江苏高校优势学科建设工程资助项目
国家自然科学基金项目（41271445、40802061）资助
中国博士后科学基金项目（20080441081）资助
中国博士后科学基金特别资助项目（200902534）资助

矿山水害空间数据挖掘与知识发现的支持向量机理论与方法

◎ 闫志刚 著

中国矿业大学出版社
China University of Mining and Technology Press

建设工程资助项目
项目(41271445、40802061)资助
中国博士后科学基金项目(20080441081)资助
中国博士后科学基金特别资助项目(200902534)资助

矿山水害空间数据挖掘与知识发现的支持向量机理论与方法

闫志刚 著

中国矿业大学出版社

内 容 提 要

支持向量机是在统计学习理论基础上发展而来的一种通用学习机器，业已广泛应用于人工智能的各个领域，其在矿山空间数据挖掘与知识发现领域也具有良好的应用前景。为了便于读者阅读和解决实际问题，本书分为理论与应用两大部分，在理论部分对支持向量机的训练参数、核函数及核参数的选择进行了探讨，研究了多类支持向量机的分类问题。在应用部分，将理论部分的研究成果应用于矿井突水水源识别、突水评价与预测、突水数据挖掘与知识发现等领域。主要内容包括支持向量机的参数选择、多类支持向量机的分析模型、多类支持向量机的建模方法、矿井突水水源识别的支持向量机模型、矿井突水知识发现的支持向量机模型、矿井突水预测的粒子群支持向量机模型、矿井水害数据挖掘与知识发现系统等。

本书可供从事空间数据挖掘、矿井水文地质、数据分析、人工智能、决策支持等领域的科技工作者、研究生和本科生参考使用。

图书在版编目(CIP)数据

矿山水害空间数据挖掘与知识发现的支持向量机理论与方法/闫志刚著. —徐州：中国矿业大学出版社，2018.10

ISBN 978-7-5646-2103-2

Ⅰ. ①矿… Ⅱ. ①闫… Ⅲ. ①向量计算机—算法理论—应用—矿山水灾—空间信息系统—数据收集 Ⅳ. ①TD745-39

中国版本图书馆 CIP 数据核字(2013)第255019号

书　　名	矿山水害空间数据挖掘与知识发现的支持向量机理论与方法
著　　者	闫志刚
责任编辑	潘俊成　孙建波
出版发行	中国矿业大学出版社有限责任公司 (江苏省徐州市解放南路　邮编 221008)
营销热线	(0516)83885307　83884995
出版服务	(0516)83885767　83884920
网　　址	http://www.cumtp.com　**E-mail**:cumtpvip@cumtp.com
印　　刷	江苏凤凰数码印务有限公司
开　　本	787×960　1/16　**印张** 9　**字数** 216 千字
版次印次	2018 年 10 月第 1 版　2018 年 10 月第 1 次印刷
定　　价	36.00 元

(图书出现印装质量问题，本社负责调换)

前　言

矿井突水预测是一个涉及水文地质、工程地质、开采条件、岩石力学等诸多因素的复杂问题，仍是当前煤矿生产中亟待解决的重大课题，这已是有关学者的共识。尽管已经有不同的理论、不同领域的学者在矿井突水预测领域做了大量工作，并取得了丰富的成果，但随着科学理论和技术的发展，应用现代科学知识探索矿井突水灾害预测预警的新方法仍具有重要的理论意义和示范作用。

为了解决突水预报的难题，矿山企业联合有关科研单位实施了突水先兆信息的探测研究，积累了时空分布广泛的海量突水监测数据，但却陷入“数据爆炸但突水认识依旧贫乏”的局面，如何有效地利用这些信息，如何从杂乱的数据中发现有效的预测因子、挖掘有价值的突水知识日益成为当前突水预测的瓶颈。目前，各矿山生产单位越来越注重科技的应用，在地质构造探测、突水防治等方面的投入逐年增加，也积累了相当丰富的防治水实例，如何从突水实例中汲取教训，从未突水实例中总结经验，从中发现有用的突水防治知识是当前需要研究的新课题。

1994 年，在加拿大渥太华举行的 GIS 国际学术会议上，李德仁院士首次提出了从空间数据中发现知识的概念，随后他带领的研究团体把发现知识进一步发展为空间数据挖掘，系统地研究了相关的理论、技术和方法。空间数据挖掘旨在解决“空间数据海量而知识贫乏”的瓶颈问题，这对解决当前矿山水害预测的难题提供了新的理论与技术。笔者长期从事矿山灾害监测与空间信息处理的研究工作，得益于工作单位——中国矿业大学在矿业、测绘、地质等学科的优势资源，能够从空间信息技术、矿井水文、采矿工程、人工智能等多视角考虑矿井突水预测预警的难题，积极探索矿山灾害空间数据挖掘与知识发现的理论与技术。通过空间数据挖掘技术可以模拟突水预测由粗到精、由繁到简、由黑到白的认知过程，能交互式地探求突水机理。另外，空间数据挖掘更侧重于从原始数据中发现有价值的、可以理解的突水预测模式，而不是单纯地建立预测模型，这非常符合当前的突水预测需要。

笔者就如何科学地分析突水预测数据，发现有价值的预测知识进行了探索

性的研究，初步建立了以支持向量机为核心的突水信息分析与预测的理论与技术体系，取得了若干创新性成果。在国内外公开发表了相关学术论文 20 多篇，其中被 SCI、Ei、ISTP 等三大检索机构收录 10 余篇。现将这些研究成果进行加工和系统化，汇集成一本较为系统的、可读性强、理论联系实践的著作。

本书共分为 8 章，第 1 章绪论，第 2 章支持向量机的推广能力分析与参数选择；第 3 章多类支持向量机基础，第 4 章多类支持向量机的改进，第 5 章基于支持向量机的矿井水源分析模型，第 6 章矿井突水分析与预测的支持向量机模型，第 7 章矿井突水预测系统的研制与应用，第 8 章为结论与展望。

本书先后获得江苏高校优势学科建设工程资助，国家自然科学基金项目资助（基于领域知识的矿山灾害感知数据时空演变过程的聚类模型及应用，41271445；基于支持向量机和流形学习的矿井突水数据挖掘与预测预警，40802061），中国博士后科学基金项目资助（基于空间数据挖掘与知识发现的矿井突水预测预警，20080441081），中国博士后科学基金特别资助（矿井突水监测信息的特征提取与知识发现，200902534），笔者对以上各方面的支持表示衷心的感谢！

笔者深知，本书所反映的研究工作虽然取得了一定进展，但是对于矿山灾害的监测与信息处理以及空间数据挖掘领域来说，其成果只是“沧海一粟”。尽管数易其稿，字斟句酌，成稿后又请不同学科的多位学者阅读，多次征求意见，集思广益，可是由于研究深度和水平所限，本书只能起到抛砖引玉的作用，书中难免存在疏漏和不足之处，敬请广大读者批评和指正。

希望本书的出版能促进支持向量机在我国各个应用领域的普及，以期能给相关领域的理论研究者和应用工作者提供一些思路和帮助。

著　者

2015 年 10 月

目　录

第1章 绪 论

1.1 研究背景、目的及意义

随着矿山信息化建设的稳步推进，矿山地理信息系统(Mine GIS，MGIS)、数字矿山(Digital Mine，DM)[1]等概念日益深入人心。数字矿山的任务是在矿业信息数据仓库的基础上，充分利用现代空间分析、数据采矿、知识挖掘、虚拟现实、可视化、网络、多媒体和科学计算技术，为矿产资源评估、矿山规划、开拓设计、生产安全和决策管理进行模拟、仿真和过程分析提供新的技术平台和强大工具[2]。其中“矿山数据挖掘与知识发现技术”是数字矿山战略实施的10项关键技术之一[3]。

目前，多数矿山企业建立了自己的矿山地理信息系统，在数字矿山建设稳步推进的同时也积累了大量的矿山各类信息，如何高效利用这些信息服务于矿山安全生产，日益成为亟待解决的问题。本书以矿井突水监测信息的处理为切入点，将最新的机器学习方法——支持向量机(Support Vector Machine，SVM)应用于矿山空间信息处理中，研究矿山数据挖掘与知识发现的支持向量机理论与技术，为数字矿山的信息处理探索新的思路与技术方法。

矿井突水是煤矿水害的主要类型之一，常引发灾害性淹井事故，给国家造成重大的损失，同时也导致人员伤亡。矿井突水预测仍是当前煤矿生产中亟待解决的重大课题，这已是有关学者的共识。尽管已经有不同领域的学者在矿井突水预测领域做了大量工作，并取得了丰富的成果，但随着科学理论和技术的发展，应用现代科学知识探索矿井突水灾害预测预警的新方法仍具有重要的理论意义和示范作用。

本书以矿井突水灾害监测信息作为研究对象，对其加以有效处理与科学分析，将最新的SVM技术应用于矿山水害监测信息的空间数据挖掘与知识发现，以描述矿井突水的认知过程，探求其机理，这不仅对煤矿的安全高效生产具有重要的现实意义，而且可为其他矿山信息的处理开辟新的技术途径和提供范例。

本章的内容安排如下：首先概述 MGIS 的基础知识，然后对矿井突水信息处理的现状进行综述；接着重点介绍 SVM 的理论基础，为以后深入研究 SVM 理论做必要的铺垫；最后给出全书的研究思路与技术路线。

1.2 MGIS 和 DM 概述

MGIS 可以定义为采集、存储、处理、分析、综合利用矿山地质、测绘、采掘、通风、安全、管理等信息的技术系统，具有典型的空间特征。利用 MGIS 处理矿井多源、多时相的时空信息，并加以有效分析与综合利用，为矿山生产提供决策支持是当前矿山信息化建设的主要课题。

在 MGIS 理论研究方面，中国矿业大学做了开创性的工作，郭达志教授最早提出 MGIS 的概念，对其特点和研发技术路线等做过较系统的论述，并创新性地将遥感与 MGIS 技术集成应用于矿产资源开发和矿区资源环境保护等领域[1]；张大顺教授等出版了第一部关于 MGIS 应用的教材[4]；随后毛善君博士对 MGIS 的数据模型进行了深入研究[5]。在三维 MGIS 研究中，毕业于中国矿业大学的陈云浩博士提出了一种适合于矿山的三维数据模型[6]，后来吴立新教授带领的研究小组提出了矿山三维“类三棱柱”数据模型[7]，并使之实用化。另外，文献[8-10]从不同侧面描述了 MGIS 的数据表达。

MGIS 系统的研发始于 20 世纪 80 年代后期，以加拿大、澳大利亚、美国、英国为代表，陆续开发了像 LYNX、MineMAP、MineTEK、MineOFT、MineCAP、VULCAN、GeoQUET 和 DATAMSNE 这样一些代表性的矿山模拟和矿业应用软件系统，并在世界许多矿业大国获得应用。20 世纪 90 年代末，我国在 MGIS 软件研发上，尤其是地测信息系统开发与应用研究取得了重要进展，如北京大学的毛善君开发出了矿山测量图形管理信息系统(MCAD)，煤炭科学研究总院西安分院的萨贤春等开发出了煤矿地测信息系统(MGS)，中国矿业大学的吴立新研发了一套符合中国矿业特色的具有自主版权的 MGIS 基础软件平台(TT-MGIS2000)等。

随着 MGIS 理论研究的日益深入，软件平台的不断成熟，其应用领域越来越广，由最初的地测制图逐步发展到矿山安全、开拓、综采、通风等各个领域，为矿山信息化建设提供了理想平台。

DM 是在 MGIS 基础上的矿山信息化建设的更高阶段。DM 是对真实矿山整体及其相关现象的统一认识与数字化再现，是一个“硅质矿山”，是数字矿区和

数字城市的一个重要组成部分。DM 的核心是在统一的时间坐标和空间框架下，科学合理地组织各类矿山信息，将海量异质的矿山信息资源进行全面、高效和有序的管理和整合。DM 最终表现为矿山的高度信息化、自动化和高效率，以至无人采矿和遥控采矿。

由于矿山空间信息的复杂性、海量性、异质性、不确定性和动态性以及多源、多精度、多时相和多尺度的特点，为了从矿山数据库中快速提取专题信息，发掘隐含规律，认识未知现象和进行时空发展预测等，必须研究高效、智能、透明、符合矿山思维、基于专家知识的空间数据挖掘技术。矿山空间数据挖掘与知识发现即是指从海量的矿山数据中挖掘和发现矿山系统中内在的、有价值的信息、规律和知识的过程。这些信息、规律和知识对矿山的安全、生产、经营与管理能发挥预测和指导作用。

本专著即是在 MGIS、DM 建设稳步推进过程中，其关键技术亟待解决的背景下展开的研究。

1.3 矿井突水预测分析方法综述

矿井突水评价及预报是一个涉及水文地质、工程地质、开采条件、岩石力学等诸多因素的复杂问题，它仍是当前煤矿生产中亟待解决的重大课题。众多学者从不同的侧面提出了一系列的评价预报方法，如斯列萨辽夫公式、突水系数法及“下三带”理论等[11]。但这些方法简化条件多或考虑的因素不够全面，仍未能深刻揭示各种影响因素与突水之间的关系，且在各种不同条件下应用也受到限制。20 世纪 90 年代以来，计算机在煤矿突水预测中的应用广泛开展，对多元突水信息的处理能力日渐增强，新兴的机器学习方法、多元信息处理技术在矿井突水预测中得到了应用。

目前，矿井突水的预测分析方法可分为两类[11]：泛决策分析理论和工程地质力学理论。应该说，工程地质力学分析是从根本上解释矿井突水的原因。最近，在矿井突水机理的研究上也取得了新进展[12-17]，各研究者从不同侧面探寻矿井突水的工程力学模型，试图从根本上解决矿井突水预测的难题。但由于矿井突水原因复杂，对突水机理的解析几乎是不可能的或者是过分简化而不精确，因此，现有矿井突水的工程力学模型普适性差，只适合于特定问题的研究。

泛决策理论以系统论为基础，将矿井突水看作是一个复杂的人—地复合系统，内部各影响因素相互关联并耦合，其作用机理相当复杂，难以满足经典数学

的处理要求。而一些软科学方法如决策论、模糊数学、随机理论、信息理论、专家系统、人工神经网络、非线性科学(非线性动力学分析、突变理论、混沌学)等对处理矿井突水问题非常独到,具有很强的生命力。当前,泛决策理论的分析方法大致有:灰色聚类法[18]、模糊综合评价法[19]、突水概率指数法[20]、模糊层次分析法[21]、人工神经网络法(Artificial Neural Networks, ANN)[22-23]、投影追踪降维法[24]以及非线性理论(突变理论、混沌学等)方法[25];在多元突水信息处理上,出现了专家系统[26]、基于 MGIS 的多元信息复合处理等方法[27]。这些方法在矿井突水预测中均取得了良好的应用效果,但由于影响矿井突水的因素众多,泛决策理论很难建立统一而有效的分析模型,并且由于决策信息的不完整和多噪声从而导致可信度低。

最近,基于机器学习的突水预测方法日益受到关注,代表性的方法有 ANN、SVM[28],与 ANN 比,而 SVM 更适合于小样本的识别问题,在预测精度上被证实一般要优于 ANN。下面对 SVM 的基础理论加以介绍。

1.4 SVM 理论基础

人的智慧中一个很重要的方面是从实例中学习的能力,通过对已知事实的分析总结出规律,预测不能直接观测的事实。在这种学习中,重要的是能够举一反三,即利用学习得到的规律,不但可以较好地解释已知的事例,而且能够对未来的现象做出正确的预测和判断。这就是我们所说的学习的推广能力(泛化能力)。

在人工智能研究中,我们希望机器也能具有上面所说的推广能力,它决定了机器学习能否真正走向实用,在这方面,统计学起着基础性的作用。但是传统的统计学所研究的主要是渐进理论,即当样本趋向于无穷大时的统计特性。在现实的问题中,我们所面对的样本数目通常是有限的,有时还十分有限,在矿山生产建设中的情况更是如此。但人们推导各种学习算法时,仍以样本数无穷多为假设,以为(希望)这样得到的算法在样本较少时也能有较好的(至少是可接受的)表现,但事实并非如此,神经网络中的过学习现象可以说就是一个典型的代表。

近年来,基于统计学习理论(Statistical Learning Theory, SLT)的 SVM 方法越来越引起人们的关注,成为机器学习、模式识别领域的研究热点。

1.4.1 统计学习理论的主要内容

结构风险最小归纳原理是统计学习理论提出的一种运用于小样本学习问题的归纳原理，它包括了学习过程的一致性、边界的理论和结构风险最小化原理等部分。它所提出的结构风险最小化归纳学习过程克服了经验风险最小化的缺点，实用中获得了更好的学习效果。下面就对此理论的一些主要内容进行介绍，主要来自于文献[31-38]。

1.4.1.1 边界理论与 VC 维

边界理论主要包含了两部分：一是非构造性边界的理论，它可以通过基于增长函数的概念获得；二是构造性边界的理论，它的主要问题是运用构造性的概念来估计这些函数。这里，主要研究的是后者。

VC 维，简而言之，它描述了组成学习模型的函数集合的容量，也就是说刻画了此函数集合的学习能力。VC 维越大，函数集合越大，其相应的学习能力就越强。

例如，对于二分类问题而言，h 是运用学习机的函数集合将点集以 2^h 种方法划分为两类的最大的点数目，即：对于每个可能的划分，在此函数集合中均存在一个函数 f_β，使得此函数对其中一个类取 $+1$，而对另外一个类取 -1。如果，取在 R^2（2 维实平面）上的 3 个点（图 1-1）[33]，3 个点分别由“●(R)”，“▲(B)”，“■(P)”这 3 个图形符号来表示（也可以用图形符号旁的英文符号表示它们）。设函数集合$\{f(\alpha,x)\}$为一组“有向线集合”。易知，3 个点最多可以存在 2^3 种划分：(RP,B)、(RB,P)、(PB,R)、(RPB,)、(B,RP)、(P,RB)、(R,PB)、(,RPB)，其中二元组的第 1 项指示的是 $+1$ 类，二元组的第 2 项指示的是 -1 类。对于任意一个划分，我们均可以在函数集合中发现一条有向线对应之，如图 1-1 给出了所有的这 8 种对应。有向线的方向所指示的是 $+1$ 类，反向所指示的是 -1 类。另外，这样的函数集合无法划分 2 维平面中任意 4 个点。所以，函数集合的 VC 维等于 3。

1.4.1.2 推广误差边界

为构造适合于小样本学习的归纳学习机，可以通过控制学习机的推广能力来达到此目的。下面给出此类学习机的推广误差边界。

结构风险最小化的归纳学习过程克服了经验风险最小化的缺点，在实用中获得了较好的学习效果。统计学习理论给出了如下估计真实风险 $R(\alpha)$ 的不等式，即对于任意 $\alpha\in\Gamma$（Γ 是抽象参数集合），以至少 $1-\eta$ 的概率满足以下不

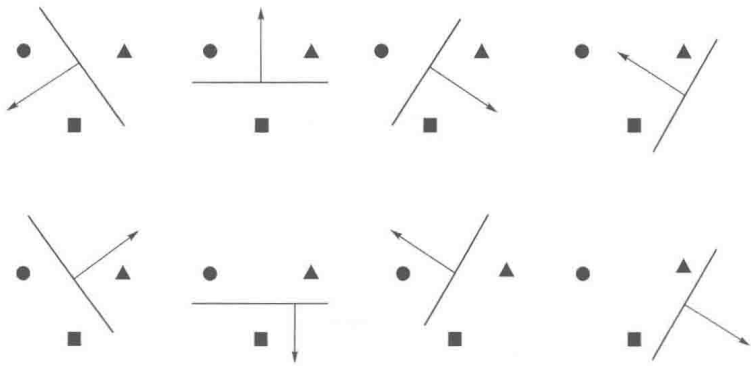

图 1-1　在 2 维平面中被有向线打散的三个点

等式：

$$R(\alpha)\leqslant R_{\text{emp}}(\alpha)+\psi\left(\frac{h}{l}\right) \tag{1-1}$$

其中：

$$\psi\left(\frac{h}{l}\right)\leqslant\sqrt{\frac{h\left(\log\frac{2l}{h}+1\right)-\log\frac{\eta}{4}}{l}} \tag{1-2}$$

α 为学习机的广义参数，$R_{\text{emp}}(\alpha)$ 表示经验风险；$\psi\left(\frac{h}{l}\right)$ 称为置信风险；l 是训练样本个数；参数 h 称为一个函数集合的 VC 维。VC 维是反映学习机学习能力(复杂度)的参数，$\psi\left(\frac{h}{l}\right)$ 随 h 的增加而增加，可见，学习机的推广能力不但与经验风险有关，而且和学习机的复杂性有关。

VC 维很难求得，对于线性分类器，Vapnik 已经证明：

$$h\leqslant\|\omega\|^2R^2+1 \tag{1-3}$$

其中，R 为包络训练数据的最小球半径。

机器学习过程不仅要使经验风险最小，还要使 VC 维尽量小，这样，对未来样本才会有较好的预测能力，这是结构风险最小化准则的基本思想。基于此，Vapnik 提出了结构风险最小化原则(Structural Risk Minimization，SRM)和一种实现它的通用学习算法，即 SVM。

1.4.1.3　结构风险最小化归纳原理

结构风险最小化归纳原理的基本想法是：如果要求风险最小，就需要不等式(1-1)中的两项相互权衡，共同趋于极小；另外，在获得的学习模型经验风险最小

的同时，希望学习模型的推广能力尽可能大，这样就需要 h 值尽可能小，即置信风险尽可能小。

根据风险估计公式(1-1)，如果固定训练样本数目 l 的大小，则，控制风险 $R(\alpha)$ 的参量有 $R_{\text{emp}}(\alpha)$ 与 h。其中：

① 经验风险依赖于学习机所选定的函数 $f(\alpha,x)$，这样，我们可以通过控制 α 来控制经验风险。

② VC 维 h 依赖于学习机所工作的函数集合。为了获得对 h 的控制，可以将函数集合结构化，建立 h 与各函数子结构之间的关系，通过对函数结构的选择来达到控制 VC 维 h 的目的。具体做法如下：

首先，运用以下方法将函数集合 $\{f(\alpha,x),\alpha\in\Gamma\}$ 结构化。考虑函数嵌套子集的集合，如图 1-2 所示(Vapnik,1995)。

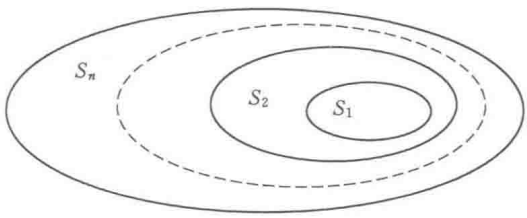

图 1-2 由函数的嵌套子集决定的函数的集合

$$S_1 \subset S_2 \subset \cdots \subset S_k \subset \cdots \subset S_n \cdots \tag{1-4}$$

其中，$S_k=\{f(\alpha,x):\alpha\in\Gamma_k\}$，并且有：

$$S^* = \bigcup_k S_k \tag{1-5}$$

结构 S 中的任何元素 S_k(或一个函数集合)拥有一个有限的 VC 维 h_k，且：

$$h_1 \leqslant h_2 \leqslant, \cdots, \leqslant h_n \cdots \tag{1-6}$$

如果给定一组样本$(x_1, y_1),(x_2, y_2),\cdots,(x_l, y_l)$，结构风险最小化原理在函数子集 S_k 中选择一个函数 $f(x,\alpha_l^k)$ 来最小化经验风险，同时，S_k 确保置信风险是最小的。

以上的思想就称为“结构风险最小化归纳原理”。为了进一步说明，请看图 1-3(Vapnik,1995)。已知一个嵌套的函数子集序列 $S_1,S_2,\cdots,S_n$，它们的 VC 维分别对应为 $h_1,h_2,\cdots,h_n$，而且有 $h_1\leqslant h_2\leqslant,\cdots,\leqslant h_n$。图 1-3 中给出了真实风险、经验风险与置信风险分别与 VC 维 h 的函数变化关系曲线。显然，随着 h 的增加，经验风险 $R_{\text{emp}}(\alpha)$ 递减，这是因为 h 增加，根据 VC 维的定义，对应的函数集合的描述能力增加，学习机的学习能力就增强，可以使有限样本的经验风险很

快地收敛，甚至变为 0；根据式(1-1)，置信风险 $\varphi\left(\frac{h}{l}\right)$ 随着 h 的增加而增加。这样，真实风险 $R(\alpha)$ 是一个凹形曲线。所以，要获得最小的真实风险，就需要折中考虑经验风险与置信风险的取值。

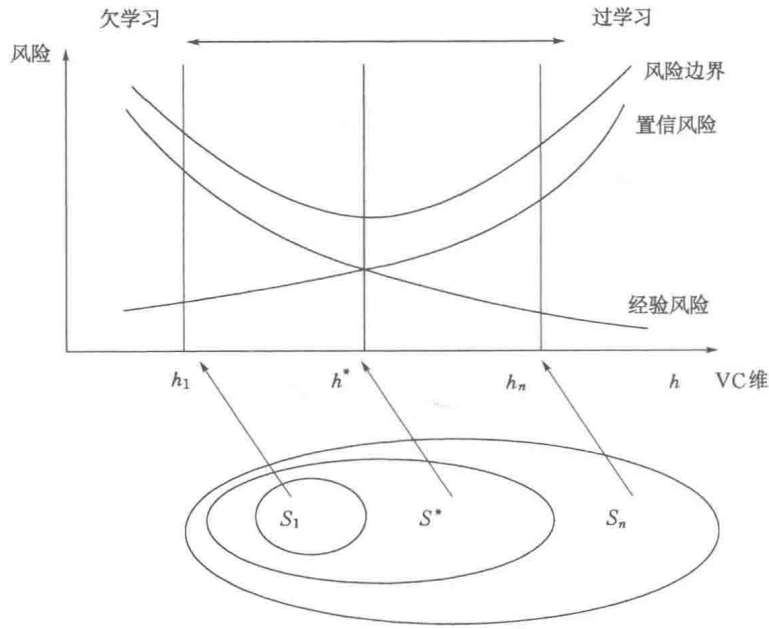

图 1-3　结构风险最小归纳原理图

根据这一分析，可以得到两种运用结构风险最小化归纳原理构造学习机的思路：

① 给定了一个函数集合，按照上面的方法来组织一个嵌套的函数结构，在每个子集中求取最小经验风险，然后选择经验风险与置信风险之和最小的子集。当子集数目较大的时候，此方法较为费时，甚至于不可行。

② 构造函数集合的某种结构，使得在其中的各函数子集均可以取得最小的经验风险(例如，使得训练误差为 0)。然后，在这些子集中选择适当的子集使得置信风险最小，则相应的函数子集中使得经验风险最小的函数就是所求解的最优函数。SVM 采用的就是方法(2)，下面将详细介绍。

1.4.2　支持向量机理论

下面，以 C-SVM(软间隔分类向量机)为例对支持向量机理论加以简介，为

后续问题的展开做必要的铺垫。本部分内容主要来自于文献[31-38]。

对于线性可分的模式识别问题，就是找到一个可计算的识别函数 $y=f(x)$，$x\in R^n$，$y\in\{-1,1\}$，对于给定的 K 个样本$(x_1,y_1),(x_2,y_2),\cdots,(x_k,y_k)$，$x\in R^n$，$y\in\{-1,1\}$，来找到一个可将样本分离的超平面(决策平面)，即 $wx+b=0$，$w\in R^n$，$b\in R$，见图 1-4 所示。通过超平面将样本分为两类，对应的识别函数：

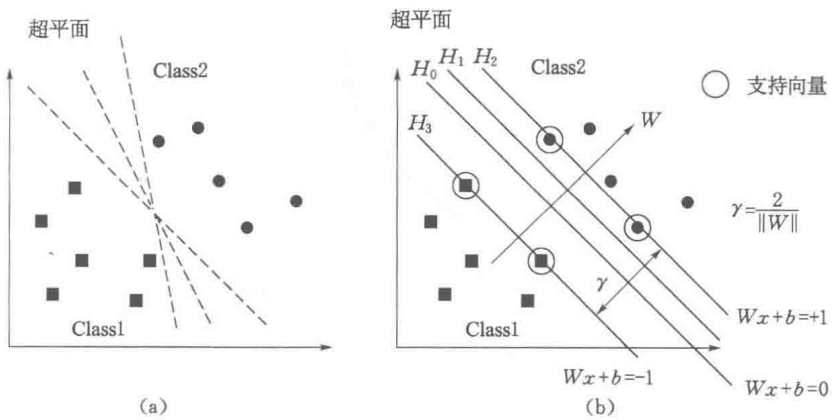

图 1-4 最优分类面示意图

$$f(x)=\mathrm{sign}[(wx)+b] \tag{1-7}$$

决策平面应满足约束：

$$y_i[wx_i+b]>0,i=1,2,\cdots,k \tag{1-8}$$

许多决策平面都可以将两类样本分开，见图 1-4(a)所示，现在的问题是如何找到最优的分类超平面。

假定划分直线的法方向已经给定，如图 1-4(b)所示。直线 H_1 是一条以 W_1 为法向量且能正确划分两类样本的直线。显然这样的直线并不唯一，如果平行地向右上方或左下方推移直线 H_1，直到碰到某类训练样本点。这样，就得到了两条极端直线 H_2 和 H_3，在直线 H_2 和 H_3 之间的平行直线都能正确划分两类样本。显然，在 H_2 和 H_3 中间的那条直线 H_0 为最好。以上给出了在已知法向量 W 的情况下构造划分直线的方法，这样就把问题归结为寻求法向量 W 的问题。

假如此时 H_0 表示为 $w_0x+b_0=0$，因为其在中间，显然 H_2 可以表示为 $w_0x+b_0=k$，H_3 表示为 $w_0x+b_0=-k$，两边同时除以 k，令 $w=\frac{w_0}{k}$，$b=\frac{b_0}{k}$，则 H_0 表示为 $Wx+b=0$，H_2 表示为 $Wx+b=+1$，H_3 表示为 $Wx+b=-1$，这个过程称

为划分直线的规范化过程。此时，两条直线 H_2 和 H_3 之间的间隔为 $2/\|w\|$。对于适当的法向量，会有两条极端的直线，这两条直线之间有间隔，最优分类直线应该是使间隔最大的那个法向量所表示的直线。最优分类超平面应该使两类之间的分类间隔最大，也就是使 $2/\|w\|$ 最大，在求解时，计算 $\frac{1}{2}\|w\|^2$ 的最小值即可。因此可得到下面的最优化问题：

$$\left.\begin{aligned}\min_{w,b}:\tau(w) &= \frac{1}{2}\|w\|^2\\ \text{s.t.}:y_i((w.x_i)+b) &\geqslant 1, i=1,2,\cdots,k\end{aligned}\right\} \tag{1-9}$$

引入拉格朗日乘子 α_i，上式求解方程为：

$$\left.\begin{aligned}\min_{w,b,\alpha}:L(w,b,\alpha) &= \frac{1}{2}\|w\|^2-\sum_{i=1}^{k}\alpha_i(y_i(w.x_i+b)-1)\\ \text{s.t.}:y_i[(w.x_i)+b] &\geqslant 1, i=1,2,\cdots,k\end{aligned}\right\} \tag{1-10}$$

对 w,b 求偏导，得到：

$$\left.\begin{aligned}\sum_{i=1}^{k}\alpha_i y_i &= 0\\ w &= \sum_{i=1}^{k}\alpha_i y_i x_i\end{aligned}\right\} \tag{1-11}$$

将式(1-11)代入(1-10)，得到：

$$Q(\alpha)=\sum_{i=1}^{k}\alpha_i-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j) \tag{1-12}$$

由优化理论中的对偶理论知，最小化式(1-10)等于最大化以约束拉格朗日乘子为变量的式(1-12)，即：

$$\left.\begin{aligned}\max:Q(\alpha) &= \sum_{i=1}^{k}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{k}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)\\ \text{s.t.}:0 &\leqslant \alpha_i, i=1,2,\cdots,k\\ \sum_{i=1}^{k}\alpha_i y_i &= 0\end{aligned}\right\} \tag{1-13}$$

式(1-13)是一个凸二次规划问题，有全局最优解。求解得到最优解 $w=\sum_{i=1}^{k}\alpha_i^* y_i x_i$，取任一 $\alpha_i\neq 0$，可求出 b。在结果中，大部分 α_i 为0，将 α_i 不为0的样本称为支持向量，见图 1-4 所示。