关注它们），以实现统一性；通过覆盖大量逻辑模型、几何模型和概率模型，以实现内容的多样性。

显然，指望在区区三百页的篇幅内深入介绍所有的机器学习内容是不现实的。后记中列出了我最终决定舍弃的，但值得进一步研究的重要领域和方向。在我看来，机器学习是统计学和知识表示“联姻”的产物，因而我在书中主题的取舍上也有意识地强化了这种观点。比如，在介绍与统计学联系更紧密的内容之前，我会用大量篇幅来介绍决策树和规则学习。纵贯全书，我都会特别关注“直观”，希望通过大量例子和图解来帮助读者培养和加强直观理解，其中许多例子都源于我对机器学习中ROC应用相关的研究工作。

如何阅读本书

本书内容是以“线性”方式呈现的，也就是说读者可按章节顺序逐页阅读。然而，这并不意味着你不能随便挑一章进行阅读，因为我在写作时已力图使内容模块化。

例如，对于那些希望尽快了解其第一个学习算法的读者，可直接从介绍两类分类的2.1节开始，然后直接跳转至第5章学习决策树的相关算法，这样在知识连贯性方面不会有什么问题。阅读完5.1节之后，还可以直接跳转到第6章的前两节去学习基于规则的分类器。

或者，对线性模型感兴趣的读者可在学习完2.1节之后，转而阅读3.2节关于回归任务的相

关内容，之后再跳转到第7章学习线性回归。第4~9章关于逻辑模型、几何模型和概率模型的编排次序有一定的逻辑，但这几章的大部分内容都是相互独立的，第10~12章关于特征、模型组合及机器学习实验的相关内容也是如此。

我还要说明的是，绪论和第1章属于导论，且都是自成体系的：尽管绪论中的确包含了一些技术细节，但对于大学预科以上水平的读者，理解起来不会有太大难度；第1章则对本书所覆盖的大部分内容给出了提纲挈领式的介绍。这两部分内容可从本书网站免费下载：www.cs.bris.ac.uk/~flach/mlbook；今后，我还将陆续添加一些其他材料，如讲义幻灯片等。考虑到本书所涉及内容的广泛性，出现一些小的错误在所难免，因此如果你希望了解已有的勘误列表，或提交新的勘误，欢迎你访问上述网站。

致谢

独自撰写一部著作难免孤独，但幸运的是，我得到了许多同事和朋友的热情帮助和鼓励。布里斯托尔的Tim Kovacs、鲁汶的Luc De Raedt以及波士顿的Carla Brodley组织了专门的阅读小组，给予了我极有价值的反馈。我还收到了来自Hendrik Blockeel、Nathalie Japkowicz、Nicolas Lachiche、Martijn van Otterlo、Fabrizio Riguzzi以及Mohak Shah的富有帮助的评价。还有许多人也以各种形式向我提供了帮助，在此我一并向他们表示感谢。

José Hernández-Orallo做了许多本职以外的工作，她不仅仔细阅读了我的手稿，而且还提出了诸多批评和出色的建议，我已一一采纳。José，我一定会找个机会请你吃饭。

感谢我在布里斯托大学的同事和合作者Tarek Abudawood、Rafal Bogacz、Tilo Burghardt、Nello Cristianini、Tijl De Bie、Bruno Golénia、Simon Price、Oliver Ray以及Sebastian Spiegler，感谢他们与我一道工作，并开展了诸多富有启发的讨论。感谢我的国际合作者Johannes Fürnkranz、César Ferri、Thomas Gärtner、José Hernández-Orallo、Nicolas Lachiche、John Lloyd、Edson Matsubara以及Ronaldo Prati，本书中的许多内容都来自或受到我们合作研究的启发。有时，本书需要推进，多亏Kerry、Paul、David、Renée和Trijntje仗义出手，我方得以逃到某个僻静之所从容写作。

剑桥出版社的David Tranah对本书加工工作的启动给予了诸多帮助，封面上隐喻“理解数据”的点彩画正是他的建议（如果你正在琢磨这是谁的画像，那我得解释一下：据David自己说，这就是个普通的剪影画，并非特指某个人）。感谢Mairi Sutherland非常细致的编辑工作。

谨以此书献给先父，若他得知本书已经完成，定会开一瓶香槟来庆祝。他归纳问题的视角虽说有些病态，但也是发人深省的：每天用来喂鸡的那只手最终会将鸡的脖子拧断（这里我要向素食读者表示歉意）。感谢父母为帮助我找到自己的人生道路所提供的一切。

最后，千言万语也不足以表达我对妻子Lisa的感激之情。在我们新婚燕尔之际，我便开始筹划本书的撰写。我们当时谁都没有预料到这本书居然会花费将近四年的时间。后见之明实在太奇妙，因为事后再看，怀疑下面几件事无法同时进行当然是合理的，但事前还是会坚信，我在完成一本书的同时，还能去组织一场国际会议，并且监督房子的大规模翻新。不过，这也见证了Lisa对我的支持、鼓励和默默忍受。所幸这三件事都已圆满成功。Dank je wel, meisje! ^①

Peter Flach
于布里斯托大学

^① 荷兰语，大意为“谢谢你，姑娘！”。——译者注

目 录

绪 论 机器学习概述	1
第1章 机器学习的构成要素	9
1.1 任务：可通过机器学习解决的问题	9
1.1.1 探寻结构	11
1.1.2 性能评价	13
1.2 模型：机器学习的输出	14
1.2.1 几何模型	14
1.2.2 概率模型	17
1.2.3 逻辑模型	22
1.2.4 分组模型与评分模型	26
1.3 特征：机器学习的马达	26
1.3.1 特征的两种用法	28
1.3.2 特征的构造与变换	29
1.3.3 特征之间的交互	32
1.4 总结与展望	33
第2章 两类分类及相关任务	37
2.1 分类	39
2.1.1 分类性能的评价	40
2.1.2 分类性能的可视化	43
2.2 评分与排序	46
2.2.1 排序性能的评价及可视化	48
2.2.2 将排序器转化为分类器	52
2.3 类概率估计	54
2.3.1 类概率估计量	55
2.3.2 将排序器转化为概率估计子	57
2.4 小结与延伸阅读	59

第3章	超越两类分类	61
3.1	处理多类问题	61
3.1.1	多类分类	61
3.1.2	多类得分及概率	65
3.2	回归	68
3.3	无监督学习及描述性学习	70
3.3.1	预测性聚类与描述性聚类	71
3.2.2	其他描述性模型	74
3.4	小结与延伸阅读	76
第4章	概念学习	77
4.1	假设空间	78
4.1.1	最小一般性	79
4.1.2	内部析取	82
4.2	通过假设空间的路径	84
4.2.1	最一般相容假设	86
4.2.2	封闭概念	87
4.3	超越合取概念	88
4.4	可学习性	92
4.5	小结与延伸阅读	94
第5章	树模型	97
5.1	决策树	100
5.2	排序与概率估计树	103
5.3	作为减小方差的树学习方法	110
5.3.1	回归树	110
5.3.2	聚类树	113
5.4	小结与延伸阅读	115
第6章	规则模型	117
6.1	学习有序规则列表	117
6.2	学习无序规则集	124
6.2.1	用于排序和概率估计的规则集	128
6.2.2	深入探究规则重叠	130
6.3	描述性规则学习	131
6.3.1	用于子群发现的规则学习	131
6.3.2	关联规则挖掘	135
6.4	一阶规则学习	139
6.5	小结与延伸阅读	143
第7章	线性模型	145
7.1	最小二乘法	146

7.1.1	多元线性回归	150
7.1.2	正则化回归	153
7.1.3	利用最小二乘回归实现分类	153
7.2	感知机	155
7.3	支持向量机	158
7.4	从线性分类器导出概率	164
7.5	超越线性的核方法	168
7.6	小结与延伸阅读	170
第8章	基于距离的模型	173
8.1	距离测度的多样性	173
8.2	近邻与范例	178
8.3	最近邻分类器	182
8.4	基于距离的聚类	184
8.4.1	K 均值算法	186
8.4.2	K 中心点聚类	187
8.4.3	silhouette	188
8.5	层次聚类	190
8.6	从核函数到距离	194
8.7	小结与延伸阅读	195
第9章	概率模型	197
9.1	正态分布及其几何意义	200
9.2	属性数据的概率模型	205
9.2.1	利用朴素贝叶斯模型实现分类	206
9.2.2	训练朴素贝叶斯模型	209
9.3	通过优化条件似然实现鉴别式学习	211
9.4	含隐变量的概率模型	214
9.4.1	期望最大化算法	215
9.4.2	高斯混合模型	216
9.5	基于压缩的模型	218
9.6	小结与延伸阅读	220
第10章	特征	223
10.1	特征的类型	223
10.1.1	特征上的计算	223
10.1.2	属性特征、有序特征及数量特征	227
10.1.3	结构化特征	228
10.2	特征变换	229
10.2.1	阈值化与离散化	229
10.2.2	归一化与标定	234

10.2.3 特征缺失	239
10.3 特征的构造与选择	240
10.4 小结与延伸阅读	243
第 11 章 模型的集成	245
11.1 Bagging 与随机森林	246
11.2 Boosting	247
11.3 集成学习进阶	250
11.3.1 偏差、方差及裕量	250
11.3.2 其他集成方法	251
11.3.3 元学习	252
11.4 小结与延伸阅读	252
第 12 章 机器学习的实验	255
12.1 度量指标的选择	256
12.2 量指标的获取	258
12.3 如何解释度量指标	260
12.4 小结与延伸阅读	264
后记 路在何方	267
记忆要点	269
参考文献	271

机器学习概述

你也许尚未意识到，自己很可能已经是一名机器学习技术的普通用户了。例如，目前绝大多数电子邮件客户端都采用了识别和滤除垃圾邮件(spam^① e-mail)的算法。早期的垃圾邮件过滤器(spam filter)采用的是基于人工规则(如正则表达式)的模式匹配方法。但人们很快发现，这样的系统不仅难以维护，而且也缺乏灵活性。这很容易理解，因为绝对意义上的“垃圾”邮件是不存在的。有些邮件对某些人来说是“垃圾”，但对其他人来说则可能极有价值。如今，机器学习技术在邮件分类领域已得到了广泛应用，其所带来的自适应性和灵活性远非早期方法可以企及。

SpamAssassin是一款非常流行的开源垃圾邮件过滤器，其工作机制是依据一组内置规则或“测试”(SpamAssassin的专用术语)为每封到来的邮件进行评分。如果某封邮件的得分不低于5分，则SpamAssassin会自动为其添加一个“junk”(即垃圾邮件)标签，并在该邮件标头(email header)中添加一段简要说明。下面展示了我收到的某封邮件的简要说明。

```
-0.1 RCVD_IN_MXRATE_WL      RBL: MXRate recommends allowing
                             [123.45.6.789 listed in sub.mxrate.net]
0.6 HTML_IMAGE_RATIO_02     BODY: HTML has a low ratio of text to image area
1.2 TVD_FW_GRAPHIC_NAME_MID BODY: TVD_FW_GRAPHIC_NAME_MID
0.0 HTML_MESSAGE            BODY: HTML included in message
0.6 HTML_FONT_FACE_BAD      BODY: HTML font face is not a word
1.4 SARE_GIF_ATTACH         FULL: Email has a inline gif
0.1 BOUNCE_MESSAGE          MTA bounce message
0.1 ANY_BOUNCE_MESSAGE      Message is some kind of bounce message
1.4 AWL                      AWL: From:address is in the auto white-list
```

在该图中，从左到右依次为某项测试的评分、测试标识符以及包含对邮件相关部分引用的简短描述。可以看到，不同测试给出的评分可能为负(表明该邮件为普通邮件而非垃圾邮件)也可

^① spam这个词原本是香料火腿(spice ham)的缩写，该肉制品之所以“声名狼藉”，完全要归咎于电视剧《蒙提·派森的飞行马戏团》(Monty Python's Flying Circus)在1970年某一集中的讽刺。

能为正。由于该邮件的综合得分为5.3分，表明该邮件有可能是一封垃圾邮件。这封邮件实际上传达的是来自中介服务器的一条通知，即另外一封邮件（其得分高达14.6分）被认定为垃圾邮件而被拒收。由于该条“退信”消息中包含了原始邮件，故继承了后者的某些特征，如低文本-图像比(text-to-image ratio)，所以它的得分超过了给定的阈值——5分。

下面再举一例。这次所涉及的是一封我本人期盼已久的重要邮件。但不幸的是，我后来在“垃圾箱”中找到了它。

```
2.5 URI_NOVOWEL          URI: URI hostname has long non-vowel sequence
3.1 FROM_DOMAIN_NOVOWEL  From: domain has series of non-vowel letters
```

这封邮件事关我与研究小组的一名同事提交到欧洲机器学习会议(ECML)和欧洲数据库中的知识发现原理与实践会议(PKDD)的一篇文章(自2001年起，这两个会议一直是联合举办)。虽然这两个会议在2008年所使用的域名ecmlpkdd2008.org对每位机器学习研究者来说几乎是耳熟能详，但却因连续出现了11个“非元音字符”而引起了SpamAssassin的怀疑！这个例子表明对于不同的用户，某项SpamAssassin测试的重要性可能不同。机器学习是一种可为用户量身打造程序的绝佳方法。



至此，你一定十分希望了解SpamAssassin确定每项测试分值或“权值”的依据到底是什么。我要告诉你的是，这正是机器学习的用武之地。假设现有一个规模较大的电子邮件“训练集”(training set)，该集合中的所有邮件都带有人工标注的类别标签，即“垃圾邮件”或“普通邮件”，且每封邮件的全部测试结果均已给出。我们的目标是每项测试找到一个恰当的权值，以使所有垃圾邮件的得分均超过5分，而所有普通邮件的得分均低于5分。在本书后面的章节中，你将接触到大量适于解决该问题的机器学习技术。下面我们先通过一个简单的示例来阐述这些方法的主要思想。

例1(线性分类) 假设我们只考虑两项测试。给定的邮件训练集中共有4封邮件，其中含3封普通邮件和1封垃圾邮件，如表1所示。可以看出，对于垃圾邮件(对应表中第1行)，两项测试都是成功的；对其中一封普通邮件(表中第2行)，两项测试均失败；而对其余两封普通邮件(对应表中第3、4行)，两项测试均只有一项成功。显然，将这两项测试的权值全部设为4，便可以对上述4封邮件正确分类。依据“背景知识1”所引入的数学符号，我们可将上述分类器表示为 $4x_1 + 4x_2 > 5$ 或 $(4, 4) \cdot (x_1, x_2) > 5$ 。实际上，采用任意介于2.5和5之间的权值都能够确保仅当两项测试均成功时，邮件的最终得分才高于阈值5。我们甚至可以考虑为两项测试分配不同的权值(只要满足每个权值均小于5，且总和超过5即可)，尽管也许很难从训练数据看出这种做法的合理性。

表1 供SpamAssassin使用的一个规模较小的训练集。 x_1 和 x_2 两列分别对应两项测试的结果，第4列表明哪些邮件为垃圾邮件，而最右列则表明如果将5作为判别函数 $4x_1 + 4x_2$ 的阈值，可轻易地将垃圾邮件分离出来

电子邮件	x_1	x_2	垃圾邮件	$4x_1 + 4x_2$
1	1	1	1	8
2	0	0	0	0
3	1	0	0	4
4	0	1	0	4

背景知识1 用数学语言描述SpamAssassin的工作原理

在本模块中，我们将借助数学语言来描述SpamAssassin的原理。在后面所有类似的模块中，我会对一些重要、有用的概念和符号做出提示。如果你对所涉及的知识不够熟悉，则需要花费一些时间来回顾（既可借助书本，也可参考一些像维基百科或MathWorld这样的在线资源），以便充分理解本书的其余内容。

借助数学语言，可以给出刻画SpamAssassin分类器的多种有用形式。若将对给定邮件的第 i 项测试结果记为 x_i （如果测试成功，则 $x_i = 1$ ，否则 $x_i = 0$ ），并记第 i 项测试的权值为 w_i ，则邮件的总分可表示为 $\sum_{i=1}^n w_i x_i$ 。这里用到这样一个事实：仅当 $x_i = 1$ （即对该封邮件的测试成功）时， w_i 方计入总分。若用 t 表示阈值（即如果某封邮件的总分超过 t ，则将其判为垃圾邮件。本例中 $t = 5$ ），则“决策规则”可表示为 $\sum_{i=1}^n w_i x_i > t$ 。

可以看到，上述不等式左边的部分对变量 x_i 是线性的。这意味着如果 x_i 的增量为 δ ，则总和的增量为一个不依赖于 x_i 的数—— $w_i \delta$ 。若 x_i 在该不等式中以平方项或其他指数不为1的形式出现，则该结论不成立。

借助线性代数的相关知识，我们可对该决策规则的符号做进一步简化。设 $\mathbf{w} = (w_1, \dots, w_n)$ ， $\mathbf{x} = (x_1, \dots, x_n)$ ，则上述不等式可简化为内积形式： $\mathbf{w} \cdot \mathbf{x} > t$ 。将该不等式修改为等式，即可得到用于判决某封邮件是否为垃圾邮件的“决策边界”（或“决策面”）方程： $\mathbf{w} \cdot \mathbf{x} = t$ 。不难看出，由于该方程等号左端为线性项，故在由变量 x_i 所张成的空间中，决策边界为一平面。向量 \mathbf{w} 与该平面垂直，并指向垃圾邮件样本的方向，图1给出了数据点和决策面在2D空间中的可视化结果。

有时可通过引入额外的常量来进一步简化决策边界、决策规则的符号。记 $\mathbf{x}^0 = (1, x_1, \dots, x_n)$ ， $\mathbf{w}^0 = (-t, w_1, \dots, w_n)$ ，即为 \mathbf{x} 引入额外分量1，将对应的权值设为 $-t$ 。这样决策规则即为 $\mathbf{w}^0 \cdot \mathbf{x}^0 > 0$ ，上述决策面方程就可整理为更紧凑的形式： $\mathbf{w}^0 \cdot \mathbf{x}^0 = 0$ 。借助齐次坐标，在新的坐标系中，决策面成为一个经过原点的平面，代价只是增加一个额外的维度（但请注意，这并不会对数据产生任何实际影响，因为所有的数据点和“真实”的决策边界均位于平面 $x_0 = 1$ 中）。

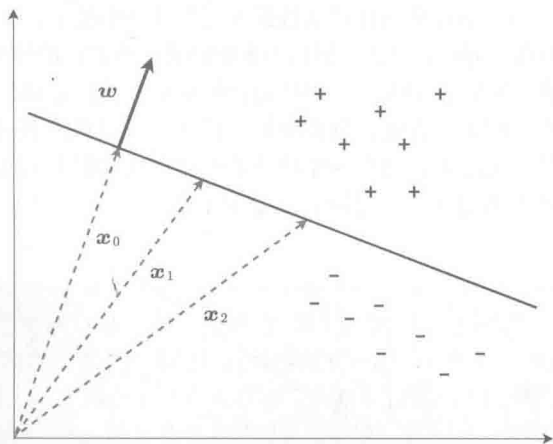


图1 2D空间中的线性分类示例。图中将正例（“+”）和反例（“-”）分离的直线（决策边界）方程为 $\mathbf{w} \cdot \mathbf{x}_i = t$ ，该式中 \mathbf{w} 为一个与决策边界垂直的向量，其方向指向正例， t 为决策阈值，而 \mathbf{x}_i 对应于决策边界上的某一个点。特别地， \mathbf{x}_0 的方向与 \mathbf{w} 一致，故有 $\mathbf{w} \cdot \mathbf{x}_0 = \|\mathbf{w}\| \cdot \|\mathbf{x}_0\| = t$ （ $\|\mathbf{x}\|$ 表示向量 \mathbf{x} 的长度），因此决策边界可表示为 $\mathbf{w} \cdot (\mathbf{x} - \mathbf{x}_0) = 0$ 。该形式在一些场合中使用时尤为方便，特别是这种符号能够更清晰地表达如下重要事实——决定决策边界位置的是 \mathbf{w} 的方向，而非其长度