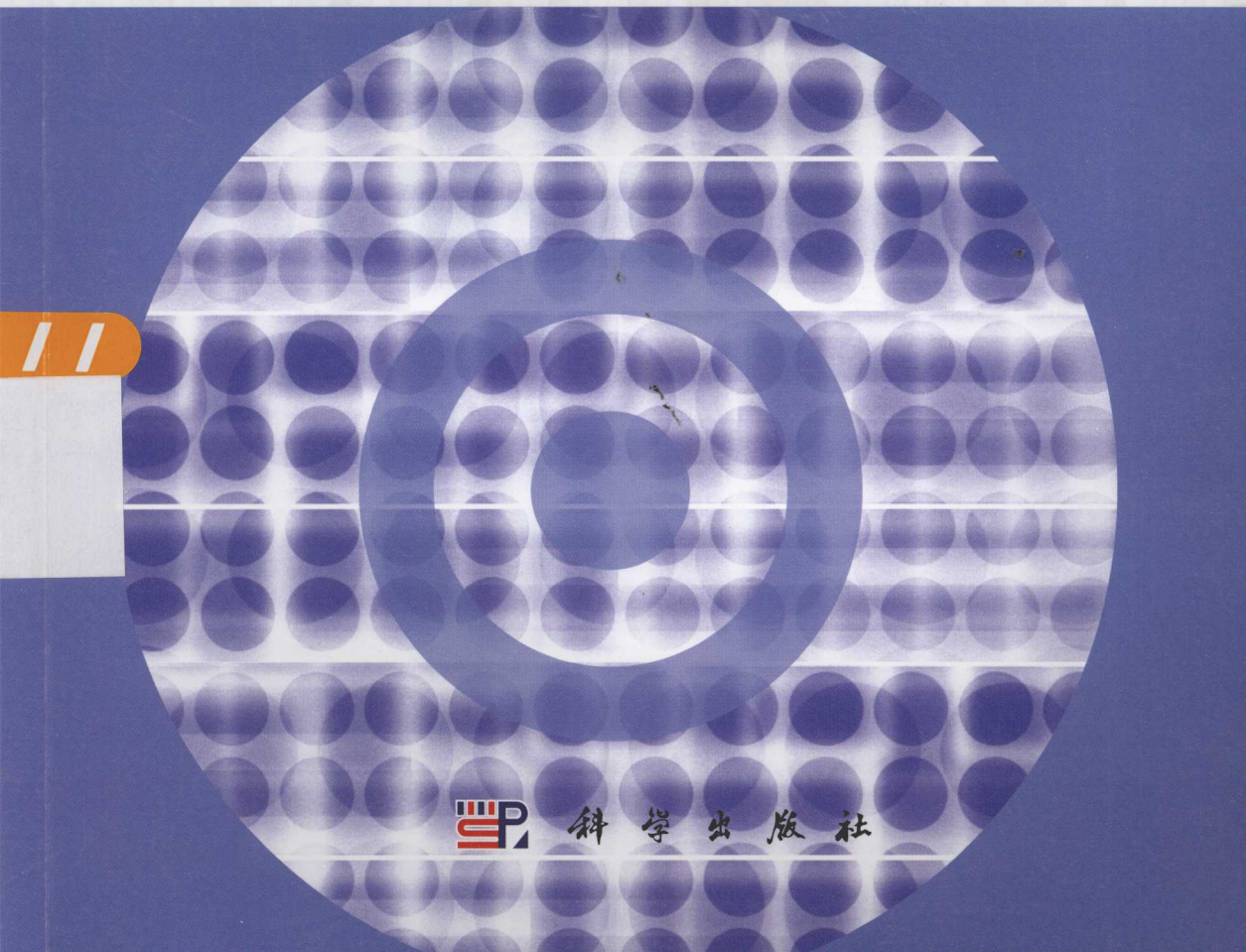
信息科学技术学术著作丛书

支持向量机建模及应用

王文剑　门昌骞 著



科学出版社

信息科学技术学术著作丛书

支持向量机建模及应用

王文剑　门昌骞　著

科学出版社
北京

内 容 简 介

本书在支持向量机学习框架下，通过融合新的理论和机器学习研究成果，系统阐述了支持向量机的建模方法，探索了解决支持向量机的模型选择、效果加速、泛化能力提高、应用范围拓展等问题的新途径。本书从多种角度重新审视支持向量机这一热点机器学习方法，可为解决实际应用问题和改进其他学习方法提供借鉴。

本书可供计算机、自动化及相关专业机器学习领域的研究人员、教师、研究生和工程技术人员参考。

图书在版编目(CIP)数据

支持向量机建模及应用/王文剑，门昌骞著. —北京：科学出版社，2014
(信息科学技术学术著作丛书)

ISBN 978-7-03-040167-0

Ⅰ. 支… Ⅱ. ①王… ②门… Ⅲ. 向量计算机 Ⅳ. TP338

中国版本图书馆 CIP 数据核字(2014)第 047276 号

责任编辑：魏英杰 / 责任校对：彭 涛
责任印制：张 倩 / 封面设计：陈 敬

科学出版社 出版

北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2014 年 3 月第 一 版 开本：720×1000 B5
2014 年 3 月第一次印刷 印张：18
字数：363 000

定价：80.00 元

(如有印装质量问题，我社负责调换)

《信息科学技术学术著作丛书》序

21世纪是信息科学技术发生深刻变革的时代，一场以网络科学、高性能计算和仿真、智能科学、计算思维为特征的信息科学革命正在兴起。信息科学技术正在逐步融入各个应用领域并与生物、纳米、认知等交织在一起，悄然改变着我们的生活方式。信息科学技术已经成为人类社会进步过程中发展最快、交叉渗透性最强、应用面最广的关键技术。

如何进一步推动我国信息科学技术的研究与发展；如何将信息技术发展的新理论、新方法与研究成果转化为社会发展的新动力；如何抓住信息技术深刻发展变革的机遇，提升我国自主创新和可持续发展的能力？这些问题的解答都离不开我国科技工作者和工程技术人员的求索和艰辛付出。为这些科技工作者和工程技术人员提供一个良好的出版环境和平台，将这些科技成就迅速转化为智力成果，将对我国信息科学技术的发展起到重要的推动作用。

《信息科学技术学术著作丛书》是科学出版社在广泛征求专家意见的基础上，经过长期考察、反复论证之后组织出版的。这套丛书旨在传播网络科学和未来网络技术，微电子、光电子和量子信息技术、超级计算机、软件和信息存储技术，数据知识化和基于知识处理的未来信息服务业，低成本信息化和用信息技术提升传统产业，智能与认知科学、生物信息学、社会信息学等前沿交叉科学，信息科学基础理论，信息安全等几个未来信息科学技术重点发展领域的优秀科研成果。丛书力争起点高、内容新、导向性强，具有一定的原创性，体现出科学出版社“高层次、高质量、高水平”的特色和“严肃、严密、严格”的优良作风。

希望这套丛书的出版，能为我国信息科学技术的发展、创新和突破带来一些启迪和帮助。同时，欢迎广大读者提出好的建议，以促进和完善丛书的出版工作。

中国工程院院士
原中国科学院计算技术研究所所长
李国杰

序

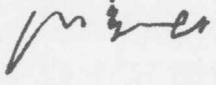
随着传感器技术、存储技术、计算机技术和网络技术的迅猛发展，数据采集和数据传输变得更为便利和快捷，使得数据的膨胀趋势日益加剧，因而信息技术的瓶颈已不在于数据的获取与传输，而在于数据的加工与利用。由于对大规模高维复杂数据分析、处理及管理的迫切需求，甚至产生了数据科学和密集型数据科学等说法。在现实世界中，数据呈现出规模的海量性、表示的高维性、结构的复杂性和时空的动态性等特征，是目前数据处理所面临的最大问题。

在诸多的数据处理方法中，机器学习是最有效的方法之一。现有的机器学习算法在理论上都比较成熟，早期的 PAC 理论和近期的统计学习理论都给出了学习算法的可靠性分析，然而许多优秀的机器学习算法在面向复杂数据的实际应用中并不像在数值实验中那样能表现出良好的性能。

作为公认泛化能力突出的一种机器学习方法，支持向量机具有其他机器学习方法难以比拟的优点，如具有坚实的理论基础、更强的泛化能力、自适应性可有效提供对给定数据集的信息压缩表示等。多年来，许多学者在支持向量机的理论分析和算法实现等方面开展了大量的研究工作，如支持向量机的泛化能力分析、期望误差分析、模型推广、损失函数研究、代数和几何求解方法、预处理方法等。在应用方面，许多学者的研究工作中展示了大量关于支持向量机在生物信息学、计算语言学、计算机视觉等领域的具体应用。目前，已经建立了专门的支持向量机和核学习等网站，许多国际、国内学术会议如机器学习、人工智能、数据挖掘等也将支持向量机列为专题，支持向量机的理论、算法及应用研究一直是相关领域的热点。

该书汇集了王文剑教授、门昌骞博士从事支持向量机研究以来的主要研究成果。书中详细介绍了作者在支持向量机学习框架下，结合其他机器学习技术，对模型选择、效率加速、泛化能力提高、应用范围拓展等一系列重要问题的研究成果，内容丰富、文献翔实，既注重实际应用，又注重理论和方法的正确性。相信该书的出版将对我国支持向量机这一重要的机器学习方法的研究与应用起到积极的推动作用。

中国科学院院士徐宗本



前　言

支持向量机是近年来受到广泛关注的一类机器学习算法，它以统计学习理论(statistical learning theory，SLT)为基础，具有简洁的数学形式、标准快捷的训练方法。与传统的基于经验风险最小化(experience risk minimization，ERM)原则的机器学习算法，如神经网络、决策树等不同，支持向量机是基于结构风险最小化(structure risk minimization)原则，考虑的是经验风险和置信界之和的最小化。因此，相较于基于 ERM 原则的机器学习算法，支持向量机具有更坚实的理论基础、更强的泛化能力，性能也更加优异，因而成为主流的机器学习算法之一。目前支持向量机已被广泛应用于模式识别、函数估计和时间序列预测等数据挖掘问题。

传统的支持向量机学习算法的研究侧重于对方法的探索，对数据集本身的特点关注程度不够。随着机器学习研究的逐步深入，研究者发现脱离实际问题本身的背景去单纯研究学习算法，在实际应用中效果并不是很好，如学习问题本身是一个半监督问题，单纯使用有标记数据进行支持向量机学习并不能达到理想的效果；再如实际问题面向的是海量复杂数据，传统的支持向量机学习算法学习效率与数据规模有关，因此，处理这类问题可能导致算法运行较慢无法求解，在实际应用中可能并没有使用价值。因此，如何从实际问题出发去探索支持向量机算法的建模和改进是一个值得深入研究的问题。除此之外，近年来一些新的机器学习框架已建立起来(典型的如集成学习框架等)，这类学习框架从某种意义上是对机器学习理论的重新认识，因此将支持向量机学习算法纳入到这些框架中也会得到许多新的结论，有较好的推广价值。

本书从多种角度重新审视支持向量机这一热点机器学习方法，以支持向量机为核心、以基于支持向量机的各种新的建模技术为主线，给出了支持向量机研究内容、研究方法和有效应用的新思路，以解决支持向量机研究中的一些问题，如模型选择、算法加速、泛化能力提高、应用范围拓展等，为基于支持向量机的实际应用提供指导和借鉴。全书共 8 章。第 1 章给出支持向量机的基本原理，介绍了支持向量机研究的一些基本概念；第 2 章深入探讨支持向量机应用中重要的模型选择方法，提出一些有效可行的支持向量机模型选择算法；第 3 章探讨如何在具体应用中结合领域知识设计新型的支持向量机学习算法，以有效提高支持向量机的学习能力；第 4 章通过粒度计算思想的引入，改进了传统支持向量机模型，使之更好地模拟人类对现实世界的主观认知，提升学习器的学习效率；第 5 章针对现实问题中遇到的大量半监督问题，给出支持向量机与半监督学习问题融合的方案，可以有效解

决半监督学习问题；第 6 章从集成学习的角度出发，设计集成支持向量机的学习算法，可以有效提高学习器的泛化性能；第 7 章针对大规模高维复杂数据，给出高效的支持向量机改进算法，可以有效降低复杂度，使支持向量机有较高的推广价值；第 8 章在空气质量预测、邮件分类、图像分割实际应用问题中说明支持向量机新模型的有效性。

本书是作者及课题组十余年来在支持向量机理论与算法研究等方面所做工作的总结。在研究工作的开展中，作者有幸承担了国家自然科学基金项目（60975035、60673095、61273291）、教育部博士点专项科研基金项目（20091401110003）、教育部新世纪优秀人才支持计划（NCET-07-0525）、教育部科学技术重点项目（208021）、山西省青年学术带头人支持计划（2007）、山西省自然（青年）科学基金项目（2009011017-2、20041014）、山西省回国留学基金项目（2012-008、2008-14、2003-04）、山西省留学人员科技活动择优资助项目、山西省教育厅科技开发项目、太原市科技明星项目（08121020），并且作为主要成员参与了国家自然科学基金重点项目（71031006）、国家 863 计划项目（2007AA01Z165、2004AA115460）、国家自然科学基金项目（70471003、60275019）等。这些项目的研究成果为本书的编写提供了关键支持。多年来，西安交通大学徐宗本教授、山西大学梁吉业教授、李德玉教授、郑家恒教授等为作者提供了多方面的支持和帮助。研究生郭虎升、王平、郭金铃、侯岩、马亮、张妤、唐超、张鑫、王敏、张文浩、王亚贝、田云等为本书的完成做了许多工作，山东理工大学张瑞为第 5 章的完成做了部分工作，研究生张荣、毕敬业、梁志、程凤伟、潘世超等做了文字校对方面的工作。山西大学计算智能与中文信息处理教育部重点实验室及山西省智能信息处理重点实验室为本书中部分工作的实验提供了数据资源和实验环境，谨在此一并表示感谢。

由于作者水平有限，本书遗漏和不妥之处在所难免，恳请读者批评指正。

作　者

2013 年 4 月

目　录

第 1 章　支持向量机方法

支持向量机(support vector machine,SVM)是 Vapnik 等提出的一类通用有效的机器学习方法,它被广泛地应用于模式识别(分类)、函数估计(回归)、时间序列预测等数据挖掘问题,目前已成为机器学习的研究热点,并在很多领域,如手写数字识别、人脸图像识别、时间序列预测等得到成功的应用。支持向量机本质上是基于统计学习理论的一种机器学习方法。

1.1　统计学习理论

基于数据的机器学习是机器智能研究的重要方面,它研究从数据(样本)中寻找规律,并利用这些规律对新数据或无规则的数据进行预测。迄今为止,关于机器学习还没有公认的理论框架,目前广泛使用的研究方法大致可分为以下三类。

第一类,经典(参数)统计学方法。现有的机器学习的理论基础之一是统计学,参数统计方法是基于传统统计学的机器学习方法。在这种方法中,参数的相关形式是已知的,训练样本用以估计参数值。该方法有很大的局限性:首先,它需要知道已知样本的分布形式,这在大多数应用中是不现实的;其次,传统统计学研究的是样本趋于无穷时的渐近理论(大样本学习),现有的学习方法也大多基于此。对于实际问题,样本的数目往往是有限的。因此,一些理论上很优秀的学习方法却可能表现出很差的实用性能。

第二类,经验非线性方法。这类方法通过对已有的基于传统统计学原理的方法进行修正,或利用启发式方法设计某些巧妙算法,对已知样本建立非线性模型。这类方法克服了传统参数估计方法的困难,可以解决许多实际问题,但该方法缺乏统一的数学理论,表现时好时坏无法控制。

第三类,统计学习理论[1,2]。这是一种专门研究小样本情况下机器学习规律的理论。该理论针对小样本统计问题建立了一套新的理论体系,在这种体系下统计推理规则不仅考虑了对渐近性能的要求,而且追求在现有有限信息条件下达到最优。Vapnik 等从 20 世纪六七十年代开始致力于此方面的研究,直到 90 年代中期,随着学习理论的不断发展和成熟,也由于神经网络等学习方法在理论上缺乏实质性进展,统计学习理论及其在此理论上发展起来的 SVM 便迅速得到人们的重视。

由于统计学习理论为系统研究有限样本情况下的机器学习问题提供了坚实的理论基础,SVM 方法常表现出令人向往的优良特征。越来越多的学者认为,关于

统计学习理论和 SVM 的研究将出现飞速发展的阶段，而且由于其出发点更符合实际情况(有限样本假设)，有理由相信这个研究热潮将会持续升温，并对机器智能研究产生具有深远意义的影响。

统计学习理论由于最初解决模式识别、回归估计等问题时趋于保守且数学上比较艰涩，而且没有能够将其理论用于实践的较好方法，发展到目前在有限样本情况下的机器学习理论研究逐渐成熟，已经形成一个完善的理论体系，使得人们能够很好地理解机器学习的本质。在此基础上所发展的 SVM 方法，在解决小样本、非线性及高维问题中所表现出许多特有的优势，其特点主要表现在以下方面。

① 算法专门针对有限样本设计，其目标是获得现有信息下的最优解，而不是样本趋于无穷时的最优解。

② 算法最终转化为求解一个二次凸规划问题，因而能求得理论上的全局最优解，解决了一些传统方法无法避免的局部极值问题。

③ 算法将实际问题通过非线性变换映射到高维特征空间，在高维特征空间中构造线性最佳逼近来解决原空间中的非线性逼近问题。这一特殊性质保证了学习机器具有良好的泛化能力，同时巧妙地解决了维数灾难问题(特别的是，其算法复杂性与数据维数无关)。

由于具有坚实的理论基础，SVM 在应用中常表现出颇具有竞争力的优势。手写体识别是其最早的应用之一[3~5]，在 MNIST 数据库上，结合了平移不变性的 SVM 的测试误差已经达到 0.56%[6]，与其他方法相比，这一识别率已达到非常高的程度；在文本分类、目标识别、基因分析等生物信息领域的实际应用中，SVM 也取得了极大成功[7~12]；基于 SVM 的回归估计也已成功地应用到许多实际问题中，如波士顿住房问题[13]、时间序列预测问题等[14~16]。在这些应用中，所获得的结果达到或超过了其他方法的水平。另外，由于 SVM 需要调整的参数较少，使用也非常方便。

学习问题是指“依据有限观测，寻找蕴含在数据中的相依性关系”这一科学问题，学习问题的形式化定义如下：

定义 1.1 令 $X\subseteq R^n$ 是一个模式空间，$Y\subseteq R^m$ 是输出空间，$Z=X\times Y$ 是样本空间，X 中的每一个 x 称为是一个模式，$y\in Y$ 称为是一个输出，有序对 $z=(x,y)$ 称为是一个观测值或一个训练数据(更简单地称为一个样本)。假设在模式和输出之间存在一个固定的但未知的关系 F(或者说一个未知的联合概率分布 $F(x,y)=F(x)F(y|x)$)，给定一个函数集合

$$\Im=\{f(x,\alpha):\alpha\in\Lambda,\Lambda\text{ 是一个参数集合}\} \tag{1.1}$$

和一个包含 l 个样本的有限集合

$$Z_0=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\} \tag{1.2}$$

Z_0 和 Z 独立同分布于未知的 $F(x,y)$，则学习问题是指从给定的函数集 $\Im$ 中

选择一个函数 $f^*(x,\alpha^*)$，使得它在一定意义上逼近 Z 上的未知概率分布 F。给定的函数集$\Im$ 称为一个学习机器；Z_0 为训练样本集。如果 Y 取有限的离散值，则学习问题为典型的模式分类问题，否则为回归问题。

为了解决学习问题，一般用风险泛函来度量$\Im$ 中的一个函数逼近 F 的程度，它定义为一个给定模式 x 的真实值 y 和计算值 $f(x,\alpha)$之间损失函数 $L(y,f(x,\alpha))$的期望值，即

$$R(f)=\int L(y,f(x,\alpha))\mathrm{d}F(x,y) \tag{1.3}$$

式中，$R(f)$为实际风险泛函。

因为 $F(x,y)$未知，$R(f)$不能直接计算，所以通常用如下定义的基于训练集构造的经验风险泛函 $R_{\mathrm{emp}}(f)$来逼近，即

$$R_{\mathrm{emp}}(f)=\frac{1}{l}\sum_{i=1}^{l}L(y_i,f(x_i,\alpha)) \tag{1.4}$$

实际风险 $R(f)$也叫做泛化误差的逼近误差，因为它是度量损失函数在整个样本空间 Z 上的均值。对应地，由于经验风险 $R_{\mathrm{emp}}(f)$是度量损失函数在训练样本集 Z_0 上的均值，所以也叫做样本逼近误差或拟合误差。

损失函数的不同表示对应不同的学习问题，如模式识别问题的损失函数可以定义为

$$L(y,f(x,\alpha))=\begin{cases}0, & y=f(x,\alpha)\\ 1, & y\neq f(x,\alpha)\end{cases} \tag{1.5}$$

回归问题的损失函数则可以定义为

$$L(y,f(x,\alpha))=(y-f(x,\alpha))^2 \tag{1.6}$$

通过实际风险泛函 $R(f)$可以形式化地描述学习问题的数学模型，即在 $\Im$ 中寻找函数 f^* 满足

$$R(f^*)=\inf_{f\in\Im}R(f):=\mathrm{opt}_Z(\Im) \tag{1.7}$$

或简单地表示为 $f^*=\arg\inf_{f\in\Im}R(f)$。

一般地，如果 $R_{\mathrm{emp}}(f^*)\leqslant\mathrm{opt}_Z(\Im)+\varepsilon$ 至少以概率 $1-\delta$ 成立，$\varepsilon,\delta\in(0,1)$，则称 f^* 为式(1.7)的一个(ε,δ)解。

学习原理是设计学习算法的归纳推理方法。通常有两种归纳学习原理：一种是众所周知的经验风险极小化(empirical risk minimization，ERM)原理，另一种即是近年来备受广泛关注的结构风险极小化(structural risk minimization，SRM)原理。

1.1.1　经验风险极小化原理

基于 ERM 原理的学习是用经验风险 $R_{\mathrm{emp}}(f)$的极小化来逼近实际风险 $R(f)$的极小化，即在$\Im$ 中寻找一个函数 f_l^* 使得下式成立，即

$$R_{\text{emp}}(f_l^*)=\inf_{f\in\Im}R_{\text{emp}}(f):=\text{opt}_{Z_0}(\Im) \tag{1.8}$$

不同于式(1.7),式(1.8)可以直接求解,因此 $R_{\text{emp}}(f)$就能作为实际风险$R(f)$的替代。很容易看出,只要取不同的损失函数,传统的 BP 神经网络学习算法、回归估计中最小二乘法、密度估计中的极大似然方法等都是 ERM 原理的具体应用。直观上 ERM 原理是有效的当且仅当固有逼近误差,即

$$\varepsilon_{\Im}(l,\delta)=R_{\text{emp}}(f_l^*)-R(f^*)=\text{opt}_{Z_0}(\Im)-\text{opt}_Z(\Im)$$

在 l 趋于无穷时趋近于 0,也就是说,如下二式同时依概率收敛,即

①当 $l\to\infty$时,$R_{\text{emp}}(f)\to R(f)$,即经验风险趋近于实际风险。

②当 $l\to\infty$时,$\text{opt}_{Z_0}(\Im)\to\text{opt}_Z(\Im)$,即样本逼近误差趋近于机器$\Im$ 的逼近误差。

第一个性质与泛函空间的大样本定律有关,第二个性质就是学习理论的一致性问题。Vapnik[1,2]和 Chervonenkis 建立的 ERM 学习的一系列基本定理说明 ERM 原理只对大样本问题或是学习机不太复杂(对应的 VC 维数不大)时才有效。对于现实世界的绝大多数应用问题,训练集的规模是固定的,不可能无穷大,所以一个基于 ERM 原理的学习算法根本不能保证收敛到实际风险的最小值。

1.1.2 结构风险极小化原理

给定一个学习机器$\Im$ 和其中的一个函数 f,记置信风险 $\varepsilon(l,\delta,f)$为

$$\varepsilon(l,\delta,f)=\text{opt}_Z(\Im)-R_{\text{emp}}(f) \tag{1.9}$$

则

$$\text{opt}_Z(\Im)=\text{opt}_{Z_0}(\Im)+\varepsilon_{\Im}(l,\delta)=\inf_{f\in\Im}\{R_{\text{emp}}(f)+\varepsilon(l,\delta,f)\} \tag{1.10}$$

这说明要想极小化实际风险 $R(f)$,必须极小化 $R_{\text{emp}}(f)+\varepsilon(l,\delta,f)$,而不是仅仅极小化 $R_{\text{emp}}(f)$。但这又是不现实的,因为置信风险 $\varepsilon(l,\delta,f)$一般不能直接计算。为此,通常用 $\varepsilon(l,\delta,f)$的一个可计算的粗糙界来代替 $\varepsilon(l,\delta,f)$,即

$$\varepsilon(l,\delta,f)\leqslant\phi(l,\delta,h) \tag{1.11}$$

式中,h 为$\Im$ 的 VC 维;$\phi(l,\delta,h)$是一个置信界。

例如,根据文献[1],$\phi(l,\delta,h)$可以取为

$$\phi(l,\delta,h)=\sqrt{\frac{1}{l}\left[h\left(\ln\frac{2l}{h}+1\right)-\ln\frac{\delta}{4}\right]} \tag{1.12}$$

SRM 原理就是在 $\Im$ 中寻找一个函数 f^* 极小化$\{R_{\text{emp}}(f)+\phi(l,\delta,h)\}$。令

$$R_{\text{upb}}(f)=R_{\text{emp}}(f)+\phi(l,\delta,h) \tag{1.13}$$

式中,$R_{\text{upb}}(f)$为上界风险。

实际上,SRM 原理就是一种极小化上界风险 $R_{\text{upb}}(f)$的方法,它在经验风险和置信界之间(在给定样本的逼近能力和逼近函数的复杂性之间)定义了一种折中。

实际风险、经验风险和置信界与 VC 维 h 的关系如图 1.1所示。SRM 原理对任何规模的样本学习问题都有效，尤其对于固定规模的样本学习问题，它可以得到最佳的可能逼近使得学习机具有最好的泛化能力。当学习机结构固定不变时，ERM 原理可以看做是 SRM 原理的特例。

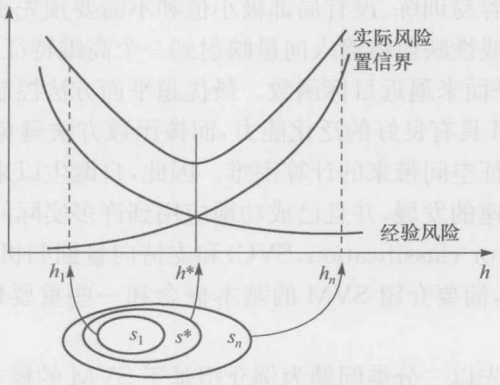


图 1.1　实际风险是经验风险和置信界之和

为了在学习算法中实现 SRM 原理，必须通过控制经验风险 $R_{\text{emp}}(f)$和置信界 $\phi(l,\delta,h)$来优化给定损失函数集 $\mathcal{S}$的上界风险 $R_{\text{upb}}(f)$。有两类方法可以实现 SRM 原理，其一是使置信界保持不变(通过选择学习机合适的拓扑结构实现)，然后极小化经验风险，神经网络就是这种方法的典型代表；其二是固定经验风险的值(如设为 0)，然后极小化置信界。SVM 是第二类方法的典型代表。两类方法得到的学习机的泛化能力与学习机的 VC 维有关，有些学习机的 VC 维可以估计出来(一些典型的神经网络模型的 VC 维估计可参见文献[17]～[23])。学习机的 VC 维很难估计。下面的定理给出了线性分类超平面的 VC 维界估计。

定理 1.1[1]　如果向量 $x\in R^n$ 包含在半径为 R 的超球中，超平面 $w^{\mathrm{T}}x-b=0$ 为 Δ-margin 分离超平面，它对 x 按如下方式分类，即

$$y=\begin{cases}1, & w^{\mathrm{T}}x-b\geqslant\Delta\\ -1, & w^{\mathrm{T}}x-b\leqslant-\Delta\end{cases}$$

则 Δ-margin 分离超平面集合$\mathfrak{I}_\Delta$的 VC 维满足以下不等式，即

$$\mathrm{VC}(\mathfrak{I}_\Delta)\leqslant\min\left\{\left(\frac{R^2}{\Delta^2}\right),n\right\}+1 \tag{1.14}$$

应该注意到，$\left(\frac{R^2}{\Delta^2}\right)$可能比 n 小，所以 $\mathrm{VC}(\mathfrak{I}_\Delta)$可能小于 $n+1$。定理 1.1 说明就复杂性和容量来讲，函数类中的某一个函数可能会比整个函数类简单得多，这也强调了结构选取的重要性。

1.2 支持向量机学习方法

SVM是一种基于统计学习理论的新的、有效的机器学习方法，与神经网络相比具有泛化能力强、容易训练、没有局部极小值和不需要预先确定网络拓扑结构等优点。SVM利用非线性映射将输入向量映射到一个高维特征空间，然后在该空间中构造一个最优超平面来逼近目标函数。最优超平面方法控制学习机所实现的函数类的容量，使SVM具有良好的泛化能力，而核函数方法避免了显式的非线性映射，可以克服高维特征空间带来的计算困难。因此，自诞生以来，SVM受到广泛的重视，理论上得到迅速的发展，并且已成功地应用到许多实际问题中。以支持向量分类机(support vector classification，SVC)和支持向量回归机(support vector regression，SVR)为例，简要介绍SVM的基本概念和一些重要性质，为以后的研究奠定基础。

不失一般性，首先以二分类问题为例介绍基于SVM的模式分类方法。

1.2.1 基本形式

对于给定的训练样本集 $G_0=\{(x_i,y_i):x_i\in R^n,y_i\in\{-1,1\}\}_{i=1}^{l}$，将模式分为：$x_i^+$ 和 x_i^-，目的是寻找一个判别函数 f，使得

$$\operatorname{sgn}(f(x_i))=\begin{cases}+1, & x_i\in x_i^+\\ -1, & x_i\in x_i^-\end{cases} \tag{1.15}$$

即 $\exists\delta>0$，使下式成立

$$y_i f(x_i)\geqslant\delta>0,\quad i=1,2,\cdots,l \tag{1.16}$$

如果存在一个线性函数 f 使得上式成立，则分类问题称为是线性可分的，否则称为是线性不可分的。

SVC在选定的特征空间中构造最优超平面，使分类间隔最大化。这种模型为

$$\begin{aligned}&\min_{w,b}\frac{1}{2}\|w\|_2^2\\ &\text{s.t.}\quad y_i((w,\Phi(x_i))-b)\geqslant 1,\quad i=1,2,\cdots,l\end{aligned} \tag{1.17}$$

进一步，Cortes等提出了C-support vector classification(C-SVC)[24]模型，通过引入松弛变量 ξ 描述分类间隔错误，并利用正则化参数 C 使分类间隔和分类错误达到某种折中。这种模型为

$$\begin{aligned}&\min_{w,b,\xi}\left(\frac{1}{2}\|w\|_2^2+C\sum_{i=1}^{l}\xi_i\right)\\ &\text{s.t.}\quad y_i((w,\Phi(x_i))-b)\geqslant 1-\xi_i\\ &\qquad\xi_i\geqslant 0,\quad i=1,2,\cdots,l\end{aligned} \tag{1.18}$$