

单位代码: 10183

学 号: 2002621016

吉林大学博士学位论文

徐红敏

分类号 X524 P333.9

基于支持向量机理论的水环境质量预测与评价方法研究

STUDY ON THE PREDICTION AND ASSESSMENT METHODS
OF WATER ENVIRONMENT QUALITY BASED ON SUPPORT
VECTOR MACHINES THEORY

徐红敏

指导教师姓名: 杨天行教授 吉林大学地球探测科学与技术学院

专 业 名 称: 地球探测与信息技术

论文答辩日期: 2007.6.2

授予学位日期: _____

答辩委员会主席: 郭建勋

论文评阅人: 张学林 荣发吉 王秉忱 陈生 马生忠

2007 年 3 月

吉林大学博士学位论文原创性声明

本人郑重声明：所呈交学位论文，是本人在指导教师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：徐红敏

日期： 2007 年 2 月 15 日

提 要

本文针对水环境系统的复杂性、非线性、小样本等特征,以水环境质量预测与评价为研究目标,立足于对支持向量机分类和回归估计的理论模型及其在实际问题研究中的应用进行完善、推广和创新。提出新的或改进的支持向量机多值分类算法和回归估计算法,构建水环境水质预测模型及其质量综合评价模型,并进行应用性的研究与分析。该研究丰富和拓展了支持向量机理论及其应用范围,不仅丰富了水环境质量预测与评价的理论与方法,而且对水资源的合理规划和可持续地开发利用提供了科学的决策依据和技术支持,具有实际应用价值。

本文的主要工作是:

- 1) 对 **SVM** 算法理论的发展历史、研究现状及未来的发展方向,特别是对目前主要研究内容及其应用作研究性综述。通过理论分析与数值实验探讨了各种 **SVM** 算法的优缺点和适用范围。
- 2) 剖析了水环境的水质预测和评价方法的国内外研究现状,阐述了支持向量机应用于水质预测与评价研究的合理性和可行性。
- 3) 拓展 **SVM** 的应用范围,提出改进的 **SVM** 多值分类模型,构建一种新的水库湖泊水质质量综合评价方法。
- 4) 提出改进的 **SVR** 算法,并建立一种新的水质预测模型。

目 录

第一章 绪论	1
§ 1.1 研究背景及意义	1
§ 1.2 目前国内外的研究现状	3
1.2.1 水质预测研究现状	3
1.2.2 水环境质量综合评价研究现状	6
1.2.3 支持向量机算法的发展历史与研究现状	10
§ 1.3 支持向量机在水质预测与综合评价中应用的可行性分析	15
§ 1.4 本文研究的主要内容	16
第二章 统计学习理论与支持向量机概述	19
§ 2.1 统计学习理论发展概述	20
2.1.1 统计学习理论基本概念	20
2.1.2 统计学习理论的发展历史	21
§ 2.2 统计学习理论的基本内容	22
2.2.1 机器学习的基本问题	22
2.2.2 统计学习理论的核心内容	27
§ 2.3 支持向量机概述	31
2.3.1 广义最优分类面	31
2.3.2 线性支持向量机	32
2.3.3 非线性支持向量机	37
2.3.4 核函数	39
2.3.5 标准支持向量回归机	41

§ 2.4 本章小结	44
第三章 支持向量机分类方法的基本理论与算法研究	45
§ 3.1 分类问题的一般描述	46
§ 3.2 C-支持向量机分类算法及其变形算法	47
3.2.1 C-SVM 算法	47
3.2.2 C-SVM 的二次松弛变量算法	49
3.2.3 BSVM 算法	50
3.2.4 无约束的 C-SVM 算法	51
§ 3.3 ν -支持向量机分类算法及其变形算法	52
3.3.1 ν -SVM 算法	52
3.3.2 $B\nu$ -SVM 算法	54
§ 3.4 线性规划形式的支持向量机分类算法	55
§ 3.5 W-支持向量机分类算法	56
3.5.1 加权的 C-SVM 算法	56
3.5.2 FSVN 算法	57
3.5.3 双 ν -SVM 算法	58
§ 3.6 LS-支持向量机分类算法	59
§ 3.7 超球面支持向量机分类算法	60
3.7.1 One-class 支持向量机分类算法	61
3.7.2 超球面二值分类算法	62
§ 3.8 各种改进的支持向量机算法的比较	63
§ 3.9 本章小结	65
第四章 支持向量回归机的基本理论与算法研究	67
§ 4.1 支持向量回归机的基本原理简介	68
4.1.1 基本思想	68

4.1.2	ε – 不敏感损失函数	69
§ 4.2	ε – 支持向量回归机	70
§ 4.3	ν – 支持向量回归机	74
§ 4.4	LS – 支持向量回归机	76
§ 4.5	W – 支持向量回归机	78
§ 4.6	基于线性规划的支持向量回归机	80
§ 4.7	单参数约束下的支持向量回归机	81
4.7.1	单参数回归模型	81
4.7.2	标准与单参数支持向量回归机的约束条件比较	84
4.7.3	算法实现与计算实例	85
§ 4.8	支持向量回归算法中的预测信任度	87
§ 4.9	本章小结	89
第五章	支持向量机多值分类方法研究	91
§ 5.1	基于两值分类的 SVM 多值分类方法	91
5.1.1	“一对多”方法	92
5.1.2	“一对一”方法	93
5.1.3	DDAG 多值分类方法	94
§ 5.2	基于超球面的 SVM 多值分类方法	95
5.2.1	分类算法	95
5.2.2	分解算法	98
5.2.3	数值实验	99
§ 5.3	基于线性规划的 SVM 多值分类方法	100
§ 5.4	二叉树的 SVM 多值分类方法	101
5.4.1	训练样本的选择方案	101
5.4.2	算法实现	102

5.4.3 决策方案	103
5.4.4 数值实验	103
§ 5.5 本章小结	104
第六章 水质预测的 SVR 模型研究	105
§ 6.1 预测学习方法	106
§ 6.2 支持向量回归机函数拟合的基本原理	107
§ 6.3 支持向量回归机的拟合特性分析	108
§ 6.4 改进的函数拟合与预测方法的研究思路	109
6.4.1 改进的加权支持向量回归机	109
6.4.2 算法实现	111
6.4.3 数值实验	111
§ 6.5 水质预测的 SVR 模型研究	112
§ 6.6 应用实例	115
6.6.1 预测模型的参数选择与训练	116
6.6.2 预测结果分析	117
§ 6.7 本章小结	118
第七章 基于 SVM 的水环境质量综合评价方法研究	121
§ 7.1 改进 SVM 多值分类方法的提出	121
7.1.1 训练样本的选择方案	122
7.1.2 算法实现	123
7.1.3 支持向量机的训练方案	123
7.1.4 预抽取两类样本的相对边界向量	124
7.1.5 训练支持向量机的循环迭代方案	125
7.1.6 支持向量机多值分类的决策方案	126
7.1.7 实施步骤	127

§ 7.2 北京市湖泊水库的 SVM 综合评价	128
7.2.1 北京市水资源概况	128
7.2.2 水质评价指标	129
7.2.3 水质评价标准	130
7.2.4 评价方法的选择	132
7.2.5 SVM 综合评价	132
§ 7.3 本章小结	136
第八章 总结与展望	137
§ 8.1 全文总结	137
§ 8.2 存在的问题和研究前景展望	140
参考文献	143
攻读博士期间发表的学术论文及参加的主要科研项目	158
致 谢	159
摘 要	I
ABSTRACT	V

第一章 绪 论

§ 1.1 研究背景及意义

水环境的水质预测与评价是水资源规划和开发利用的重要组成部分。随着人口与经济的增长,世界水资源的需求量不断增加,水环境受到不断恶化,水资源紧缺已成为世人共同关注的全球性问题^[1]。1997 年 1 月,联合国在《对世界淡水资源的全面评价》的报告中指出:缺水问题将严重地制约 21 世纪经济和社会发展,并可能导致国家间的冲突。另外,水污染、水灾害也是困扰许多国家的问题。因此世界范围内对水环境问题的研究越来越重视。

由于我国地理位置的特殊性、地质地貌的复杂性、气候条件的季风性、生态系统的多样性,加之人口多,经济增长快和城市工农业用水的不断迅速增长,使得我国水环境问题更加突出。工业和城市的快速发展,大量废弃物排入江、河、湖、海,给人类赖以生存的水体环境造成了严重的污染^[2]。与世界许多国家相比,我国水环境问题十分严重,特别是改革开放以来国民经济快速发展已成为世界之最,对水资源的需求不断增加,在自然和社会两方面的压力下,出现水环境问题的地区越来越多,而且问题的性质也各不相同,情况复杂。水环境污染问题已经成为环保领域的重要课题,它直接影响到人类的生存与发展。要解决水环境问题,必须树立科学发展观,遵循生态学的一般原理、规律和特点,保护水生态系统的完整性,从而保持生态系统的服务功能,建立经济社会和自然资源的协调关系,实现可持续发展。

水环境质量预测与评价是多个环境要素优劣的综合概念,是一个多因素多水平耦合作用的复杂系统^[3]。从功能上看,环境质量预测与评价是一种辅助和改

善各种社会经济、环境等决策过程的有利工具，其目标是分析人类拟议行动对各种潜在的、显著的环境影响的后果，为确立行动方案的可行性，减轻和消除任何不利影响的对策，使得全球的、国家的、区域的开发建设活动建立在环境协调和可持续发展的基础上。水质预测的目的在于掌握水质现状及其发展趋势，分析判断事故原因及危害性，也为开展水环境水资源质量评价、预测预报提供基础数据和手段^[4]。目前，如何针对水质监测数据特点研究确立较好的数学模型越来越显得更加重要。而水质评价的目的是评价水体污染程度，划分其污染等级，确定其主要污染物。自上世纪 60 年代以来，国内外就不断有文献讨论水质评价的方法^[6-23]，其中最常见、常用的是单项指数法、综合指数法、模糊综合评价法、灰色关联分析法、灰色聚类分析法、人工神经网络算法以及由上述几种数学方法的综合运用而得到的一些综合评价方法。但这些方法中大部分本质上都是通过建立线性模型来实现水质评价的，虽然简单，但是也存在明显的不足，因为影响水质的因素较多，评价因子与标准级别之间的关系是极其复杂和非线性的^[4, 18]。针对线性评价方法的不足，许多学者采用基于软计算的非线性方法进行水质评价，以避免建立显式关系方程确定系数的困难，如采用神经网络(BF)方法^[24-26]。但是由于实际布设的水质监测点的数量有限，导致用来训练神经网络的样本偏少，所以在学习过程中神经网络容易陷入“过学习”和局部极小点，从而降低了泛化能力和水质评价效果的可靠性。

传统的统计学研究的是假定样本数目趋于无穷大时的渐近理论，现有的机器学习方法大多是基于这个假设^[27, 28]。然而在实际的问题中，样本数目往往是有限的。现有的基于传统统计学的学习方法在有限样本的情况下难以取得理想的效果。统计学习理论是在有限样本情况下新建立起来的统计学理论体系，为人们系统地研究小样本情况下机器学习问题提供了有力的理论基础。

支持向量机(Support Vector Machine, SVM)是由 AT & T 贝尔实验室的 Vapnik 及其研究小组于 1995 年在统计学习理论(Statistical Learning Theory, SLT)

的基础上提出来的一类新型的机器学习方法^[29]。它较好地解决了以往困扰很多学习方法的小样本、非线性、过学习、高维数、局部极小点等实际问题，具有很强的推广能力。目前，统计学习理论和支持向量机作为小样本学习的最佳理论，开始受到越来越广泛的重视，正在成为人工智能和机器学习领域新的研究热点。

SVM 是结构风险最小化(Structural Risk Minimization, SRM)准则基本思想的具体实现。为了最小化期望风险，同时做到经验风险最小化 (Empirical Risk Minimization, ERM) 和置信范围的最小化，即以训练误差作为优化问题的约束条件，并且以置信范围值最小化作为优化问题的目标来实现，所以支持向量机的泛化能力 (generalization capacity) 要明显优于神经网络等传统的学习方法。

总而言之，无论在理论上还是在应用前景上，SVM 都具有其它机器学习方法难以比拟的优越性，因此基于可持续发展的理论与方法，对环境变化条件下的水环境质量预测与评价的理论方法及应用进行研究，对于补充完善水环境质量的预测与评价的理论有重要的科学意义和实际应用价值。

水质评价和水质预测问题，是水体分类和水质参数回归的问题，因此新兴的机器学习支持向量机能够解决水质评价和水质预测问题。本文基于这一出发点，详细研究了支持向量机的理论与方法，探讨将支持向量机算法应用于水环境的水质预测及其质量综合评价的可行性，并运用于实际问题的研究，这将对水资源的合理规划和可持续地开发利用提供决策依据和技术支持。

§ 1.2 目前国内外的研究现状

1.2.1 水质预测研究现状

随着我国工业化的发展、城市化进程的加快以及人口的增长，人类赖以生存的淡水越来越短缺，时空分布又不均匀，造成了我国资源性缺水。同时，我国水环境污染状况相当严重，不少地区污染物排放量已明显超过水环境承载能

力,而且区域性水域富营养化情况日益严重,具有明显的季节和地域变化以及年际差异,使得我国又伴随着水质型缺水。因此,水质监测与预测作为水环境水资源管理和污染控制的主要手段之一,目的在于掌握水质现状及其发展趋势,分析判断事故原因、危害,同时为开展水环境水资源质量评价、预测预报提供基础数据和手段,为水资源的合理规划与利用提供科学的决策依据。

对于水质预测,作者的导师及王秉忱先生在八十年代初,在国家科技攻关项目《湘江重金属水环境容量研究》中就有涉及。有学者应用指数平滑法进行了某些尝试,该方法是一种非统计型的方法,在整个预测过程中,始终不断地利用预测误差来纠正新的预测值,预测结果接近于实际,具有较为满意的精度,适用于短期或中一短期的预测。而在其它领域,指数平滑法也得到广泛的应用,例如王季震^[8]等提出了以季节性指数平滑法对城市用水量进行预测模拟,建立了城市用水量预测模型。赵洪宾将它用在了给水管网漏失预测模型的研究,都得到了较好的预测精度。

近年来,国内推广了邓聚龙教授创立的灰色系统理论^[5],其中的GM(1,1)模型是灰色系统理论常用的预测模型,也是灰色系统理论应用中的重要内容,在水质预测中的到了广泛的应用,解决了许多一般方法难以解决的水质预测问题。目前国内也有多人,如王开章^[6]等将灰色系统用于大武水源的水质预测。天津大学的张大海,大连理工的李云贵等人,都对灰色系统理论进行了不同程度的改进,从而提高了该模型的预测精度。吕海涛、赵林明提出了应用灰色模型(GM)时注意的一些问题^[7]。

随着人工智能算法的发展,人工神经网络在水环境的水质预测方面的应用进展迅速。人工神经网络(ANN)是目前国际上的前沿交叉学科,是借助人脑和神经系统存储和处理信息的某些特征而抽象出来的一种数学模型。目前应用最多、研究比较成熟的多层前馈网络反传算法模型(BP模型)。BP模型是一种较特殊的非线性映射方法,它通过一元函数的多次复合来逼近多元函数。在国外,A.

Lapedes 和 R. Farber 最早发表了将神经网络应用于预测的文章。随着神经网络在各个领域的应用,也逐渐用来解决水文地质问题,如 Holger R. Maier 和 Graeme C. Dan^[9]对澳大利亚南部 Murray 河上某段的水质参数矿化度进行了预测,取得了一定成效。河海大学徐立中,王建颖等对水环境监测数据融合处理^[10]。西安邮电学院的蒋林将人工神经网络技术引入了组合预测领域,提出了基于人工神经网络的非线性组合预测模型^[11],在实际应用中取得了良好的效果。

● 指数平滑法预测

指数平滑法预测模型是通过计算平滑值来构建预测模型的一种预测方法,是一种时间序列预测方法。该法假定未来预测值与过去已知数据有一定关系,近期数据对预测值的影响较大,远期数据对预测值的影响较小,影响力呈几何级数减少,是在移动平均法的基础上发展起来的一种趋势分析方法。其实质上是一种以特殊的等比数列为权数的移动平均法,即以本期实际值和上期指数平滑值的加权平均值作为本期指数平滑值,作为下一期的预测值^[12]。

● 灰色系统预测

在水环境系统预测方法中,灰色系统预测模型不依赖于水文地质参数信息,可以在缺乏水文地质参数时,根据已观测地下水资料建立预报因子与影响因素之间的某种近似函数关系即预报模型。

灰色系统是部分信息已知,部分信息未知的系统。地下水系统既有已知信息部分,如有限年代的水位、水质等,更有未知信息部分,如未来年代的水位、水质等,因此它是一种灰色系统。

灰色预测模型以灰色系统理论为基础,以微分方程为描述形式,揭示事物发展的连续过程,符合地下水质量的渐变规律。模型的建立过程即是对地下水系统的辨识过程,运用已知的水质数据去揭示系统的动态特性,但由于其它信息不明确,因而这个过程是灰色的逆过程。

● 人工神经网络预测

人工神经网络在模拟人脑的思维方式下，用于水质的预测。该方法具有较快的收敛速度，预测结果趋于稳定，预测精度高，能够满足于水质的预测。RBFN神经网络应用于水质预测的特点：描述水环境水质的复杂非线性关系；模型建立主要依赖于多年常规观测资料，不需要专门试验和识别参数；模型有很强的学习功能，当系统水环境发生变化时，可操作性强。

总体来说，目前对水质预测数据进行处理，在多源监测数据与水质参数浓度之间建立映射关系(模型)。建立这种映射关系的方法归结起来可分为两类：一是建立显式方法(如：概率法、指数法、多项式回归等)；二是建立隐式方法(如：神经网络方法、遗传算法等)。显式方法大部分本质上都是通过建立线性模型来实现水质预测的，虽然简单，但是也存在明显的不足，利用线性回归方法对水质状况进行估计，精度差，计算量大，没有充分利用数据之间的互补性，处理效果不明显^[18]。作为隐式的BP网络是人工神经网络研究中最重要成果之一，也是迄今为止应用最为广泛的BP网络^[24]。实践证明它具有很强的映射能力，可以解决许多实际问题，但算法本身的缺陷也是无法回避的，它的权重是基于误差梯度下降的原则进行修改的，网络容易隐于局部极小点，算法收敛速度慢，隐层个数以及隐层神经元节点个数难以确定，尽管人们在BP算法和BP网络结构上提出许多改进的方法，但至今尚无一个公认的结果。

1.2.2 水环境质量综合评价研究现状

水质评价是水环境质量综合评价的简称。水环境质量的综合评价是多个环境要素优劣的综合概念，是一个多因素多水平耦合作用的复杂系统^[3]。从功能上看，环境质量的评价是一种辅助和改善各种社会经济、环境等决策过程的有利工具，其目标是分析人类拟议行动对各种潜在的、显著的环境影响的后果，为确立行动方案的可行性，减轻和消除任何不利影响的对策，使得全球的、国家的、区域的开发建设活动建立在环境协调和可持续发展的基础上。

水环境质量评价是指对水环境各个要素进行单项和综合评价。由于水环境质量不仅包括了水体质量，而且涉及影响水体质量的其它因子，因此确定影响水体质量的评价因子，研究其发生、发展的规律及相互关系和质量水平是水环境质量评价的基础^[5]。

水质评价的主要内容是根据水体的用途及水的物理、化学、及生物的性质，按照一定的水质、质量标准和评价方法，将参数数据转化为确定水质状况信息，获得水环境现状及其水质分布状况，对水域的水质或水体质量进行定性或定量的评定。评价水体污染程度，划分其污染等级，确定其主要污染物，准确地指出水体的污染程度，了解掌握主要污染物对水体水质的影响程度以及将来的发展趋势，是进行水资源的保护和综合应用提供原则性的方案和依据的前提。

早期用于水质评价的方法主要有生物学评定分类方法和专家评价法。生物学评定分类方法是 20 世纪初(1902-1909 年)柯克维支和莫松(kolkuitz&Morson)等提出来的^[3]，是从河流受污染的结果来分析，提供水质污染所引起的后果，这种方法可以为河流的水的利用提供依据。专家评价法就是组织环境领域或相关领域专家，运用专业方面的知识和经验对水质进行评价的一种方法，专家评价法的最大特点是对于某些难以定量化的因素给予考虑和作出评价：在缺乏足够的资料的情况下，难以作出定性和定量的可靠判断与估计。进入 20 世纪 60 年代，随着对水质评价研究深入，逐步提出用数学模型来进行水质评价。此后，随着数学理论与方法本身的发展，以及计算机技术的发展，各种各样的新的水质评价方法蓬勃发展，不断涌现。常用水质综合评价方法分析有：

◆ 多项参数综合评价法

多项参数综合评价法即把选用的若干参数综合成一个概括的指数来评价水质，又称指数评价法。常用的指数评价法有布朗水质指数(WQI)、普拉特水质指数(Prati)、罗斯水质指数(Ross)、内梅罗水质指数(Nemerow)、综合污染指数地图叠加法等。它们是用水体各监测项目的监测结果与其评价标准之比作为该项目

的污染分指数，然后通过各种数学手段将各项目的分指数综合得到该水体的污染指数，作为水质评定尺度。

水质评价方法中最为关键的还是各评价因子权重的确定，因为各种污染因子对于环境污染的贡献是不一样的，而且各因子在水环境质量体系中所占的地位(权重)应该是客观存在的，不应与其浓度高低有关。综合污染指数法中各因子做等权处理；内梅罗污染指数法中污染最重因子权重最大，占二分之一，其余因子等权处理；单因子评价法中污染最重因子占 100% 权重，其余因子权重为零；其它评价方法中因子权重由权重公式计算得到。在这些权重的处理方法中基本上有两种：①综合污染指数法是在评价前先确定评价因子的权重。②内梅罗污染指数法和单因子评价法是根据污染因子的浓度来确定其权重。综合污染指数法对各因子做等权处理，不考虑各因子对水环境影响的差异性，可能有局限性。

◆ 模糊综合评价方法与灰色聚类评价方法

水质综合评价中的污染程度、水质类别、分类界限等都是一些客观存在着的模糊现象，具有一定的人为性和不确定性。因此，应用模糊数学方法和灰色聚类方法进行综合评价的相对较多^[14-16]，这两种方法的评价过程完全一致。首先，二者都要对实测数据进行无量纲处理，对处理后的数据，模糊综合评判方法要依据评价标准构造隶属函数，而灰色聚类方法要依据评价标准构造白化函数，这两个函数的构造方式也很相似，通常都采用线性的梯形式、三角形式；其次都要依据污染指标对水环境的影响程度确定其权重，不同的是灰色聚类方法的权重考虑了不同级别，不同指标间的差异；最后是组合运算。隶属函数与白化函数的不同构造方式以及权重的不同确定方法，形成了各种不同的模糊综合评判法及灰色聚类方法。

由于模糊数学与灰色系统自身理论的限制，级别区间内污染值（评价值）的变化特征是无法体现的，而实际上即使在同一级别区间内，其污染值也是有差异的。这就促使人们寻求更为合适的评价方法。

◆ 人工神经网络评价方法

前述的方法都要事先假定模式或主观规定的一些参数。如模糊综合评判法要给定个水质参数的权值，模糊聚类分析要给定隶属函数等，评价结果具有很强的主观性。鉴于这些评价方法存在诸多问题^[18]，近年来国内外都在一直努力探索更为合理的评价方法，其中人工神经网络评价法（BP）被广泛使用，人工神经网络模拟人类思维方式，对事物的判断、分类不需要预先建立某种模式，只根据事物的本质特征、采用直观的推理判断^[17, 18]。因此用人工神经网络进行水质综合评价，不会受某种模式的约束，使评价结果更具有客观性。目前用于水质评价的人工神经网络多为 BP 网络，即误差反向传播网络和 Hopfield 网络。BP 网络通常由输入层、一个隐层和输出层构成。其中输入层神经元节点个数由评价指标个数即污染指标个数决定，输出层神经元节点个数与水质分级标准数目相同，或者综合为一个输出层节点，其中隐层神经元节点数目的确定尚没有一个固定的依据，多由经验或参考其他较为成熟的网络确定。BP 网络是人工神经网络研究中最重要成果之一，也是迄今为止应用最为广泛的网络。实践证明它具有很强的映射能力，可以解决许多实际问题，但算法本身的缺陷也是无法回避的。首先权重是基于误差梯度下降的原则进行修改的，网络容易隐于局部极小点；其次算法收敛速度慢；再次隐层个数以及隐层神经元节点个数难以确定，尽管人们在 BP 算法和 BP 网络结构上提出许多改进的方法^[16, 24]，但至今尚无一个公认的结果。在实际应用中，由于实际布设的水质监测点的数量有限，导致用来训练神经网络的样本偏少，所以在学习过程中易陷入“过学习”和局部极小，从而降低了泛化能力和水质评价结果的可靠性。

许多实际工作者已经注意到，数学模型的有效性并不是由数学模型本身的繁杂程度来决定。有些数学模型模式很复杂，模型本身看来很完善，却忽视了水质评价的特性，难以在实际工作中发挥作用。因此水质综合评价工作应从实际出发，要从简练明了入手，达到评价目的即可，应发展标准化的计算方法和

评价方法，使评价结果有代表性和可比性。

水质综合评价是一个复杂的问题，涉及到多个学科，需要跨学科的人员共同攻关。现有的水质评价方法都有其优点，也有其不足之处。它虽然都得到了一定的应用，仍需对现有的方法不断改进、完善。所以，寻求针对复杂水质综合评价的更好的方法仍然成为广大水文研究者及其它领域科技工作者不断探索和深入研究的目标和方向。

1.2.3 支持向量机算法的发展历史与研究现状

1.2.3.1 支持向量机的理论基础---统计学习理论简介

统计学习理论是一种专门研究小样本情况下机器学习规律的理论^[28]。Vapnik等人从六、七十年代开始致力于统计学习理论的研究^[29-35]。与传统统计学相比，该理论针对小样本统计问题建立了一套新的理论体系，在这种体系下的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的条件下得到最优结果。1963年 Vapnik、Lerner 以及 Chervonenkis 开始研究描述学习理论的非线性普遍算法，他们开创了实现统计机器学习算法的先河^[39]。从此，以 Vapnik 为代表的许多学者对统计机器学习的具体内容不断的丰富和发展。1968-1971 年间，Vapnik 和 Chervonenkis 在《The Necessary and Sufficient Conditions for the Uniforms Convergence of Averages to Expected Values》一文中，提出了一个重要的理论基础---VC 维理论^[30]。利用 VC 熵和 VC 维的概念，发现对指示函数空间的大数定律及其与模式识别问题的关系，并创造出对于经验风险最小化原则下的模式识别的一个一般的非渐近理论。1965-1973 年间，创造了对随机逼近归纳推理的一个一般的渐近学习理论。1974 年，形成结构风险最小化原则 Structural Risk Minimization, SRM)，并且在 1982 年，Vapnik 在《Estimation of Dependences Based on Empirical Data》^[43]一书中首次提出这一具有划时代意义的结构风险最小化原理，堪称为支持向量机(Support Vector Machines, SVM)算法的基石。

1974-1979年,创造基于ERM和SRM原则上的一般的非渐近学习理论。1981年,把大数定律推广到实值函数空间^[31]。1987年,Vidyasaga出版了《A Theory of Learning and Generalization》专著,从数学角度详细讨论了统计学习理论,以及统计学习理论分析神经网络的推广能力。1991年,发现ERM原则与最大似然方法一致的充分必要条件,完成了对经验风险最小化归纳推理的分析。经过30年对学习过程的分析,机器学习中的统计本质已为人们所认识,为统计学习理论的形成完善奠定了基础,对一般学习现象的研究,也积累了一定的算法。到20世纪90年代,统计学习理论已基本成熟。

1995年,Vapnik完成专著《The Nature of Statistical Learning Theory》^[34]。这是统计学习理论走向成熟和得到正式承认的标志。1998年,Vapnik出版的《Statistical Learning Theory》一文^[46],充实了1995年Vapnik论著的理论证明和一些新的成果。与渐近理论形成对照,关于小样本理论(非渐近理论)的讨论广泛地开展起来。

在最近的几年里,在统计学习理论的激发下,在学习方法论上出现了一些新思想。与受学习过程的生物模拟所启发的构造学习算法的旧思想不同,新的思想是受最小化错误率的理论界限的努力激励的,这些界限是作为学习过程的形式化分析的结果得到的。而实用的方法是由对统计学界限深入的理论分析所得到的结果,支持向量机算法可以说是一个典型的成功实例,在解决小样本、非线性和高维的机器学习问题中表现出了许多特有的优势。

1.2.3.2 支持向量机算法的提出和研究现状

统计学习理论的VC维理论和结构风险最小化原则的提出都为支持向量机算法打下了理论基石。1992年,Boser, Guyon和Vapnik等人在《A Training Algorithm for Optimal Margin Classifiers》一书中,提出了最优边界分类器算法^[32],这也是支持向量机算法的最初模型。1993年,Cortes和Vapnik在《The Soft Margin

Classifier》一书中，进一步探讨了非线性情况下的最优边界分类问题^[33]。接着，Vapnik 在 1995 年发表的《The Nature of Statistical Learning Theory》一书中，完整地提出了基于统计学习理论的支持向量机器学习算法。1997 年，Vapnik，Gokowich 和 Smola 发表的《Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing》一文中，详细介绍了基于支持向量机方法的回归估计方法（Support Vector Regression, SVR）和信号处理方法^[35]。

支持向量机成功地解决了高维问题和局部极值问题。支持向量机使用大间隔因子来控制学习机器的训练过程，使其只选择具有最大分类间隔的分类超平面，又叫最优超平面（在不可分情况下，又引入松弛因子来控制经验风险），从而使其在满足分类要求的情况下，又具有最高的推广能力。寻找最优超平面的过程最终转化为二次型优化问题(Quadratic Programming, QP)，从理论上说，得到的是全局最优解。

在处理非线性分类问题上，与传统的学习机器不同的是，支持向量机是将输入空间映射到高维的特征空间，仍然使用大间隔因子在高维特征空间中寻找最大间隔超平面。事实上，高维特征空间中的超平面对应着输入空间中的非线性分类面。实际中，支持向量机的优化过程并没有真正在高维特征空间中进行，而是通过一些具有特殊性质的核函数^[52, 56, 57]，将高维特征空间中的内积运算转化为原始空间中核函数的运算，从而巧妙地避免了在高维特征空间中处理问题的困难。

作为统计学习理论的 VC 维理论和结构风险最小化原则的具体实现算法，支持向量机集优化、核、最佳推广能力等特点于一身。支持向量机算法的几个主要优点有：

- 1.它是专门针对有限样本情况的，其目标是得到现有信息下的最优解而不仅仅是样本数趋于无穷大时的最优值；
- 2.算法最终将转化成为一个二次型寻优问题，从理论上说，得到的将是全局

最优解，解决了在神经网络方法中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转换到高维的特征空间(Feature Space)，在高维特征空间中构造线性判别函数来实现原始空间中的非线性判别函数，特殊性质的核函数能保证机器有较好的推广能力，同时也巧妙地解决了高维问题，其算法的复杂度与样本维数无太大相关；

4.由于其优化目标是结构风险最小化，同时考虑了经验风险和置信范围的最小化，因而支持向量机具有非常好的推广能力。

◆ 支持向量机的理论研究

虽然支持向量机发展时间较短，但是由于它的产生是基于统计学习理论的，因此具有坚实的理论基础。近几年涌现出的大量理论研究成果，更为其应用研究奠定了坚实基础。如 Anthony 等人^[52]给出了关于硬邻域支持向量机学习误差的严格理论界限；Shawe-Taylor 等人^[53]也给出了类似的关于软邻域支持向量机和回归情况下的误差界限；Weston^[55]和 Vapnik^[38]等研究了支持向量机的推广性能及其在多值分类和回归问题的扩展问题；Smola 和 Schölkopf^[55]提出了支持向量机一般意义下的损失函数数学描述。

随着支持向量机在理论上的深入研究，出现了许多变种支持向量机。Vapnik 提出的可调参数 C 的 SVM 统称为 C -SVM 系列^[38, 50, 56, 57]。Schölkopf 提出的用于分类和回归的 ν -SVM 系列^[53, 55]，引入反映超出 ϵ 管道的样本数据点个数（即边界支持向量数量）和支持向量数的新参数 ν ，从而简化 SVM 的参数调节。另外，一些学者还扩展了支持向量机概念，如 Mangasarian^[100]等人的通用支持向量机（Generalized SVMs）。支持向量机通过最优超平面，将样本划分为互相对立的两个类。然而在实际应用中，在某些情况下，每个样本并不是完全能够划归到某一类，即样本与类别之间存在着某种模糊的隶属关系，由此，Takuya 和 Lin 等人提出了一种模糊支持向量机^[111]，通过给样本增加一个模糊隶属关系来充分地利用样本的信息。1999 年，Suykens 提出了一种新型支持向量机---最小二乘支

持向量机^[87]。最小二乘支持向量机引入最小二乘线性系统到支持向量机中，代替传统的支持向量机采用二次规划方法解决函数估计问题。在此基础上提出的周期最小二乘支持向量机已用于时间序列的预测^[67]。Chew^[116]提出的加权支持向量机可以解决普通支持向量机不能根据每个采样点数据的重要性区别对待的缺陷，还可以解决类别补偿问题。最有意思的是，One-class SVM 算法，它解决的是只有正训练样本一个类别的分类问题，Schölkopf 最初是用它来对高维分布进行估计^[94]，而 Tax^[118, 119]通过用超球面代替超平面来划分样本，用球来包含所有的无类别信息的样本，从而解决只有一类训练样本的分类问题。另外简化支持向量机^[73, 74]、小波向量机^[85]、Lagrangian 支持向量机^[86]以及其他一些新的 SVM 模型^[89-108, 144-161]。

◆ 支持向量机的应用研究

统计学习理论从七十年代末诞生，到九十年代之前都处在初级研究和理论准备阶段，近几年才逐渐得到重视，其本身也趋向完善，并产生了支持向量机这一将这种理论付诸实现的有效的机器学习方法。目前，SVM 算法在模式识别、回归估计、概率密度函数估计等方面都有应用。例如，在模式识别方面，对于手写字识别^[22, 23, 94-96]、语音识别^[97-101]、文本分类^[24, 102, 103]、人脸图像识别^[63, 104-108]、遥感图像分析^[109, 110]等问题，SVM 算法在精度上已经超过传统的学习算法或与之不相上下。

支持向量机在工业领域的应用研究正逐渐受到研究者的重视。文^[111, 112]将支持向量机用于系统辨识，进行线性和非线性动态系统的辨识。De Kruijff^[113]将支持向量机用于前馈学习控制，Suykens^[87]将最小二乘支持向量机应用于非线性系统的最优控制。文^[115-117]将支持向量机用于工业故障诊断。

支持向量机还被应用到其它一些领域。支持向量机在时间序列的预测和混沌系统的动态重构中显示出了强大的优势^[118, 43]。在信号处理方面，Chen^[146]对多路径通道 DS-CDMA 信号传输构造了一个基于支持向量机的自适应多用户检测

器。

§ 1.3 支持向量机在水质预测与综合评价中应用的可行性分析

用于分类与回归的方法很多，如传统的统计分析方法以及神经网络方法^[11]等。这些方法虽然在实际应用中占据主导地位，但人们也发现它们还存在着许多不足之处^[4, 7, 18, 25]。比如，传统的统计方法一般需要事先知道样本的先验分布，并要求有足够多的样本数据，而这些要求在实际应用中往往难以达到，这就使其在实际应用中效果往往并不理想。神经网络方法虽然很好地解决了非线性问题，但由于其自身存在着结构不易确定、易陷入局部极小等固有的缺陷，从而限制了其实际应用。

依据水质标准进行水环境质量评价，实质上可以看作是一个水质分类的模式识别问题；水质的预测是依据现有的水质监测数据预测水质的变化趋势。但是在实际水质数据监测与评价中，难以布设足够多的水质监测点以满足网络训练数据量的要求，导致用来训练神经网络的样本偏少，所以在学习过程中神经网络容易陷入“过学习”和局部极小点，从而降低了泛化能力和水质监测和评价效果的可靠性^[24]。另外，神经网络的学习算法仅仅试图使经验风险最小化，并没有使期望风险最小化，与传统的最小二乘法相比，在原理上缺乏实质性的突破，这也是神经网络过拟合现象产生的原因，从而导致了其推广能力的下降。

水质评价和水质预测问题是水体分类和水质参数回归的问题，其恰恰对应于支持向量机多值分类和回归估计算法。水质参数的样本是有限样本、非线性、多维问题。因此新兴的机器学习支持向量机能够解决水质评价和水质预测问题。

Vapnik^[46]等人提出的统计学习理论是一种针对小样本情况研究统计学习规律的理论，该理论的核心思想是通过引入结构风险最小化准则，控制学习机器的容量，从而刻画了过度拟合与泛化能力之间的关系。支持向量机不仅有完善的理论基础，而且其应用领域也在不断扩大。

作为一门新兴的学科，有关支持向量机方面的很多研究在国外也是刚刚起步，国内的有关研究只是在近几年才逐渐引起人们的注意。与传统方法相比，支持向量机所具有的优势体现在，既克服了传统方法的大样本要求，还有效地克服了维数灾难及局部极小问题，并在处理非线性问题时计算简便、误差较小，显示了其卓越的优越性。它非常适合水环境的水质预测与质量综合评价的要求。

支持向量机的主要思想可以概括为两点：

1) 它开始是针对线性可分情况进行分析的，后来对于线性不可分的情况，通过使用非线性映射算法将低维输入空间线性不可分的样本映射到高维属性空间使其线性可分，使得在高维属性空间采用线性算法对样本的非线性特性进行分析成为可能；

2) 它通过使用 **SRM** 准则在属性空间构造最优分割超平面，使得机器学习得到全局最优化，解决了“过学习”问题，对样本具有较好的泛化能力^[27]。

另外，由于支持向量机的训练问题本质上是一个经典的二次规划问题，避免了局部最优解，有效地克服了维数灾难，而且可以利用最优化理论中许多成熟的算法。**SVM** 算法有很多成功的应用，这些都说明这种基于 **VC** 维理论而发展起来的结构化学习方法所具有的潜在优势。更重要的是：作为 **SVM** 算法基础的 **VC** 维理论和结构最小化原则也为进一步完善传统的统计预测方法和经验非线性预测方法提供了理论基础和统一的理论框架^[27-59]。

本文正是在这一理论背景下，在深入探讨与研究支持向量机理论基础上，在二次规划和线性规划的框架下对分类与回归中的基本方法进行研究，研究与分析分类与回归之间的关系，并构造新的分类及回归方法。主要是建立满足水质预测与水环境综合评价的客观要求的数学模型及其求解方法，发挥 **SVM** 的优势，探讨支持向量机在时间序列预测和分类评价中的应用。

§ 1.4 本文研究的主要内容

支持向量机(Support Vector Machines, SVM)是复杂非线性科学和人工智能科学的研究前沿,其在水资源的水质预测与评价领域的应用研究在国内外尚处于初创阶段。本文在全面分析评述了水质预测与评价研究现状,在分析阐述了 SVM 的基本原理与算法基础上,在国内首次将 SVM 方法引入水环境质量预测及其质量综合评价中的应用领域,主要在水质预测和水质评价的 SVM 建模方法以及算法方面进行了一些创新性的研究工作,为提高水质评价和水质模拟的智能化水平做出了努力。

全文共分八章,具体内容安排如下:

第一章阐述了选题的意义。在对目前关于水环境质量预测及其评价的研究文献进行系统总结的基础上,进一步评述了:①目前国内外的水环境质量的预测及其评价的研究现状;②概述了支持向量机算法的发展历史与研究现状;③支持向量机理论与方法在水环境质量的预测与评价中的可行性与合理性。在此基础上,形成了本文的总体研究思路。

第二章研究分析了支持向量机的理论基础——统计学习理论的若干原理,并对支持向量机算法的发展历史和现状及主要研究内容作了详细的论述。通过对统计学习理论进行回顾,概述统计学习理论研究的基本问题及主要内容。重点探讨 SVM 的基本概念、理论与方法以及标准支持向量机的基本模型。

第三章详细讨论的是两类问题的支持向量机算法及其变形算法。通过对各种支持向量机分类算法进行讨论与研究,比较分析它们的特点及其适用范围,以期对这些方法有一个客观、全面地认识,为构造基于支持向量机分类理论的水环境质量综合评价模型做好理论准备。

第四章重点阐述用于曲线拟合的支持向量回归机(SVR)学习算法,以及其他几种在优化式上加以改进的用于回归估计的 SVR 优化算法;同时,建立了在 ε -不敏感函数下的单参数约束下的支持向量回归模型,并进行数值实验;最后探

讨并建立支持向量回归算法优劣的评价依据，从而丰富了支持向量回归机的优化理论基础和决策依据。

第五章支持向量机多值分类法的研究，全面总结了目前存在的基于支持向量机的多类别分类方法，包括“一对多”方法、“一对一”方法、线性规划的支持向量机多值分类方法、超球面多值分类算法、二叉树的支持向量多值分类等若干方法，比较了它们的优缺点及性能；并通过数值实验对其中的几种常用的方法进行了验证与比较。

第六章在综合研究 SVR 算法的基础上，针对水环境中污染物浓度的监测的问题，将水质参数平均浓度值的监测看成一个时间序列预测问题，建立了基于支持向量机的水质监测模型。然后，利用支持向量回归机对实验数据进行回归分析，并与基于 BP 神经网络法的水质预测模型作了比较，最后给出了结论分析。

第七章在进行了深入的研究与分析的基础上，根据水质评价的原则与要求，结合预抽取支持向量以及循环迭代算法，提出了一种改进的循环决策二叉树的支持向量机多值分类方法，成功地克服了其他方法存在的问题及缺点，提高支持向量机的训练和决策速度，并且解决了目前支持向量机多值分类方法的不可分区域的分类问题，取得了较为理想的效果。运用改进的循环决策二叉树的支持向量机多值分类方法，建立了北京市湖泊水库的水质评价模型，并对其结果进行了比较分析。

第八章全文总结。对全文进行概括性总结，并对本文工作中存在的问题和有待进一步研究的内容进行归纳和展望。

第二章 统计学习理论与支持向量机概述

机器学习是现代智能技术的重要方面，主要研究如何从一些观测数据(样本)出发寻找目前尚不能通过原理分析得到的规律，利用这些规律去分析客观现象，对未来数据或无法观测的数据进行预测。现实世界中确实存在大量的人类尚无法准确认识但却可以进行观测的事物，因此机器学习在现代科学技术、社会、经济等各领域有着十分重要的应用。将研究的规律抽象成分类关系时，这种机器学习问题就演变成分类问题，或称模式识别；而将研究的规律抽象成预测或估计问题时，就称为回归问题。包括模式识别、神经网络等在内，现有机器学习方法共同的重要理论基础之一是统计学。传统统计学研究的是样本数目趋于无穷大时的渐近理论，现有学习方法也多是基于此假设。但在实际问题中，样本数目往往是有限的，因此一些理论上很优秀的学习方法实际中表现却可能不尽人意^[46]。

统计学习理论(Statistical Learning Theory, 简称 SLT)是一种专门研究小样本情况下机器学习规律的理论^[46]。V.Vapnik 等^[39]人早在 20 世纪 60 年代就已提出了统计学习理论，它是支持向量机的理论基础，而 90 年代以前并没有提出能够将其理论用于实践的较好的方法。加之当时正处于其他学习方法飞速发展的时期，因此，这些研究一直没有得到充分的重视。直到 90 年代中期，有限样本情况下的机器学习理论研究逐渐成熟起来，它目前形成了一个较完善的理论体系。

在统计学习理论的基础上发展而来的新的机器学习方法---支持向量机(Support Vector Machine, 简称 SVM)，在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势，并能够推广应用到函数拟合等其他机器学习

问题中。**SVM** 以统计学习理论为基础，建立在计算学习理论的结构风险最小化原则之上，具有简洁的数学形式，能进行直观的几何解释，并具有良好的泛化能力，避免了局部最优解，且需要人为设定的参数少，便于使用，为小样本机器学习提供了一种新方法，现已成为新的研究热点，并成功应用于模式识别^[66]、回归分析和函数逼近等领域^[53]。

本章首先对统计学习理论进行回顾，概述统计学习理论研究的基本问题及主要内容。重点探讨 **SVM** 的基本概念、理论与方法以及标准支持向量机的基本模型。

§ 2.1 统计学习理论发展概述^[29]

2.1.1 统计学习理论基本概念

统计方法是从观测自然现象或者专门安排的实验所得到的数据去推断该事物内在的规律性。科学的规律性一般总是隐藏得比较深，最初总是从其在数量表现上通过统计分析看出一些线索，然后提出一定的假说，作进一步深入的研究。当理论研究提出一定的结论时，就需要在实践中加以检验。统计学习理论是在研究小样本统计估计和预测的过程中发展起来的一种新兴理论，它的发展及其在随后的应用中涉及到一些基本的概念如下：

(1) 学习——如果一个系统能够通过执行某一过程而改进它的性能，我们就称之为学习。

(2) 机器学习——主要研究从采集样本出发，得出目前尚不能通过进行原理分析得到的规律，并利用这些规律对未来数据或无法观测的数据进行预测。它是现代智能技术中十分重要的一个方面。

(3) 模式——通常我们把通过对具体的个别事物进行观测所得到的具有时间和空间分布的信息称为模式，而把所属的类别或同一类中模式的总体称为模式类(或简称为类)。

(4) 模式识别——对表征事物或现象的各种形式(数值, 文字和逻辑关系)信息进行处理和分析, 以对事物或现象进行描述、辨认、分类和解释的过程, 是信息科学和人工智能的重要组成部分。它诞生于 20 世纪 20 年代, 在 60 年代初发展成一门学科。

(5) 统计学习理论——是一种研究小样本估计和预测的理论。

2.1.2 统计学习理论的发展历史

虽然感知器的思想已经在神经生理学领域中被讨论了多年, 但直到 F. Rosenblatt(1958, 1962)把它作为一个学习机器模型, 才标志着人们对学习过程进行数学研究的真正开始^[46]。不寻常的是, F. Rosenblatt 把它表现为一个计算机程序, 通过实验说明这个模型可以被推广。1962 年 Novikoff 证明了关于感知器的第一个定理, 这个定理实际上是统计学习理论的开始, 它在创建学习理论的过程中起到了重要的作用。它在一定意义上把导致学习机器具有推广能力的原因和最小化训练集上的错误数的原则联系了起来。Tikhonov(1963)、Ivanov(1962)、Phillips(1962)发现了关于解决不适定问题的正则化原则, 这种思想及其正则化技术在统计学领域产生了深远地影响。为了从范围宽的集合(非参数集合)中估计概率密度, 必须采用某种新的推理方法, 其中利用了正则化技术。在 20 世纪 60 年代, Rosenblatt(1956)、Parzen(1962)、Chentsov(1963)提出了几种此类(非参数)算法, Vapnik 和 Stefanyuk(1978)发现了创建此类算法的一般途径, 它是建立在解决不适定问题的标准解法基础上的, 从而一举奠定了非参数统计学的基础。Vapnik 和 Chervonenkis 在 1968 年, 针对指示函数集(即模式识别), 提出了 VC 熵和 VC 维的概念, 这是统计学习理论的核心概念, 利用这些概念, 发现了泛函空间的大数定理(频率一致收敛于其概率的充分必要条件), 研究了它与学习过程的联系, 并且得到了关于收敛速度的非渐近界的主要结论。在 1971 年, Vapnik 和 Chervonenkis 发表了这些工作的完全证明。所得到的这些界使得 Vapnik 和

Chervonenkis 在 1974 年提出了结构风险最小化归纳原则，从而完成了模式识别的学习理论^[85]。Solomonoff(1960)、Kolmogorov(1965)和 Chaitin(1966)提出了算法复杂度的思想，这是统计学和信息论中最伟大的思想之一，开创了研究推理问题的信息论方法。在这些思想上，Rissanen 在 1978 年提出了对于学习问题的最小描述长度(MDL)归纳原则。

虽然有上述这些研究成果，但这些研究尚不十分完善，得到的收敛速度的非渐近界往往趋于保守，且在数学上艰涩难懂，直到上世纪 90 年代以前都没有提出能够将其理论付诸实践的较好方法，因此这些研究一直没有得到充分重视。直到 90 年代中期，有限样本情况下的学习机器理论研究逐渐成熟起来，形成了较完善的理论体系——统计学习理论(SLT)^[46]。而同时，由于诸如神经网络等较新兴的机器学习方法的研究遇到了一些困难，在这种情况下，试图从更本质上研究学习机器问题的统计学习理论逐步得到重视。

§ 2.2 统计学习理论的基本内容^[46]

2.2.1 机器学习的基本问题

一、学习问题的表示

学习问题一般都可以归结为一个包括以下三部分的函数估计模型（见图 2.1）：

(1)产生器 (G) ,产生随机向量 $x \in R^d$, R 是实数, d 是向量的维数，它们是从固定但未知的概率分布函数 $F(x)$ 中独立抽取得到的。

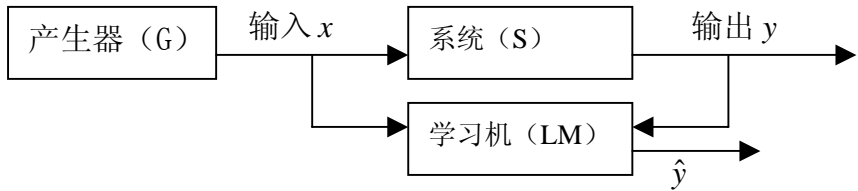


图 2.1 机器学习的基本模型

(2)系统 (S), 对每个输入向量 x 返回一个输出值 y , 产生的依据是同样固定但未知的条件分布 $F(y/x)$ 。

(3)学习机 (LM), 它能够实现一定的函数 $\{f(x, \alpha), \alpha \in \Lambda\}$, 其中 Λ 是参数集合, 即, 它能对每个输入向量 x 产生一个输出值 \hat{y} , 但是 \hat{y} 不是唯一的, 而是满足 $\hat{y} = f(x, \alpha), \alpha \in \Lambda, \Lambda$ 为指标集。

学习的问题就是从给定的函数集 $\{f(x, \alpha), \alpha \in \Lambda\}$ 中选择出能够最好地逼近系统反馈的函数。这种选择是基于训练集的, 训练集由根据联合分布 $F(x, y) = F(x)F(y|x)$ 抽取的 l 个独立同分布的观测值 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ 组成。为此引入一个损失函数 $L(y, f(x, \alpha))$ 以度量在给定输入 x 下系统反馈 y 与学习机给出的反馈 $\hat{y} = f(x, \alpha), \alpha \in \Lambda$ 之间的差异, 并对该损失函数求期望得到风险泛函:

$$R(\alpha) = \int L(y, f(x, \alpha)) dF(x, y) \quad (2.2.1)$$

学习目标就是在联合概率分布 $F(x, y)$ 未知、所有可用信息都包含在训练集 $\{(x_i, y_i)\}, i = 1, 2, \dots, l$, 中的情况下, 寻找函数 $f(x, \alpha_0)$, 使它在 $\{f(x, \alpha), \alpha \in \Lambda\}$ 上最小化风险泛函 $R(\alpha)$ ^[83]。

二、三个主要的学习问题

上面描述的学习问题是非常一般化的情形, 它包括了很多的特殊问题, 但典型的学习问题有三种: 模式识别、回归函数估计、概率密度估计。

(1)模式识别 令系统的反馈值 $y \in \{0, 1\}$, $\{f(x, \alpha), \alpha \in \Lambda\}$ 为指示函数集(指

示函数仅取 0, 1 两个值)。

考虑损失函数

$$L(y, f(x, \alpha)) = \begin{cases} 0 & y=f(x, \alpha) \\ 1 & y \neq f(x, \alpha) \end{cases} \quad (2.2.2)$$

此时, (2.2.1) 式的泛函 $R(\alpha)$ 实际上就是分类错误率, 模式识别问题就是寻找一个函数 $f(x, \alpha_0)$ 使其极小化分类错误率。

(2) 回归估计 此时, $y \in R, f(x, \alpha) \in R, \alpha \in \Lambda$, 常取损失函数:

$$L(y, f(x, \alpha)) = (y - f(x, \alpha))^2 \quad (2.2.3)$$

回归估计就是寻找一个函数 $f(x, \alpha_0) = \int y dF(y/x)$, 使其极小化 $R(\alpha)$, 而 $f(x, \alpha_0)$ 就是所求的回归函数估计。

(3) 密度估计 在给定独立同分布 $F(X)$ 的数据 x_1, x_2, \dots, x_l 的情况下, 从密度函数集合 $\{p(x, \alpha), \alpha \in \Lambda\}$ 中寻找某一函数 $p(x, \alpha_0)$ 使得风险泛函

$$R(\alpha) = \int L(p(x, \alpha)) dF(x) \quad (2.2.4)$$

其中 $L(p(x, \alpha)) = -\log p(x, \alpha)$ 。

这样, 我们得到了学习问题的一般表示: 给定独立同分布 $F(Z)$ 的训练样本集 $\{z_1, z_2, \dots, z_l\}$ 及学习机所能执行的函数集 $\{f(x, \alpha), \alpha \in \Lambda\}$, 学习目标就是最小化风险泛函

$$R(\alpha) = \int Q(z, \alpha) dF(z) \quad (2.2.5)$$

其中 $Q(z, \alpha)$ 代表损失函数。

因此, 机器学习问题也可以表示为: 从一组独立同分布的观测样本出发, 通

过最小化风险泛函 $R(\alpha)$ ，确定学习机器的广义参数 α 的过程。

三、经验风险最小化原则

在未知分布 $F(Z)$ 下，我们无法最小化 (2.2.5)。在实际问题中，联合分布是未知的，所能利用的信息只有观测样本。因此，期望风险无法直接计算和最小化。

根据概率论中的大数定律的思想，人们自然想到用算术平均

$$R_{emp}(\alpha) = \frac{1}{l} \sum_{i=1}^l Q(z_i, \alpha) \quad (2.2.6)$$

来逼近 (2.2.5) 中的期望风险。由于 $R_{emp}(\alpha)$ 是用已知的训练样本 (即经验数据) 定义的，因此称为经验风险。用经验风险 $R_{emp}(\alpha)$ 最小化来代替期望风险 $R(\alpha)$ 最小化，从而求得学习机器的参数 α 的方法，就是所谓的经验风险最小化 (ERM) 原则。经典的模式识别分类器都是基于经验风险最小化原则的。

仔细研究经验风险最小化原则和机器学习问题中的期望风险最小化要求，可以发现，使用经验风险代替期望风险并没有可靠的理论依据，所凭借的只是人们的直觉。

首先 $R_{emp}(\alpha)$ 和 $R(\alpha)$ 都是函数，概率论中的大数定律要求只有当样本数目趋向于无穷大时， $R_{emp}(\alpha)$ 将在概率意义下趋近于 $R(\alpha)$ ，并不能保证 $R_{emp}(\alpha)$ 和 $R(\alpha)$ 在同一个点上取得最小值。

其次，没有理论保证在样本数目无穷大条件下得到的学习机器，在样本数目有限的情况下仍能具有好的效果。

多年来，经验风险最小化原则作为解决模式识别等机器学习问题的基本思想，几乎统治了这个领域内的所有研究。大部分的研究者们把注意力集中在如

何更好地逼近最小化经验风险的最优解。

四、复杂性与推广能力

经验风险最小化原则希望通过最小化训练误差来实现最小化测试误差的目的。在早期的神经网络研究中，人们总是把注意力集中在如何使 $R_{emp}(\alpha)$ 更小，但很快发现，一味追求训练误差最小并不总能达到好的预测效果。实际情况是，只要学习机器足够复杂，训练时间足够长，则训练误差可以任意减小。最极端的情况是，学习机器“记住”所有训练样本，则训练误差为零，但是实际上这种学习机器几乎不具有泛化能力。某些情况下，训练误差过小反而可能会导致泛化能力下降。这就是几乎所有神经网络研究者都曾经遇到的所谓“过学习”（overfitting）问题。

例如，考虑使用 BP 网络对二维平面上两类数据的分类问题^[17]。设有两类数据是分别从两个不同的数据发生器产生的。实验中，首先从这个数据发生器中各采集一部分作为训练数据，另外再采集一部分作为测试数据。所使用的 BP 网络采用典型的三层结构(输入层，隐层和输出层)。设定网络的隐层节点数，对 BP 网络训练并测试，重复进行 50 次，并记录平均训练误差和测试误差。改变隐层节点数，重复前一步骤。实验结果列于表 2.1 中。

表 2.1 BP 网络分类实验

隐层节点	2	5	10	20	50
训练误差	8.9%	8.6%	7.8%	6.1%	6.4%
测试误差	9.2%	9.0%	9.9%	11.4%	14.3%

由表 2.1 可见，随着隐层节点数的增加，训练误差越来越小，这是意料之中的。然而，测试误差在五个隐层节点时达到最低，之后随着隐层节点的增加却呈上升趋势，这是一种典型的“过学习”现象。从这里可以得到如下结论：

1. 经验风险(或者说训练误差)对学习机器的性能有一定的影响,但不起决定作用。执行经验风险最小化原则(即最小化训练误差),并不总能提高学习机器的泛化性能(或者说降低测试误差)。

2. 复杂度高的学习机器,往往具有较低的经验风险。因此,经验风险最小化原则的结果,将使学习机器变得越来越复杂(隐层节点数增多)。

3. 学习机器的复杂度对其性能有较大影响,泛化性能好的学习机器应该具有与实验所研究问题相对应的复杂度。在这个仿真的例子中,具有五个隐层节点的三层 BP 网络是比较理想的结构。

因此,如何根据实际问题,在学习机器的经验风险和模型复杂度之间取得合理的折衷,从而使学习机器具有更高的泛化性能,是非常重要的问题。因此人们研究了很多解决办法,比如采用正则化、模型选择、噪声干扰或者有限温度学习等方法以控制学习机器的复杂度。但这些方法都缺乏完善的理论基础。

2.2.2 统计学习理论的核心内容

统计学习理论被认为是目前针对小样本统计估计和预测问题的最佳理论。主要包括四个方面:

- (1) 基于经验风险原则的统计学习过程的一致性理论;
- (2) 学习过程收敛速度的非渐近理论;
- (3) 控制学习过程推广能力的理论;
- (4) 构造学习算法的理论。

其中最具有指导性的理论结果是推广能力的界,与此相关的一个核心概念是 VC 维。

一、VC 维

为了研究经验风险最小化函数集的学习一致收敛速度和推广性,统计学习理论定义了一些指标来衡量函数集的性能,其中最重要的就是 VC 维

(Vapnik-Chervonenkis Dimension)。VC 维在模式识别方法中的直观定义是：对于一个指示函数集，如果存在 h 个样本能够被函数集里的函数按照所有可能的 2^h 种形式分开，则称函数集能够把 h 个样本打散。函数集的 VC 维就是能够打散的最大样本数目 h 。如果对任意的样本数，总有函数能打散它们，则函数集的 VC 维就是无穷大。有界实值函数的 VC 维可以通过使用阈值将函数转化成指示函数来定义^[85]。

VC 维是统计学习理论的核心概念，它是目前为止对函数集学习性能的最好描述指标，一般而言，VC 维越大，学习机器的学习能力就越强，但学习机器也越复杂（容量越大）。但遗憾的是，目前还没有通用的关于计算任意函数集的 VC 维理论，只有对一些特殊函数集的 VC 维可以准确知道，例如， n 维实数空间中线性分类器和线性实函数的 VC 维是 $n + 1$ 。而对于一些比较复杂的学习机器（比如神经网络），其 VC 维除了与函数集的选择有关外，通常也受学习算法等影响。因此要确定其 VC 维将更加困难。对于给定的学习函数集，如何用理论或实验的方法计算其 VC 维是当前统计学习理论研究中有待解决的一个问题。

二、推广性的界

对于各种类型的函数集，统计学习理论系统地研究了经验风险和实际风险之间的关系，即推广性的界^[84]。它是分析学习机器的性能和发展新的学习算法的重要基础。根据统计学习理论中关于函数集推广性界的结论，对于指示函数集中所有的函数，经验风险 $R_{emp}(\alpha)$ 和实际风险 $R(\alpha)$ 之间至少以概率 $1 - \eta$ 满足如下关系^[46, 46]：对任意给定的 $\eta (0 < \eta \leq 1)$ ，不等式

$$R(\alpha) \leq R_{emp}(\alpha) + \Phi\left(\frac{l}{h}\right) \quad (2.2.7)$$

以 $1 - \eta$ 成立。其中 h 是指示函数集 $\{Q(z, \alpha), \alpha \in \Lambda\}$ 的 VC 维， l 是给定的训练

样本数, $R_{emp}(\alpha)$ 是经验风险, $\Phi(\frac{l}{h})$ 是置信区间, 且:

$$\Phi(\frac{l}{h}) = \sqrt{\frac{h(\ln \frac{2l}{h} + 1) - \ln \frac{\eta}{h}}{l}} \quad (2.2.8)$$

由式(2.2.8)知, $\Phi(\frac{l}{h})$ 是 $\frac{l}{h}$ 的单调减函数。当 l 固定时, 它是 VC 维 h 的单调增函数; 当 VC 维 h 不变时, 它是样本数 l 的单调减函数。而 VC 维与经验风险 $R_{emp}(\alpha)$ 是一对矛盾, 它们不可能同时都很小, 所以我们要选择一个适当的 h 使两者之和最小。这可以通过结构风险最小化原理 (SRM) 来实现。

在有限的训练样本下, 学习机器的 VC 维越高(复杂性越高), 则置信范围越大, 导致实际风险与经验风险之间的可能的差别越大。这就是为什么会出现“过学习”现象的原因。机器学习过程不但要使经验风险最小, 还要使 VC 维尽量小, 以缩短置信范围, 才能取得较小的实际风险, 即对未来样本有较好的推广性。需要指出, 推广性的界是对于最坏情况的结论, 在许多情况下是较松的, 尤其当 VC 维较高时更是如此, 而且这种界只在对同一类学习函数进行比较时才有效。它可以指导我们从函数集中选择最优的函数, 在不同函数之间比较不一定成立。Vapnik 指出, 寻找更好地反映学习机器能力的参数和得到更紧的界是学习理论今后的研究方向之一。

三、结构风险最小化

传统机器学习方法中普遍采用的经验风险最小化原则在样本数目有限时是不合理的。从(2.2.8)可以看出, 我们需要同时最小化经验风险和置信范围。其实在传统的机器学习中, 我们选择机器学习模型和算法的过程就是优化置信范围的过程, 如果选择了适合现有样本的学习模型, 就可以取得比较好的效果。

但由于缺乏理论上的指导，只能依赖先验信息和经验来选择模型和算法，这造成了诸如神经网络等学习方法对使用者“技巧”的过分依赖。

统计学习理论提出了一种新的策略来解决上述问题，也就是把函数集 $S = \{f(x, w), w \in \Omega\}$ 分解为一个函数子集序列(或叫子集结构)：

$$S_1 \subset S_2 \subset \dots \subset S_k \subset S \tag{2.2.9}$$

使各个子集能够按照 Φ 的大小来排列，也就是按照 VC 维的大小来排列，即满足

$$h_1 < h_2 < \dots < h_k < \dots \tag{2.2.10}$$

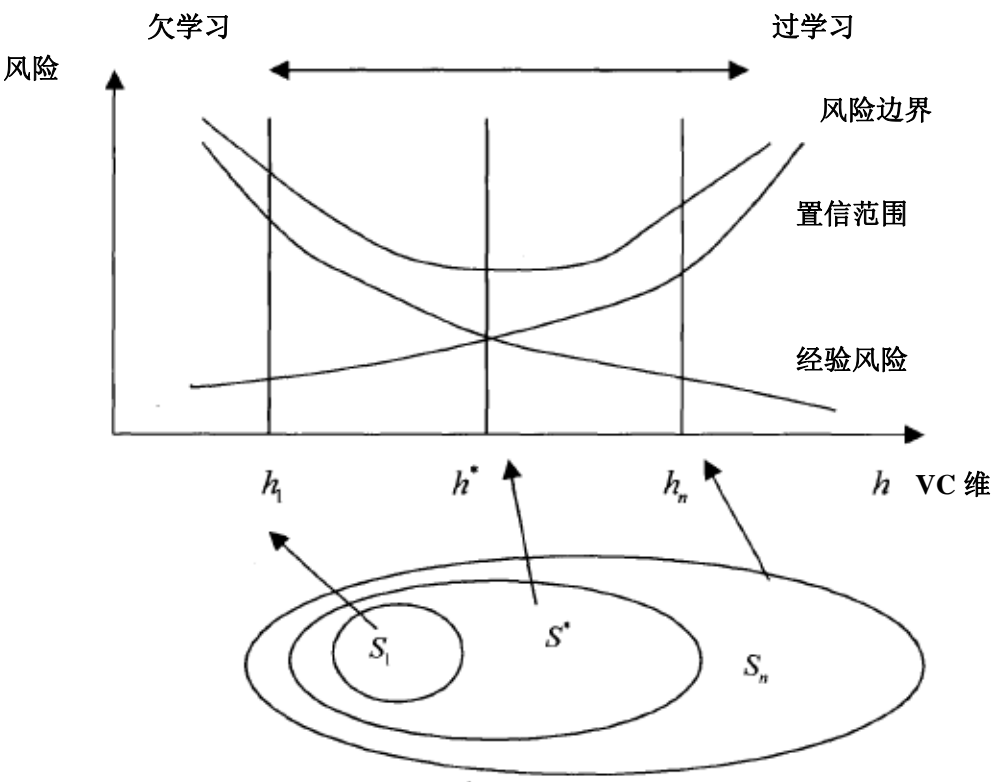


图 2.2 结构风险最小化原理图示

在每个子集中寻找最小经验风险，在子集间折衷考虑经验风险和置信范围，使实际风险最小。这种思想称作有序风险最小化或结构风险最小化，简称 SRM 原

则，如图 2.2 所示。

结构风险最小化原则为我们提供了一种不同于经验风险最小化的更科学的机器学习设计原则，但它最终是在式 (2.2.8) 的两个求和项之间进行折衷，因此实现这一原则并不容易。一般有两种思路，一是在每个子集中求经验风险最小化，然后选择使经验风险最小和置信范围之和最小的子集，显然，这种方法在子集数目很大甚至是无穷时是不可行的。第二种思路是设计函数集的某种结构使每个子集中都能取得最小的经验风险，然后选择使置信范围最小的子集，则这个子集中使经验风险最小的函数就是最优函数。支持向量机实际上就是这种思路的体现。

§ 2.3 支持向量机概述

支持向量机的理论最初来自于对数据分类问题的处理。对于线性可分数据的二值分类，如果采用神经网络来实现，其机理可以简单描述为：系统随机的产生一个超平面并移动它，直到训练集合中属于不同类别的点正好位于该超平面的不同侧面，就完成了对网络的设计要求。但是这种机理决定了不能保证最终所获得的分割平面位于两个类别的中心，这对于分类问题的容错性是不利的。

保证最终所获得的平面位于两个类别的中心对于分类问题的实际应用是很重要的。支持向量机方法很巧妙地解决了这一问题。该方法的机理可以简单描述为：寻找一个满足分类要求的最优分类超平面，使得超平面在保证分类精度的同时，能够使超平面两侧的空白区域最大化。理论上来说，支持向量机能够实现对线性可分数据的最优分类。为进一步解决非线性问题，Vapnik 等人通过引入核映射方法将低维空间中的非线性问题转化为高维空间的线性可分问题来解决^[47]。

2.3.1 广义最优分类面

SVM 是从线性可分情况下的最优分类面发展而来的，基本思想可用图 2.3 的两维情况说明。图中，圆点和叉点代表两类样本，H 为分类线，H₁, H₂ 分别为过各类中离分类线最近的样本且平行于分类线的直线，它们之间的距离叫做分类间隔(margin)。所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为 0)，而且使分类间隔最大。前者是保证经验风险最小(为 0)，后者是为了使置信风险最小，从而使实际风险最小，这是保证结构风险最小化原则的具体实现。推广到高维空间，最优分类线就成为最优超平面(Optimal Hyperplane)。

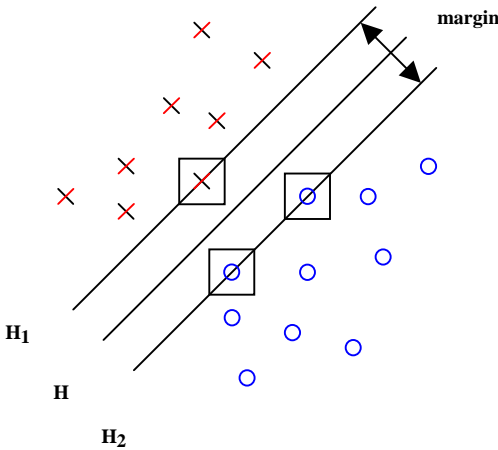


图 2.3 线性可分情况下的最优分类线

2.3.2 线性支持向量机

给定训练样本集 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, $x_i \in R^d$, $y_i \in R$ ，它们是从某一确定但未知的概率分布函数 $F(x, y)$ 中随机独立抽取的。为了使分类面对所有样本正确分类并且具备最大的分类间隔（图 2.4 和图 2.5），就要求它满足如下约束：

$$\left. \begin{array}{ll} w \cdot x_i + b \geq +1 & \text{for } y_i = +1 \\ w \cdot x_i + b \leq -1 & \text{for } y_i = -1 \end{array} \right\} \Leftrightarrow y_i(w \cdot x_i + b) - 1 \geq 0 \quad (2.3.1)$$

其中 x_i 是输入值, w 是可调的权值函数, b 是阈值, $w \cdot x$ 表示向量 w 与 x 的内积。

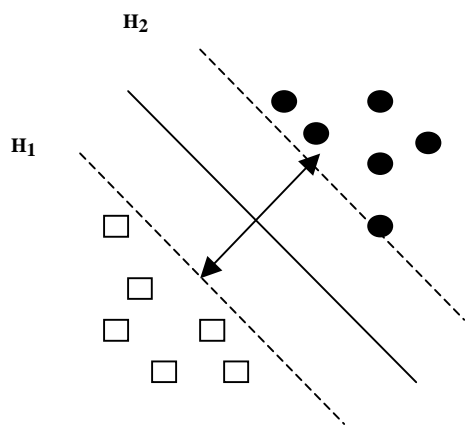


图 2.4 分类间隔较小的分类面

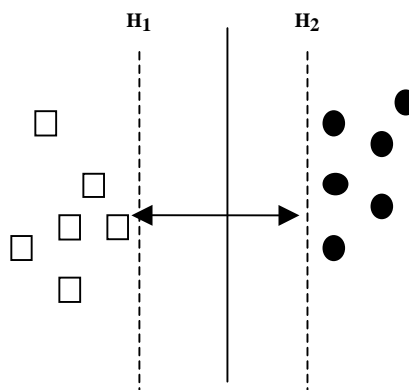


图 2.5 最大间隔的最优分类面

在线性可分条件下, 可以计算出分类间隔为

$$\min_{\{x_i | y_i = +1\}} \frac{w \cdot x_i + b}{\|w\|} - \max_{\{x_i | y_i = -1\}} \frac{w \cdot x_i + b}{\|w\|} = \frac{2}{\|w\|} \quad (2.3.2)$$

现在的目标就是在服从约束 (2.3.1) 的条件下最大化分类间隔 $2/\|w\|$, 这可以通过最小化 $\|w\|^2$ 的方法来实现。这样, 求解最优超平面问题就可以表示成如下的约束优化问题: 即在条件 (2.3.1) 的约束下, 最小化函数:

$$\Phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \quad (2.3.3)$$

这个约束优化问题称为原问题 (primal problem)。它的特点如下:

- (1) 代价函数 $\Phi(w)$ 是 w 的凸集;
- (2) 约束条件关于 w 是线性的。

因此我们可以使用 Lagrange 乘子方法解决约束最优问题。

首先，我们建立 Lagrange 函数

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot x_i + b) - 1] \quad (2.3.4)$$

其中辅助变量 $\alpha_i \geq 0$ 称作 Lagrange 乘子。约束最优问题的解由 Lagrange 函数

$L(w, b, \alpha)$ 的鞍点决定，此函数对 w 和 b 必定最小化，对 α 必定最大化。 $L(w, b, \alpha)$

对 w 和 b 求偏微分并置结果等于零，我们得到下面两个最优化条件：

$$\text{条件 1: } \frac{\partial L(w, b, \alpha)}{\partial w} = 0$$

$$\text{条件 2: } \frac{\partial L(w, b, \alpha)}{\partial b} = 0$$

由条件 1 可得到

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.3.5)$$

即最优超平面的权系数向量是训练样本向量的线性组合。

由条件 2 可得到

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (2.3.6)$$

这样就把上述问题转化为一个较简单的“对偶”问题。即

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.3.7)$$

这是一个不等式约束的二次函数极值问题(Quadratic Programming, 简称 QP)。根据最优性条件---Karush-Kuhn-Tucker 条件(简称 KKT 条件)，这个优化问题的解必须满足：

$$\alpha_i \{[(w \cdot x_i) + b]y_i - 1\} = 0 \quad (2.3.8)$$

因此，对多数样本 α_i 将为零，取值不为零的 α_i 对应使式(2.3.8)中等号成立的样本，即**支持向量**(Support Vectors)，如图 2.3 中用方框标出的样本点所示。对学习过程而言，支持向量是训练集中的关键元素，它们离决策边界最近。如果去掉所有其它训练点，再重新进行训练，得到的分类面是相同的。

求解上述问题后得到的最优分类函数是：

$$f(x) = \text{sgn}[\sum_{i=1}^l y_i \alpha_i (x_i \cdot x) + b] \quad (2.3.9)$$

由于非支持向量对应的 α_i 均为 0，因此上式的求和实际上只对支持向量进行。 b 是分类的阈值，可以由任意一个支持向量用式(2.3.8)求得或通过两类中任意一对支持向量取中值求得。

在线性不可分模式的情况下，给定这样一组训练数据，不可能建立一个不具有分类误差的分离超平面。然而，我们希望找到一个最优超平面，它对整个训练集合平均的分类误差的概率达到最小。

在类之间的分离边缘称为是软的，如果数据点 (x_i, y_i) 不满足下面的条件

$$y_i [(w \cdot x_i) + b] \geq 1, \quad i = 1, 2, \dots, l$$

这种违反条件以下面两种方式之一出现：

- 1) 数据点 (x_i, y_i) 落在分离区域之内，但在决策面正确的一侧；
- 2) 数据点 (x_i, y_i) 落在决策面错误的一侧。

注意，在情况 1 下，我们有正确的分类；但在情况 2 下，分类是错误的。因此在必要的时候放宽上面的约束条件，这可以通过在上式中引入一个松弛变量 $\xi_i \geq 0$ 来实现，此时的约束条件就变为：

$$y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, l \quad (2.3.10)$$

当分类出现错误时, $\xi_i \geq 0$, 因此, $\sum_{i=1}^l \xi_i$ 是训练集中错分样本数的上界。它们度量一个数据点对模式可分的理想条件的偏离程度。对于 $0 \leq \xi_i \leq 1$, 数据点落入分离区域的内部, 但是在决策面的正确一侧, 对于 $\xi_i > 1$, 数据点落到分离超平面的错误一侧。支持向量是那些精确满足式 (2.3.8) 的特殊数据点, 即使 $\xi_i > 0$ 。注意, 如果一个 $\xi_i > 0$ 对应的样本被遗弃在训练集外, 决策面就要改变。因此, 支持向量的定义对线性可分和不可分的情况都是相同。

这样就需要在目标函数中为分类误差分配一个额外的代价函数, 即引入错误惩罚分量。所以, 现在的目标函数改为:

$$L(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (2.3.11)$$

满足约束条件:

$$\begin{cases} y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (2.3.12)$$

这样, w 范数的平方被认为是一个关于不可分离点的联合最小化中一个数量项, 而不是作为强加在关于不可分离点数量的最小化上的一个约束条件。此公式与结构风险最小化原则完全吻合。其中 $C > 0$ 是一个常数, 它控制对错分样本的惩罚程度, 控制机器的复杂性和不可分离点数之间的平衡, C 越大表示对错误的惩罚越重。

使用 Lagrange 乘子的方法, 我们可以得到不可分模式对偶问题的表示如下:

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.3.13)$$

满足约束:

$$\begin{cases} \sum_{i=1}^l y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{cases} \quad (2.3.14)$$

松弛变量 ξ_i 和它们的 Lagrange 乘子都不出现在对偶问题里。其 KKT 条件被定义为：

$$\begin{cases} \alpha_i [y_i (w \cdot x_i + b) - 1 + \xi_i] = 0 \\ \xi_i (\alpha_i - C) = 0 \end{cases} \quad (2.3.15)$$

从上面的介绍可以看出，线性不可分情况和线性可分情况的差别就在于可分模式中的约束条件 $\alpha_i \geq 0$ ，在不可分模式中换成了更为严格的条件 $0 \leq \alpha_i \leq C$ 。除了这一修正，线性不可分情况的约束最优化问题中，权值 w 和阈值 b 的最优值的计算都和线性可分情况中的过程是相同的。线性可分情况的最优化问题可以视作一种特殊情形，包含在线性不可分的情况的最优化问题之中。在式 (2.3.9) 和式 (2.3.10) 中对所有 i 令 $\xi_i = 0$ ，即可得到相应的线性可分情况时的形式。

2.3.3 非线性支持向量机

在很多情况下，训练数据集是线性不可分的，Vapnik 等人提出了用广义分类面来解决这一问题，也就是折衷考虑最小错分样本和最大分类间隔。对于非线性问题，可以通过非线性变换将它转化为某个高维空间中的线性问题，在这个高维空间中寻求最优分类面。但是在低维输入空间向高维特征空间映射过程中，由于空间维数急速增长，这就使得在大多数情况下难以在特征空间直接计算最优分类平面。支持向量机通过定义核函数 (Kernel Function)，巧妙地将这一问题转化到输入空间进行计算^[54]，其具体机理如下：

注意到在上面的问题求解中都只涉及内积运算，因此可以假设有非线性映射：

$$\varphi: R^d \rightarrow H \quad (2.3.16)$$

把数据从原空间 R^d 映射到更高维的特征空间 H ，在特征空间中构造最优超平面时，训练算法仅使用特征空间中的点积，即 $\varphi(x_i) \cdot \varphi(x_j)$ 。所以，若能找到一个函数 $K(\cdot, \cdot)$ ，使得 $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ ，这样，在特征空间中公式的推导类似于线性情况，只是要把原空间中的 x 或 x_i ，全部用 $\varphi(x)$ 或 $\varphi(x_i)$ 代替即可。在高维空间中实际上只需要进行内积运算，甚至不必知道变换中的具体形式。

根据泛函的有关理论，只要一种函数 $K(x_i, x_j)$ 满足 Mercer 条件，它就对应某一变换空间的内积。统计学习理论指出，根据 Hilbert-Schmidt 原理，只要核函数满足 Mercer 条件，它就对应某一变换空间中的内积^[46, 79]。因此，在最优分类面中采用满足 Mercer 条件的内积函数 $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) = (\varphi(x_i), \varphi(x_j))$ 就可以实现某一非线性变换后的线性分类，而计算复杂度却没有增加，此时的拉格朗日函数变为：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot \varphi(x_i) + b) - 1] \quad (2.3.17)$$

把上述问题转化为“对偶”问题，即：

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (2.3.18)$$

求解上述问题后得到的最优判别函数（即分类函数）为

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^m y_i \alpha_i K(x, x_i) + b \right\} \quad (2.3.19)$$

$$b = \frac{1}{N_{NSN}} \sum_{x_i \in SV} (y_i - \sum_{x_j \in SV} \alpha_j y_j K(x_j, x_i)) \quad (2.3.20)$$

其中 m 是支持向量的个数， N_{NSN} 为标准支持向量数， SV 为支持向量的集合。

由于最终判别函数中实际只包含支持向量的内积以及求和，因此识别时的计算复杂度取决于支持向量的个数。

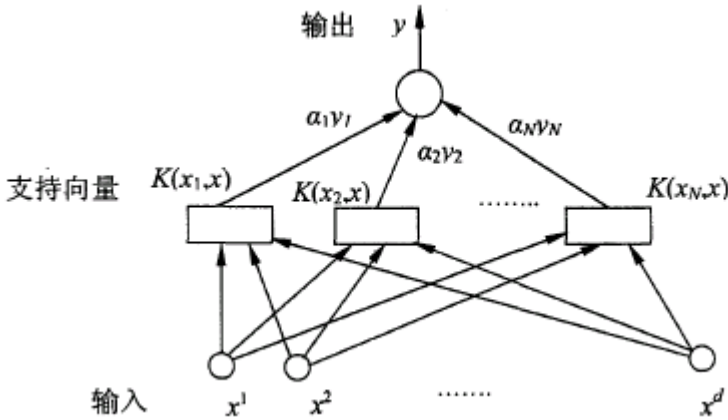


图 2.6 支持向量机示意图

由图 2.6 可以看出，支持向量机求得的决策函数形式上类似于一个神经网络，其输出是若干中间层节点的线性组合，而每一个中间层节点对应于输入样本与一个支持向量的内积，因此被称为是支持向量网络。

2.3.4 核函数

在前面的讨论中可知：在实际计算中，我们无须确切地知道映射 φ 的结构，只要找一个核函数使得它可以写成同一函数的内积形式。如何找核函数呢？下面的定理回答了这个问题。

定理 2.1 (Mercer 定理) 设 $K \in L_\infty(\chi^2)$ ，由它定义的积分算子

$$T_K : L_2(\chi) \rightarrow L_2(\chi),$$

使得 $T_K f(\cdot) = \int_\chi K(\cdot, x) f(x) d\mu(x)$ 是正的。再设 $\varphi_j \in L_2(\chi)$ ， λ_j 是 T_K 的特征函数与相应特征值，即

$$(T_K \varphi_j)(x) = \lambda_j \varphi_j(x),$$

则 $K(x, x') = \sum_{j \in N} \overline{\varphi_j(x)} \varphi_j(x')$, $\overline{\varphi_j}$ 表示 φ_j 的共轭。

由于上面定理不易验证, 我们往往采用简化 Mercer 条件来判断一个函数是否为核函数。

定理 2.2 ^[46] (Mercer 条件) 对于任意的对称函数 $K(x, x')$, 它是某个特征空间中的内积运算的充分必要条件是, 对于任意的 $\varphi(x) \neq 0$ 且 $\int \varphi^2(x) dx < \infty$, 有

$$\iint K(x, x') \varphi(x) \varphi(x') dx dx' > 0 \quad (2.3.21)$$

由定理可知, 任何满足 Mercer 条件的函数都可以做为支持向量机的核函数。

推论 1 ^[46] 设 $K_1(x, x'), K_2(x, x')$ 是支持向量机的核函数, $C_1 > 0, C_2 > 0$, 则 $K(x, x') = C_1 K_1(x, x') + C_2 K_2(x, x')$ 也是支持向量核。

常见的用于 SVM 中的核函数的具体形式如下:

(1) p 阶多项式核

$$K(x, x') = (x \cdot x')^p, \quad p = 1, 2, \dots \quad (2.3.22)$$

$$K(x, x') = ((x \cdot x') + C)^p, \quad p = 1, 2, \dots \quad (2.3.23)$$

其中 C 是常数, 常取 1。

(2) 高斯径向基核

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (2.3.24)$$

(3) 多层感知器, 又称 Sigmoid 核, 双曲正切核

$$K(x, x') = \tanh((x \cdot x') - \theta) \quad (2.3.25)$$

其中 θ 为正常数。

(4) 样条核函数

$$K(x, x') = \sum_{i=0}^p x^i x'^i + \sum_{j=1}^N (x - r_j)_+^p (x' - r_j)_+^p \quad (2.3.26)$$

其中

$$x_+^p = \begin{cases} x^p, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

关于核函数我们有如下的定理：

定理 2.3 对于任意给定的核函数 $K(x, x') = (\varphi(x), \varphi(x'))$ ，映射 $\varphi: R^d \rightarrow H$ 是不唯一的^[29]。

采用不同的内积函数将导致不同的支持向量机算法，目前对支持向量机算法的研究得到了长足的发展。

2.3.5 标准支持向量回归机

与统计学习理论中的很多结论一样，虽然 SVM 方法是通过分类问题提出的，但它同样可以很好地应用于回归函数估计问题^[46]，支持向量机回归有线性回归和非线性回归两种情形。对于线性回归，考虑用线性回归函数

$$y = f(x) = w \cdot x + b \quad (2.3.27)$$

拟合数据 $\{(x_i, y_i), i = 1, 2, \dots, l\}, x_i \in R^d, y_i \in R$ 的问题，为保证 (2.3.27) 的平坦，必须寻找一个最小的 w ，为此，采用最小化欧几里德空间的范数。其中 w 和 b 分别为线性回归函数的法向量及偏移量，并假设所有的训练数据在精度 ε 下无误差地用线性函数拟合，我们可以解下面的优化问题，即

$$\min \quad \frac{1}{2} \|w\|^2 \quad (2.3.28)$$

约束条件

$$\begin{cases} y_i - w \cdot x_i - b \leq \varepsilon \\ w \cdot x_i + b - y_i \leq \varepsilon \end{cases} \quad i = 1, 2, \dots, l \quad (2.3.29)$$

在不能完全满足约束条件(2.3.29)下，可以引入松弛变量 ξ_i 和 ξ_i^* ，此时优化问题可以转化为下面的问题，即

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2.3.30)$$

约束条件

$$\begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \end{cases} \quad i = 1, 2, \dots, l$$

$$\xi_i, \xi_i^* \geq 0,$$

$$C > 0 \quad (2.3.31)$$

上述的优化函数为二次型，约束条件是线性的，因此是个典型的二次规划问题，可用拉格朗日乘子法求解。引入拉格朗日乘子 $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ ，

$$\begin{aligned} L(w, b, \xi_i, \xi_i^*) = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + w \cdot x_i + b) \\ & - \sum_{i=1}^l \alpha_i^* (y_i + \varepsilon + \xi_i^* - w \cdot x_i - b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (2.3.32)$$

在最优解处有：

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot x_i = 0 \\ \frac{\partial L}{\partial b} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \\ \frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \end{cases} \quad (2.3.33)$$

将式 (2.3.33) 代入式 (2.3.32)，将线性可分条件下的原问题变换为其对偶问题：

$$\max \quad \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i \cdot x_j) \quad (2.3.34)$$

满足条件：

$$(1) \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0;$$

$$(2) \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l,$$

$$0 \leq \alpha_i^* \leq C, \quad i = 1, 2, \dots, l,$$

其中 C 为用户给定的参数。

求解方程 (2.3.33) 和 (2.3.34)，得到回归函数的法向量和回归函数：

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot x_i \quad (2.3.35)$$

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*)(x_i \cdot x) + b \quad (2.3.36)$$

这里， $(x_i \cdot x)$ 为向量 x_i 与 x 向量的内积。

在非线性情况下，可以把样本 x 通过非线性映射 $\varphi(x)$ 映射到高维特征空间 H ，并在 H 中求解最优回归函数。这样，在高维空间中的线性回归，就对应于低维空间中的非线性回归。因此，在最优回归函数中采用适当的核函数 $K(x_i, x)$ 代替高维空间中的向量内积 $\varphi(x_i) \cdot \varphi(x)$ ，就可以实现某一非线性变换后的线性拟合，而计算复杂度却没有增加。此时最优化问题转化为

$$\max \sum_{i=1}^l y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i \cdot x_j) \quad (2.3.37)$$

此时对应的回归函数及其法向量为：

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i \cdot x) + b \quad (2.3.39)$$

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i \cdot x) \quad (2.3.38)$$

其中 ε 和 C 是控制逼近函数 VC 维的自由参数， ε 和 C 分别是精度和惩罚系数，两者都由用户选择。

§ 2.4 本章小结

传统统计学的研究是假定样本数目趋于无穷大时的渐近理论，大多数现有的机器学习方法也是基于这个假设。现有的基于传统统计学的学习方法在有限样本的情况下难以取得理想的效果。统计学习理论是在有限样本情况下新建立起来的统计学理论体系，为人们系统地研究小样本情况下机器学习问题提供了有力的理论基础。支持向量机是在统计学习理论基础上开发出来的一种新的、非常有效的机器学习新方法。它较好地解决了以往困扰很多学习方法的小样本、非线性、过学习、高维数、局部极小点等实际问题，具有很强的推广能力。目前，统计学习理论和支持向量机作为小样本学习的最佳理论，开始受到越来越广泛的重视，正在成为人工智能和机器学习领域新的研究热点。

本章首先对统计学习理论进行了回顾，概述了统计学习理论研究的基本问题及主要内容。为了更好说明统计学习理论在实际中的实现问题，我们回顾了 SVM 的基本概念及基本理论。着重介绍了线性与非线性标准支持向量、标准支持向量回归机的基本理论与方法。

第三章 支持向量机分类方法的基本理论与算法研究

分类问题是现实生活中普遍存在的问题。分类的作用和根本目的在于，面对某一具体事物时将其正确地归于某一类，然后用同一种方法去处理同一类中的不同事物。将某一事物正确归入某一类的方法即分类方法，研究分类方法首先要确定分类标准，而对任何事物都不存在纯客观的分类标准，任何分类都带有主观性，因此，对不同的分类标准会产生不同的分类方法。目前已有各种各样的分类方法^[79, 80]，如人工神经网络、贝叶斯决策、决策树等等，本章将要研究的是分类问题的新方法，即支持向量机方法。

支持向量机(Support Vector Machine, 简称 SVM)是针对模式识别问题提出来的，它的理论最初来自于对数据分类问题的处理。对于数据分类问题，如果采用传统的神经网络方法来实现，其机理可以简单地描述为：系统随机产生一个超平面并移动它，直到训练集中属于不同分类的点正好位于平面的不同侧面。这就决定了用神经网络方法进行数据分类最终获得的分割平面将相当靠近训练集中的点，从而出现“过学习”现象，造成神经网络方法的泛化性能较差。因此在 SVM 方法中引入了最优超平面，寻找一个满足分类要求的分割平面，并使训练集中的点距离该分割平面尽可能远，也就是使分割平面两侧的空白区域最大。

SVM 方法中的大部分算法都是针对两类问题提出的。随着对支持向量机研究的深入，许多研究人员提出了一些支持向量机的变形算法^[60-77, 101-112]，在本文中按照他们的算法结构特征归纳为如下系列，如 C-SVM 系列、 ν -SVM 系列、W-SVM(weighted SVM)系列、LS-SVM(least-square SVM)和 One-class SVM 等算法^[98]。这些算法主要是通过增加函数项、变量或系数等方法使公式变形，产生

出各种有某一方面优点或者一定应用范围的算法。在本章讨论的是两类问题的支持向量机算法及其变形算法。通过对这些变形的支持向量机分类算法进行讨论与研究，比较他们各自的特点及适用范围，以便对这些方法有一个全面地认识，为找到新的支持向量机分类算法做好理论准备。

§ 3.1 分类问题的一般描述

本节将详细介绍支持向量机关于分类问题的理论基础和若干常见算法模型。下面以 2 维图形为例解释常见的两分类问题。首先看以下三种不同的情形。

从图 3.1 中可以看出两类样本分布间隔较大，很容易找到一条直线将两类样本分离，这属于线性可分问题，对应的支持向量机线性可分模型主要就是求解其最大间隔分类面。

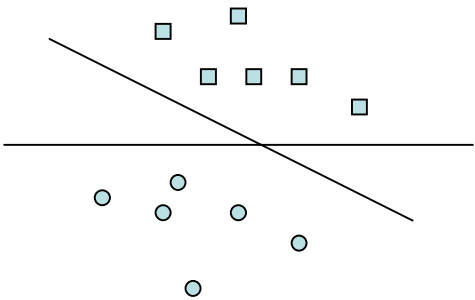


图 3.1 线性可分支持向量机

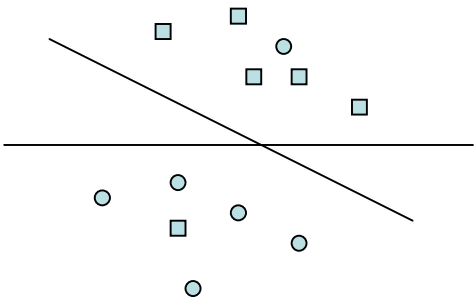


图 3.2 线性不可分支持向量机

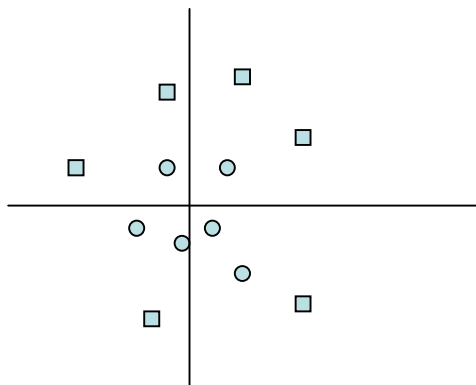


图 3.3 非线性可分支持向量机

图 3.2 描述的是线性不可分情形，用一条直线虽不能完全的分离样本集，但大致上仍能把两类样本集分离，此时仍采用线性分类机予以解决，只需在模型中加入对错分样本的惩罚项。

图 3.3 则描述了非线性可分的情形，此时用一条直线加以划分两样本集时会产生很大的误差，而若采用曲线(非线性)分类器则可顺利地分离两样本集，这就是非线性可分的问题，通过引入核函数将原先低维空间中非线性可分的样本集映射到高维特征空间中，再在高维空间中采用线性分类器将映射后的样本集加以划分。

§ 3.2 C-支持向量机分类算法及其变形算法

3.2.1 C-SVM 算法

本算法就是第二章介绍的支持向量机算法^[46]，这是 Vapnik 分别在 1995 和 1998 年提出的。为了便于比较，下面列出主要的步骤和方法。

给定训练样本集 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, $x_i \in R^d$, $y_i \in \{-1, +1\}$ ，对于这样的分类问题，我们首先引进从输入空间 R^n 到 Hilbert 空间 H 的变换：

$$\varphi: R^n \rightarrow H$$

$$x \rightarrow \varphi(x)$$

然后在 Hilbert 空间 H 中构造原始问题:

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (3.2.1)$$

约束条件:

$$\begin{cases} y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.2.2)$$

其中 $C > 0$ 是一个常数, 它控制对错分样本的惩罚程度, 控制机器的复杂性和不可分离点数之间的平衡, C 越大表示对错误的惩罚越重。它也是算法中唯一可以调节的参数。

可以看出模型 (3.2.1) 为典型的二次规划问题, 只是问题的规模巨大, 且绝大多数不具有稀疏性, 使得典型的二次规划求解算法不能直接应用。支持向量分类机常用的求解方法是通过先求解其对偶问题的解, 再利用该对偶问题的解来表示原始问题的解, 从而确定原始分类问题的最优判别函数。采用 Lagrange 乘子法求解这个具有线性约束的二次规划问题, 得到的对偶问题为:

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^l \alpha_i \quad (3.2.3)$$

约束条件:

$$\begin{cases} \sum_{i=1}^l y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C, & i = 1, 2, \dots, l \end{cases} \quad (3.2.4)$$

其中 $K(x_i, x_j)$ 为核函数:

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$$

通过求解上述对偶问题得最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$, 选取 α^* 的一个正分量

$0 < \alpha_j^* < C$ ，并据此计算阈值

$$b^* = y_j - \sum y_i \alpha_i^* K(x_i, x_j), \quad (3.2.5)$$

最后构造决策函数

$$f(x) = \text{sgn}[\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*] \quad (3.2.6)$$

该算法称为 C -支持向量分类机。

3.2.2 C -SVM 的二次松弛变量算法

这种算法也是由 Vapnik 分别在 1995 和 1998 年提出的^[46]。 C -SVM 的二次松弛变量问题的初始问题为：

$$\min \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l \xi_i^2 \quad (3.2.7)$$

约束条件：

$$\begin{cases} y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.2.8)$$

它的对偶问题为：

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j (K(x_i, x_j) + \frac{1}{C} \delta_{ij}) - \sum_{j=1}^l \alpha_j \quad (3.2.9)$$

约束条件：

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (3.2.10)$$

$$\alpha_i \geq 0, i = 1, \dots, l \quad (3.2.11)$$

其中 $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ 。

通过求解上述对偶问题得最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ ，选取 α^* 的一个正分量 $0 < \alpha_j^* < C$ ，并据此计算阈值：

$$b^* = y_j \left(1 - \frac{\alpha_j^*}{C}\right) - \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j), \quad (3.2.12)$$

最后构造决策函数：

$$f(x) = \text{sgn} \left[\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^* \right] \quad (3.2.13)$$

3.2.3 BSVM 算法

在文献[100]中讨论了标准 C -支持向量分类机解的唯一性，给出了决策函数中 b 的不唯一的条件和求解公式。如果将目标函数中加上一项 $\frac{1}{2}b^2$ ，则原始问题变为：

$$\min \quad \frac{1}{2}(\|w\|^2 + b^2) + C \sum_{i=1}^l \xi_i \quad (3.2.14)$$

约束条件：

$$\begin{cases} y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.2.15)$$

它的对偶表达式为如下所示的边界约束问题，因此，称之为 BSVM(bounded SVM) 算法：

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j [K(x_i, x_j) + 1] - \sum_{j=1}^l \alpha_j \quad (3.2.16)$$

约束条件：

$$0 \leq \alpha_i \leq C, i = 1, \dots, l \quad (3.2.17)$$

这是有唯一解的优化问题，与标准的 C -SVM 比较，BSVM 的优化目标多一

项，而约束条件少一项等式约束，变为边界约束条件下二次规划问题，适合迭代求解。同时应用矩阵分解技术，每次只需更新 Lagrange 乘子 α 的一个分量，从而不需将所有样本载入内存，提高了收敛速度。可以容易地处理没有通常线性约束的问题，而且推广误差也是可以接受的。

3.2.4 无约束的 C-SVM 算法

文献[105]将目标函数(3.2.14)中的 ξ_i 取为 ξ_i^2 ，则问题变为：

$$\min \quad \frac{1}{2}(\|w\|^2 + b^2) + \frac{C}{2} \sum_{i=1}^l \xi_i^2 \quad (3.2.18)$$

约束条件：

$$\begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.2.19)$$

可知该问题关于 ξ_i 的解 ξ_i^* 应满足：

$$\xi_i^* = [1 - y_i(w \cdot x_i + b)]_+, \quad i = 1, \dots, l \quad (3.2.20)$$

其中函数 $(\cdot)_+$ 是单变量函数

$$(\Delta)_+ = \begin{cases} \Delta, & \Delta \geq 0; \\ 0, & \Delta < 0. \end{cases} \quad (3.2.21)$$

把(3.2.20)代入到(3.2.18)中，就得到了无约束最优化问题：

$$\min \quad \frac{1}{2}(\|w\|^2 + b^2) + \frac{C}{2} \sum_{i=1}^l [1 - y_i(w \cdot x_i + b)]_+^2 \quad (3.2.22)$$

上述问题是严格凸的无约束最优化问题，它有唯一的最优解。函数 $(\cdot)_+$ 是不可微的，需要用非光滑的无约束最优化方法求解。

如果希望用通常的无约束最优化方法求解，则考虑对目标函数进行光滑化，得到近似的最优化问题：

$$\min \quad \frac{1}{2}(\|w\|^2 + b^2) + \frac{C}{2} \sum_{i=1}^l P[1 - y_i(w \cdot x_i + b), \lambda]^2 \quad (3.2.23)$$

其中 $P(\cdot, \lambda)$ 是以 λ 为参量的函数:

$$P(\Delta, \lambda) = \Delta + \frac{1}{\lambda} \log(1 + e^{-\lambda \Delta}) \quad (3.2.24)$$

当 λ 充分大时, 光滑无约束问题 (3.2.23) 的解会近似于非光滑无约束问题 (3.2.22) 的解。

通过引入核函数, 并用 $\sum_{i=1}^l \alpha_i^2$ 来代替 $\|w\|^2$, 可以得到带有核的光滑无约束问题:

$$\min \quad \frac{1}{2}(\sum_{i=1}^l \alpha_i^2 + b^2) + \frac{C}{2} \sum_{i=1}^l P[1 - y_i(\sum_{j=1}^l y_j \alpha_j K(x_j, x_i) + b), \lambda]^2 \quad (3.2.25)$$

该问题的目标函数具有连续的梯度和 Hesse 矩阵, 而且是无约束的, 可以用基本的无约束问题算法来求解, 求得最优解 (α^*, b^*) 后, 便可构造出决策函数:

$$f(x) = \text{sgn}[\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*] \quad (3.2.26)$$

§ 3.3 ν -支持向量机分类算法及其变形算法

3.3.1 ν -SVM 算法^[102]

该算法要解决的原始问题为:

$$\min \quad \frac{1}{2}\|w\|^2 - \nu\rho + \frac{1}{l} \sum_{i=1}^l \xi_i \quad (3.3.1)$$

约束条件:

$$\begin{cases} y_i[(w \cdot x_i) + b] \geq \rho - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0, & i = 1, \dots, l; \quad \rho \geq 0 \end{cases} \quad (3.3.2)$$

它的对偶问题为:

$$\min \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (3.3.3)$$

约束条件:

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (3.3.4)$$

$$0 \leq \alpha_i \leq \frac{1}{l}, i = 1, \dots, l \quad (3.3.5)$$

$$\sum_{i=1}^l \alpha_i \geq \nu \quad (3.3.6)$$

求解, 即得其最优解:

$$\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$$

选取 $j \in S_+ = \{i \mid \alpha_i^* \in (0, 1/l), y_i = 1\}$, $k \in S_- = \{i \mid \alpha_i^* \in (0, 1/l), y_i = -1\}$, 计算

$$b^* = -\frac{1}{2} \sum_{i=1}^l y_i \alpha_i^* (K(x_i, x_j) + K(x_i, x_k)); \quad (3.3.7)$$

可构造出决策函数:

$$f(x) = \text{sgn}[\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*] \quad (3.3.8)$$

ν -SVM 分类中的参数 ν 与 C -SVM 分类中的参数 C 相比, 不仅具有更直观的意义, 而且便于选择。下面的定理给出了选择参数 ν 的理论依据^[58]。

定理 3.1 设给定由 l 个样本点组成的训练集 T , 并用算法 ν -SVM 进行分类, 若所得到的 $\rho^* > 0$, 则

(1) 若记间隔错误样本点的个数为 p , 则 $\nu \geq p/l$, 即 ν 是间隔错误样本点所占总样本数的份额的上界;

(2) 若记支持向量的个数为 q ，则 $v \leq q/l$ ，即 v 是支持向量的个数所占总样本点数的份额的下界。

这里的间隔错误样本点，或者是两个超平面

$$(w^* \cdot x) + b^* = \rho^* \text{ 和 } (w^* \cdot x) + b^* = -\rho^* \quad (3.3.9)$$

形成的间隔内的点，或者是被决策函数 $f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*)$ 分错的点。

3.3.2 Bv -SVM 算法

标准 v -SVM 算法比 C -SVM 复杂，不是解决大规模问题的有效方法，因此文献[107]提出的一种变形 v -SVM 算法，在目标函数(3.3.1)中加上一项 $\frac{1}{2}b^2$ ，为了区别，称之为 Bv -SVM 算法。它不仅保留了 v -SVM 算法的优点，而且把 C -SVM 的解法可以直接借鉴过来。

原始问题关于 (w, b) 的解是唯一的，此时原始问题为

$$\min \quad \frac{1}{2}(\|w\|^2 + b^2) - v\rho + \frac{1}{l} \sum_{i=1}^l \xi_i \quad (3.3.10)$$

约束条件：

$$\begin{cases} y_i(w \cdot x_i + b) \geq \rho - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0, & i = 1, \dots, l; \quad \rho \geq 0 \end{cases} \quad (3.3.11)$$

它的对偶问题为：

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j [K(x_i, x_j) + 1] \quad (3.3.12)$$

约束条件：

$$0 \leq \alpha_i \leq \frac{1}{l}, i = 1, \dots, l \quad (3.3.13)$$

$$\sum_{i=1}^l \alpha_i \geq \nu \quad (3.3.14)$$

对偶问题中少了等式约束,减少了计算的复杂性,其解法与具有唯一解的 C -SVM 方法类似。求得最优解 (α^*, b^*) 后,便可构造出决策函数

$$f(x) = \text{sgn}[\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*] \quad (3.3.15)$$

§ 3.4 线性规划形式的支持向量机分类算法

在上述介绍的算法中,最终得到的最优化问题都是凸二次规划。鉴于线性规划问题的研究已十分成熟,并有了十分有效地求解大型和超大型问题的算法,所以人们也建立了线性规划形式的支持向量机分类算法。

例如:与 C -支持向量分类机相对应的线性规划问题为^[108]:

$$\min \sum_{i=1}^l (\alpha_i + \alpha_i^*) + C \sum_{i=1}^l \xi_i \quad (3.4.1)$$

约束条件:

$$y_i [\sum_{j=1}^l (\alpha_j - \alpha_j^*) K(x_j, x_i) + b] \geq 1 - \xi_i, \quad (3.4.2)$$

$$\alpha_i, \alpha_i^*, \xi_i \geq 0, i = 1, \dots, l \quad (3.4.3)$$

与 ν -支持向量分类机相对应的线性规划问题为:

$$\min \frac{1}{l} \sum_{i=1}^l \xi_i - \nu \rho \quad (3.4.4)$$

约束条件:

$$\frac{1}{l} \sum_{i=1}^l (\alpha_i + \alpha_i^*) = 1 \quad (3.4.5)$$

$$y_i(\sum_{j=1}^l(\alpha_j - \alpha_j^*)K(x_j, x_i) + b) \geq \rho - \xi_i \quad (3.4.6)$$

$$\alpha_i, \alpha_i^*, \xi_i \geq 0, i = 1, \dots, l; \rho \geq 0 \quad (3.4.7)$$

求解这些优化问题，可得相应的决策函数。

§ 3.5 W-支持向量机分类算法

支持向量机引入惩罚系数 C 实行对错误分类的惩罚。然而在实际应用中，某些重要样本正确分类的要求高，而某些样本正确分类的要求低，因此，在优化问题描述中，对每个采样点用不同的惩罚系数，以得到更准确的分类，这种支持向量机我们称之为**加权支持向量机**（Weighted Support Vector Machine，**W-SVM**）。另外，不同类别的样本数量差异比较大时，存在着分类结果偏向数量较多的问题^[154]，这类问题的解决方法本质上也是 **W-SVM**。

3.5.1 加权的 C -SVM 算法

如果训练集中正负类别的样本数量差异较大时，存在着分类结果偏向数量较多的那一类的问题，因此我们希望对训练集中正负训练点赋予不同的权重，另外的情况是对那些需要正确分类的重要样本希望赋予大的权重，对某些正确分类要求低的样本赋予小的权重，这样就可以得到加权的 C -支持向量分类机。其原始问题为：

$$\min \quad \frac{1}{2} \|w\|^2 + \sum_{i=1}^l C_i \xi_i \quad (3.5.1)$$

约束条件：

$$\begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.5.2)$$

其中 $C_i > 0, i = 1, \dots, l$ 。

它的对偶问题为:

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j [K(x_i, x_j) + 1] - \sum_{j=1}^l \alpha_j \quad (3.5.3)$$

约束条件:

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (3.5.4)$$

$$0 \leq \alpha_i \leq C_i, i = 1, \dots, l \quad (3.5.5)$$

求解此对偶问题, 可得决策函数。

3.5.2 FSVM 算法

Lin Chun-Fu 在 2002 年提出 Fuzzy SVM(FSVM)算法^[111], 就是给每个样本都赋一个模糊隶属度值, 这样不同的样本对决策函数的学习有不同的贡献, 可以减少外部的影响。其本质上是一种针对 $C-SVM$ 的加权 SVM, 初始问题为:

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l s_i \xi_i \quad (3.5.6)$$

约束条件:

$$\begin{cases} y_i [w \cdot \varphi(x_i) + b] \geq 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.5.7)$$

其中 $s_i > 0$ 表示样本的加权系数。它的对偶问题为:

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \quad (3.5.8)$$

约束条件:

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (3.5.9)$$

$$0 \leq \alpha_i \leq s_i C_i, i = 1, \dots, l \quad (3.5.10)$$

针对一个给定的问题，关键在于模糊隶属函数的产生。首先定义模糊隶属度的下限，其次需要选择数据集的主要特性，并找到这种特性和模糊隶属度之间的关系。

加权算法的关键是合理选择系数 s_i 。文献[111]认为在样本集中第 1 个样本的重要性最低， s_i 设定为远小于 1 的值，最后一个样本的重要性最高， s_i 设定为 1。采用线性插值方法得到其他样本点的加权系数。另外，这种算法都可以有效地补偿由于各类别样本数量之间差异而引起的倾向性问题。

3.5.3 双 ν -SVM 算法

为了消除 ν -SVM 算法中，大样本和小样本之间差异太大而造成对分类精度的影响，文献[121]提出的双 ν -SVM 算法用于补偿两种类别之间样本差异的影响，用两个 ν 来修正 ν -SVM 算法，每个 ν 对应一类，可以对每一类的支持向量数界限分别进行调整，这种调节可以补偿训练类别大小不等的影响，灵活地为每一类指定一个不同的误差比率。本质上也是一种 W -SVM 算法。该算法的初始问题为：

$$\min \quad \frac{1}{2} \|w\|^2 - \sum_{i=1}^l C_i (\nu \rho - \xi_i) \quad (3.5.11)$$

约束条件：

$$\begin{cases} y_i [w \cdot \varphi(x_i) + b] \geq \rho - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0, & i = 1, \dots, l; \quad \rho \geq 0 \end{cases} \quad (3.5.12)$$

它的对偶问题为：

$$\min \quad \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (3.5.13)$$

约束条件：

$$0 \leq \alpha_i \leq C, i = 1, \dots, l \quad (3.5.14)$$

$$\sum_{i=1}^l \alpha_i \geq v \sum_{i=1}^l C_i \quad (3.5.15)$$

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (3.5.16)$$

该算法中的 C_i 只有两个值 C_+ 和 C_- ， C_+ 表示正类对应的加权系数， C_- 表示负类对应的加权系数。根据二类别中每类样本的数量来确定 C_+ 和 C_- 的值，即 $C_+ : C_- = l_+ : l_-$ ，其中 l_+ 和 l_- 分别表示正类和负类的样本数。由此得到的算法避免了在各类别数量不均衡时引起的分类结果会偏向于数量较多一类的问题。通过给不同类别的样本设置一定比例关系的权值，就可有效地补偿类别大小差异的影响。例如在两种类别的情况下，正类和负类分别设定不同的权值来消除类别差异的影响。

§ 3.6 LS-支持向量机分类算法

在 Suykens[1999; 2002]和 Vandewalb 提出的最小二乘支持向量机(least-square SVM, LS-SVM)中^[87, 88]，优化指标采用了平方项，只有等式约束，而没有 C-SVM 的不等式约束，从而推出不同的一系列的等式约束，而不是二次规划问题^[92, 112]。其问题表示为：

$$\min \quad \frac{1}{2} \|w\|^2 + \frac{r}{2} \sum_{i=1}^l \xi_i^2 \quad (3.6.1)$$

约束条件：

$$\begin{cases} y_i [w \cdot \varphi(x_i) + b] = 1 - \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (3.6.2)$$

可得到线性方程组

$$\begin{bmatrix} 0 & y^T \\ y & Q + r^{-1}I \end{bmatrix}_{(l+1) \times (l+1)} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ e \end{bmatrix} \quad (3.6.3)$$

其中, $e \in R^l$ 是元素都为 1 的向量, I 为 l 阶单位矩阵, $\alpha = [\alpha_1, \dots, \alpha_l]^T \in R^l$, $y = [y_1, \dots, y_l]^T \in R^l$ $Q = [q_{ij}]_{l \times l}$ 是一个半正定矩阵, $q_{ij} = y_i y_j K(x_i, x_j)$ 。式 (3.6.3) 用最小二乘法即可求解。在 $LS-SVM$ 算法中, 将二次规划问题转变成线性方程组的求解, 简化了计算复杂度。文献[93]针对 $LS-SVM$ 提出了加权 $LS-SVM$ 算法。

§ 3.7 超球面支持向量机分类算法

超球面与超平面的区别在于: 超平面只是将两类样本分开, 由于超平面把空间一分为二, 两边的地位都是相等的, 对于第三类样本无法做出相应处理; 而超球面不仅可以分开两类, 而且超球面把空间分成地位不相等的两块, 对第三类样本来说, 处在超球内部和外部是不一样的。进一步而言, 第三类样本对训练的两类样本来说是未知样本, 其在空间上的分布是不定的, 因此可以假设成全空间分布。那么, 第三类样本处在超平面两侧的任何一侧的概率是相等的, 这是因为在构建这个超平面时, 并没有考虑第三类样本的存在。而处在超球的内部和外部的概率则是可以不一样的。通过控制超球的大小和范围, 使超球的含义不仅仅是分开两类, 而且还有把球里面的样本尽量包“牢”和包“纯”, 拒绝其它类样本的进入^[118, 119]。

与支持向量机的思想类似, 要求这个超球面的构造在两个条件中折中考虑: 一是要求将两类样本尽可能正确分开; 二是要求超球的半径尽可能的小, 则其它未知类别的样本进入该超球领域的可能性就越小。这种思想也称为支持向量的区域描述 (support vector domain description, 即 SVDD)。这是因为这个超球面就是一个区域, 而这个区域是由训练出来的支持向量支撑起来的。

目前超球面算法主要有两种，一种是面向单类别样本的 One-class SVM 算法^[118]，另一种是面向二类别样本的算法^[120]。前者是后者的特例。

3.7.1 One-class 支持向量机分类算法

One-class SVM 最初是用于高维分布估计，即用来寻找超平面 VC 维的估计值^[55]。这样，支持向量机不但可以实现二值分类以及回归问题，同时还可以实现这种特殊的一类分类问题，不妨将其称为一类支持向量机(one-class Support Vector Machine, one-class SVM)。异常值检测实际上可视为一类特殊的分类问题，被称为一类分类^[49]。One-class SVM 有两种实现途径^[118]，下面给出通过超球体来实现一类分类的方法。

设一个正类样本集为：

$$\{x_i, i = 1, \dots, l\}, x_i \in R^d$$

设法找一个以 a 为中心，以 R 为半径的能够包含所有样本点的最小球体。如果直接进行优化处理，所得到的优化区域就是一个超球体。为了使优化区域更紧致，这里仍然采用核映射思想，首先用一个非线性映射 φ ，将样本点映射到高维特征空间，然后在高维特征空间中求解包含所有样本点的最小超球体。为了允许一些数据点存在误差，可以引入松弛变量 ξ_i 来控制，同时将高维空间优化中的内积运算采用满足 Mercer 条件的核函数代替，即找一个核函数 $K(x, y)$ ，使得 $K(x, y) = (\varphi(x), \varphi(y)) = \varphi(x) \cdot \varphi(y)$ ，于是优化问题为：

$$\min R^2 + C \sum_{i=1}^l \xi_i \quad (3.7.1)$$

约束条件：

$$[\varphi(x_i) - a][\varphi(x_i) - a]^T \leq R^2 + \xi_i, i = 1, \dots, l \quad (3.7.2)$$

$$\xi_i \geq 0, i = 1, \dots, l \quad (3.7.3)$$

其对偶形式为：

$$\max \sum_{i=1}^l \alpha_i K(x_i, x_i) - \sum_{i,j=1}^l \alpha_i \alpha_j K(x_i, x_j) \quad (3.7.4)$$

约束条件：

$$\sum_{i=1}^l \alpha_i = 1 \quad ((3.7.5)$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, l \quad (3.7.6)$$

解优化问题 (3.7.4) 可以得到 α 的值，通常大部分 α_i 将为零，不为零的 α_i 所对应的样本仍然被称为支持向量。

根据 KKT 条件，对应于 $0 \leq \alpha_i \leq C$ ($i = 1, 2, \dots, l$) 的样本满足：

$$R^2 - [K(x_i, x_i) - 2 \sum_{j=1}^l \alpha_j K(x_j, x_i) + a^2] = 0 \quad (3.7.7)$$

其中， $a = \sum_{i=1}^l \alpha_i \varphi(x_i)$ ，因此，用任意一个支持向量根据式 (3.7.7) 可求出 R 的值。

对于新样本 z ，设

$$\begin{aligned} f(z) &= [\varphi(z) - a][\varphi(z) - a]^T \\ &= K(z, z) - 2 \sum_{i=1}^l \alpha_i K(z, x_i) + \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j K(x_i, x_j) \end{aligned} \quad (3.7.8)$$

若 $f(z) \leq R^2$ ，则 z 为正常点，否则 z 为异常点。

3.7.2 超球面二值分类算法

对于二值分类算法，文献[119]将超球面 SVM 算法的优化问题描述为：

$$\min R^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i \quad (3.7.9)$$

约束条件:

$$y_i[R^2 - (x_i - a)^T(x_i - a)] \geq -\xi_i, i = 1, \dots, l \quad (3.7.10)$$

$$\xi_i \geq 0, i = 1, \dots, l \quad (3.7.11)$$

式中 $y_i \in \{-1, 1\}$ 。采用拉格朗日乘子法求解上述优化问题, 得到其对偶表达式为:

$$\max \sum_{i=1}^l \alpha_i y_i (x_i \cdot x_i) - \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i, x_j) \quad (3.7.12)$$

约束条件:

$$\sum_{i=1}^l y_i \alpha_i = 1 \quad (3.7.13)$$

$$0 \leq \alpha_i \leq \frac{1}{cl}, i = 1, \dots, l \quad (3.7.14)$$

对于非线性的情况, 利用核函数得到的决策函数为

$$f(z) = \text{sgn}[R^2 - K(z, z) + 2 \sum_{i=1}^l \alpha_i y_i K(z, x_i) - \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j)] \quad (3.7.15)$$

§ 3.8 各种改进的支持向量机算法的比较

在以往的传统学习方法中采用的经验风险最小化准则 (ERM), 虽然可以使训练误差最小化, 但并不能最小化学习过程的推广误差, 如神经网络的学习问题。因此, Vapnik 和 Chervonenkis 提出了结构风险最小化准则 (SRM)^[46], 即通过对推广误差上界的最小化达到最大的推广能力。结构风险最小化准则的基本思想是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推广能力。Vapnik 提出的支持向量机就是这种思想的具体实现。支持向量机的基本思想是在样本空间或特征空间, 构造出最优超平面使不同类样本集之间的距离最大, 从而达到最大的推广能力。其约束条件约束了 w 和

b , 使得经验误差为零, 同时最小 $\|w\|^2$ 使 VC 维最小, 因此, 它的最优化体现了结构风险最小化准则。另外, 因为其变形算法主要是通过增加函数项, 变量或系数等方法使公式变形, 它们满足结构风险最小化原则。

在 C -SVM 算法中, 唯一可以调节的参数是常数 C , 它对 SVM 的性能有重要影响。然而, 由于不能给出 C 的直观解释, 这样在实际应用中很难选择合适的 C 。

在 ν -SVM 算法中 C 被参数 ν 取代了。文献[121]分析了 ν -SVM, 认为它是带有上下界约束和一个简单的等式约束的二次规划问题, 可以用现有的解 C -SVM 算法的方法来求解。主要优点是: (1) 使用一个参数 ν 就可以控制支持向量的个数及误差; (2) 没有 C , 避免了数值计算的麻烦; (3) ν 参数是边缘误差值与样本数之比的上界, 支持向量数与样本数之比的下界。 C -SVM 中 C 的增加导致支持向量的减少, 而 ν -SVM 中 ν 的减少导致支持向量的减少。

在 W -SVM 算法中, 通过给不同类别的样本设置一定比例关系的权值就可有效地补偿类别大小差异对分类的影响。例如在两种类别的情况下, 正类和负类 C -SVM 引入惩罚系数 C , 实行对错误分类的惩罚。然而实际应用中, 某些重要样本正确分类的要求高, 而某些样本正确分类的要求低, 因此, 在优化问题描述中, 对每个采样点数据采用不同的惩罚系数, 以得到更准确的分类, 这就是加权支持向量机算法的优势所在, 克服了 C -SVM 算法在这方面的缺陷。

LS -SVM 算法和 C -SVM 算法的主要区别是只有等式约束, 而没有 C -SVM 的不等式约束, 从而推出不同的一系列的等式约束, 不是二次规划问题。即将二次规划问题转变成线性方程组的求解, 简化了计算复杂性。

线性规划下的支持向量机算法在一定程度上解决了计算量的问题^[122-126]。支持向量机一类分类、二值分类以及回归算法都是把问题归结为一个二次规划来求解, 其中参数的数量在分类情况下等于训练样本的个数, 在回归情况下是训练样本个数的二倍。当数据量很大时, 其计算的时间和空间复杂度均很大。若

能将支持向量机算法归结为线性规划来求解无疑会大大减少计算量。

One-class SVM 算法适用于只对某一类感兴趣，其他类信息不全的情况。引进超球面的概念，和其它算法的超平面有很大的区别。异常值检测实际上可视为一类特殊的分类问题。但是，目前的超球面支持向量机算法从支持向量机理论的基础来说是有损失的。它们的目标函数中缺少了使分类间隔尽量大这个条件，而这个条件恰恰是统计学习理论中结构风险最小化的体现，反映了算法的推广能力。

§ 3.9 本章小结

本章介绍了支持向量机理论中的各种变形的支持向量算法，讨论了各自的优缺点和适用范围等情况，如 ν -SVM 算法使用新参数 ν 来控制支持向量的数目和误差，比 C 参数易于选择。 W -SVM 适用于需要对每一个样本区别对待的情况下，有一定的抗噪声作用。 LS -SVM 算法没有 C -SVM 的不等式约束，把二次规划问题转变成线性方程组的求解，简化了计算复杂性。**One-class SVM** 适用于只对某一类感兴趣，其它类信息不全的情况。由于引进了超球面的概念，和其它算法的超平面有很大的区别。这些算法的比较分析有利于研究者对目前各种变形算法有清楚全面的认识，有利于进一步深入研究。

第四章 支持向量回归机的基本理论与算法研究

支持向量机以其出色的学习性能已经在一些领域得到广泛应用，随着其应用领域的不断扩大以及在实际应用中遇到的问题，也反过来要求支持向量机本身不断完善和发展。**SVM** 的方法最早是针对模式识别问题提出的，随着 Vapnik 对 ε -不敏感损失函数的引入，已将其推广应用到非线性回归估计和曲线拟合中，得到了用于曲线拟合的支持向量回归机(Support Vector Machine for Regression, **SVR**)^[84]，并且表现出很好的学习效果。

在用于回归估计的标准 **SVR** 学习算法中，学习的目的在于构造一个回归估计函数 $f(x)$ ，使它与目标值的距离小于 ε ，同时函数的 VC 维最小，从而将线性或非线性函数 $f(x)$ 的回归估计问题，转化为一个具有线性等式约束和线性不等式约束的二次规划问题，可以得到唯一的全局最优解。同时它还具有许多优越性，例如，以可控的精度逼近非线性函数；良好的泛化能力等。目前各类 **SVR** 算法已经得到了广泛关注和研究，它在非线性系统辨识、预测预报、建模与控制等领域都有潜在的广泛应用，使得对其进行研究显得非常重要。但是其理论体系和学习算法的实现仍然有大量问题有待于发展和完善，如何设计快速有效的算法是 **SVR** 实际应用中的瓶颈问题，而且在应用方面的研究也相当缺乏。

本章重点阐述用于曲线拟合与回归估计的支持向量回归机的基本原理简介，然后详细地阐述并分析了各种支持向量回归机模型，包括 ε -支持向量回归机、 ν -支持向量回归机、LS-支持向量回归机、W-支持向量回归机、基于线性规划的支持向量回归机，同时，建立了在 ε -不敏感函数下的单参数约束下的支持向量回归模型，论证了单参数回归模型与标准回归模型的等价性，并进行数值实验；

最后探讨并建立了支持向量回归算法优劣的评价依据，从而完善支持向量回归机的优化理论基础和决策依据。

§ 4.1 支持向量回归机的基本原理简介^[84]

4.1.1 基本思想

回归分析是找若干变量之间的统计关系的一种方法，利用所找的统计关系对某一变量作出未来某一时刻的估计，即预报值。

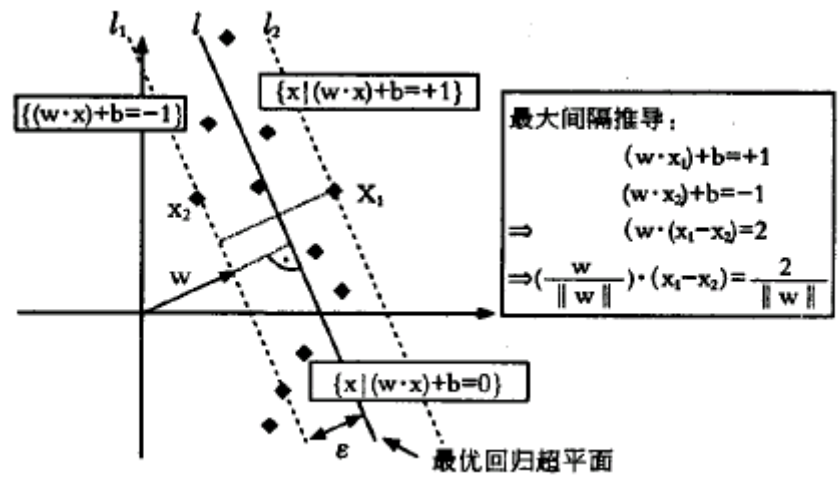


图 4.1 最优回归超平面和最大间隔

SVM 方法的核心思想之一是求解两类分类问题的最优分类面，这引出支持向量的概念——SVM 方法的另一核心思想是通过非线性映射 ϕ 把样本集映射入一个高维(以至于无穷维)的 Hilbert 空间(称为特征空间)，使在样本空间中的高度非线性问题在高维空间中应用线性分类的方法得以实现，从而解决非线性问题。

与分类问题不同的是：回归的样本点只有一类，所寻求的最优超平面不是使两类样本点分得“最开”，而是使所有样本点离超平面的“总偏差”最小，这时样本点都在两条边界线之间，求最优回归超平面同样等价于求最大间隔

(maximal margin) (图 4. 1)。

4. 1. 2 ε -不敏感损失函数

当把支持向量机推广到回归问题建立新的回归算法---支持向量回归机 (Support Vector Regression, SVR) 时, 则需要引入合适的损失函数来保证这个重要性质的存在条件^[46]。

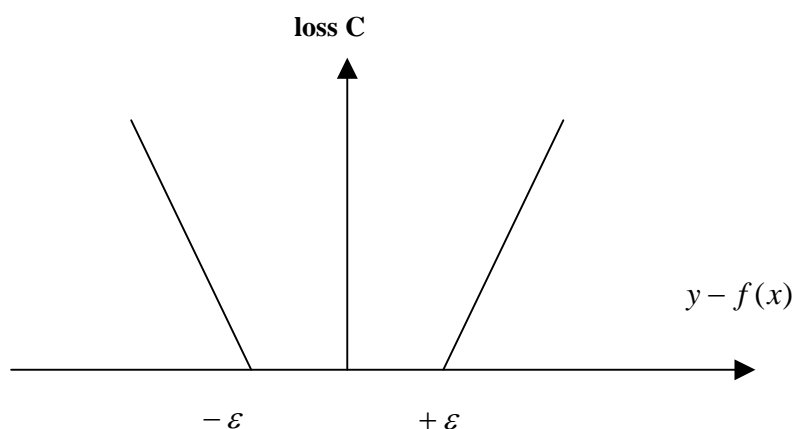


图 4. 2 ε -不敏感损失函数

SVR 采用 Vapnik 的 ε -不灵敏函数作为误差函数 (即误差小于 ε 时视为无误差)^[8]:

$$c(x, y, f(x)) = |y - f(x)|_{\varepsilon}$$

其中 $|y - f(x)|_{\varepsilon} = \max\{0, |y - f(x)| - \varepsilon\}$

这里 ε 是事先取定的一个正数, ε -不敏感损失函数的含义是, 当 x 点的观察值 y 与预测值 $f(x)$ 之差不超过事先给定的 ε 时, 则认为在该点的预测值 $f(x)$ 是无损失的, 尽管预测值 $f(x)$ 和观测值 y 可能并不完全相等。图 4. 2 中画出了损失函数的图像。

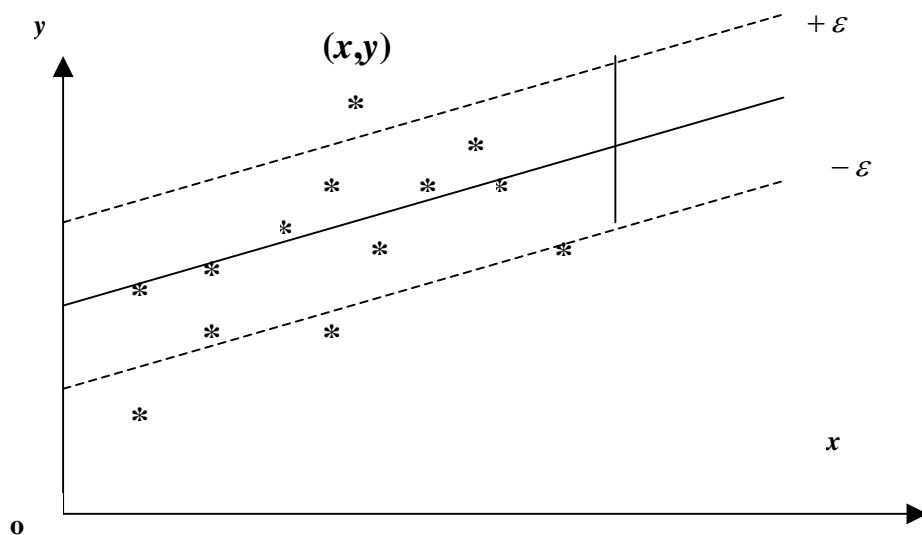


图 4.3 ε -带

ε -不敏感损失函数有一个特点：对样本点来说，存在着一个不为目标函数提供任何损失值的区域，即 ε -带（图 4.3）。这个特点是其它许多损失函数所不具备的。可以期望，在 ε -带内的样本点，不会出现在决策函数中。

§ 4.2 ε -支持向量回归机

设图 4.3 中的观测点（训练集）为

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\}$$

利用上述的 ε -不敏感损失函数，并限定在线性函数集合中的估计回归函数，基于结构风险最小化原则，当所有样本点到所求超平面的距离都可以小于 ε 时，寻求最优回归超平面的问题转化为求解如下一个二次凸规划问题：

$$\min \frac{1}{2} \|w\|^2 \quad (4.2.1)$$

约束条件：

$$\begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon \\ (w \cdot x_i) + b - y_i \leq \varepsilon \end{cases} \quad (4.2.2)$$

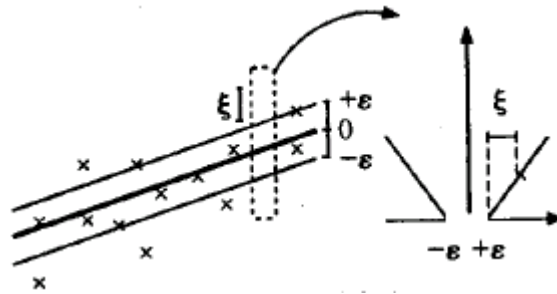


图 4.4 有偏差点的最优回归超平面图示

下面重点研究非线性情形^[84, 85]。与模式分类一样，通过非线性情况下，可以把样本 x 通过非线性映射 $\varphi(x)$ 映射到高维特征空间 H ，并在 H 中求解最优回归函数。这样，在高维空间中的线性回归，就对应于低维空间中的非线性回归。因此，在最优回归函数中采用适当的核函数 $K(x_i, x)$ 代替高维空间中的向量内积 $\varphi(x_i) \cdot \varphi(x)$ ，就可以实现某一非线性变换后的线性拟合，而计算复杂度却没有增加。当个别样本点到所求超平面的距离大于 ε 时，引入松弛变量 ξ_i ， ξ_i^* (图 4.4)，构造容错惩罚系数 C ，此时最优化问题转化为：

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (4.2.3)$$

约束条件

$$y_i - w \cdot \varphi(x_i) - b \leq \varepsilon + \xi_i \quad (4.2.4)$$

$$w \cdot \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (4.2.5)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, l \quad (4.2.6)$$

对于最优化问题(4.2.3)，由于目标函数和约束条件都是凸集，根据线性规划理论，它们都存在唯一的全局最小解。引入拉格朗日（Lagrange）函数：

$$L(w, b, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i [\varepsilon + \xi_i - y_i + w \cdot \varphi(x_i) + b] - \sum_{i=1}^l \alpha_i^* [\varepsilon + \xi_i^* + y_i - w \cdot \varphi(x_i) - b] - \sum (\eta_i \xi_i + \mu_i^* \xi_i^*) \quad (4.2.7)$$

$$\text{令 } \frac{\partial L}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (4.2.8)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.2.9)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \eta_i = 0 \quad (4.2.10)$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \quad \Rightarrow \quad C - \alpha_i^* - \eta_i^* = 0 \quad (4.2.11)$$

将(4.2.8)—(4.2.11)代入(4.2.7)，得到其对偶形式(4.2.12)：

$$\max \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \varphi(x_i) \cdot \varphi(x_j) \quad (4.2.12)$$

约束条件：

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.2.13)$$

$$\alpha_i, \alpha_i^* \in [0, C] \quad (4.2.14)$$

根据 KKT 条件，式(4.2.12)–(4.2.14)的优化问题的解在最优点处必须满

足：

$$\alpha_i[\varepsilon + \xi_i - y_i + w \cdot \varphi(x_i) + b] = 0 \quad (4.2.15)$$

$$\alpha_i^*[\varepsilon + \xi_i^* + y_i - w \cdot \varphi(x_i) - b] = 0 \quad (4.2.16)$$

$$\eta_i \xi_i = 0 \Rightarrow (C - \alpha_i) \xi_i = 0 \quad (4.2.17)$$

$$\eta_i^* \xi_i^* = 0 \Rightarrow (C - \alpha_i^*) \xi_i^* = 0 \quad (4.2.18)$$

由(4.2.15)和(4.2.16)式，可得到：

$$\alpha_i \times \alpha_i^* = 0 \quad (4.2.19)$$

故， α_i, α_i^* 的取值包含以下几种情况之一：

- ① $\alpha_i = 0, \alpha_i^* = 0$ ；
- ② $\alpha_i = C, \alpha_i^* = 0$ ；
- ③ $\alpha_i = 0, \alpha_i^* = C$ ；
- ④ $0 < \alpha_i < C, \alpha_i^* = 0$ ；
- ⑤ $\alpha_i = 0, 0 < \alpha_i^* < C$ ；

在②～⑤中， $(\alpha_i - \alpha_i^*)$ 不为零，所对应的 x_i 称为支持向量（Support Vector, SV），

其中，②和③对应的 x_i 称边界为支持向量（Boundary Support Vector, BSV），

④和⑤对应的 x_i 称为标准支持向量（Normal Support Vector, NSV）。由

式(4.2.15)～(4.2.19)，我们可以分析得到：

若 $0 < \alpha_i < C$ ，则必有 $\alpha_i^* = 0, \xi_i = 0$ ，那么

$$\varepsilon + \xi_i - y_i + w \cdot \varphi(x_i) + b = 0 \quad (4.2.20)$$

若 $0 < \alpha_i^* < C$ ，则必有 $\alpha_i = 0, \xi_i^* = 0$ ，那么

$$\varepsilon + y_i - w \cdot \varphi(x_i) - b = 0 \quad (4.2.21)$$

综合(4.1.20)、(4.1.21)式，可以得出：

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_i < C} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) \varphi(x_j) \cdot \varphi(x_i) - \varepsilon] \right. \\ \left. + \sum_{0 < \alpha_i^* < C} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) \varphi(x_j) \cdot \varphi(x_i) - \varepsilon] \right\} \quad (4.2.22)$$

上式中 SV 为标准支持向量机集合， N_{NSV} 为标准向量机数量。

从而所求的最优超平面线性回归函数为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \cdot \varphi(x) + b \quad (4.2.23)$$

§ 4.3 ν -支持向量回归机

在标准 SVR 算法中，我们通过设定 ε 的值来控制回归估计精度，但是 ε 取多少是事先无法预知的。而在 Schölkopf 和 Smola 提出的 ν -SVR 算法中^[53, 57, 59]，引入反映超出 ε 管道之外的样本数据点，即边界支持向量数量和支持向量数的新参数 ν ，简化了 SVR 的参数调节。

同样假设回归估计函数为：

$$f(x) = w \cdot \varphi(x) + b \quad (4.3.1)$$

ν -SVR 中的最优化问题为：

$$\min \quad \frac{1}{2} \|w\|^2 + C[\nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*)] \quad (4.3.2)$$

约束条件：

$$y_i - w \cdot \varphi(x_i) - b \leq \varepsilon + \xi_i \quad (4.3.3)$$

$$w \cdot \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (4.3.4)$$

$$\xi_i, \xi_i^* \geq 0, \quad i=1,2,\dots,l \quad (4.3.5)$$

构造如下拉格朗日函数:

$$L(w, b, \xi, \xi^*, \varepsilon) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i [\varepsilon + \xi_i - y_i + w \cdot \varphi(x_i) + b] \\ - \sum_{i=1}^l \alpha_i^* [\varepsilon + \xi_i^* + y_i - w \cdot \varphi(x_i) - b] - \sum (\eta_i \xi_i + \mu_i^* \xi_i^*) - \gamma \varepsilon \quad (4.3.6)$$

$$\text{令 } \frac{\partial L}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (4.3.7)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.3.8)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \eta_i = 0 \quad (4.3.9)$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \quad \Rightarrow \quad C - \alpha_i^* - \eta_i^* = 0 \quad (4.3.10)$$

$$\frac{\partial L}{\partial \varepsilon} = 0 \quad \Rightarrow \quad \sum_{i=1}^l (\alpha_i + \alpha_i^*) = Cv - \gamma \quad (4.3.11)$$

将(4.3.7)—(4.3.11)代入(4.3.6),得到其对偶形式(4.3.12):

$$\max \quad \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i \cdot x_j) \quad (4.3.12)$$

约束条件:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.3.13)$$

$$0 \leq \alpha_i \leq \frac{C}{l} \quad (4.3.14)$$

$$0 \leq \alpha_i^* \leq \frac{C}{l} \quad (4.3.15)$$

$$\sum_{i=1}^l (\alpha_i + \alpha_i^*) \leq Cv \quad (4.3.16)$$

从(4.3.12)—(4.3.16)式，我们不难发现在对偶最优化问题的求解过程中不需要 ε 的值，因此不需要预先规定 ε 取多少，简化了 SVR 学习过程中的参数调节问题。

通过学习训练得到的回归估计函数为：

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i \cdot x) + b \quad (4.3.17)$$

类似标准 SVR 算法，我们可以求出：

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_i < C/l} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) - \varepsilon] \right. \\ \left. + \sum_{0 < \alpha_i < C/l} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) + \varepsilon] \right\} \quad (4.3.18)$$

上式中 N_{NSV} 为标准支持向量数量。

§ 4.4 LS-支持向量回归机

J.A.K.Suykens 提出的最小二乘支持向量机算法^[36, 88, 93]，采用最小二乘线性系统代替标准 SVR 算法，用二次规划方法解决模式识别和函数估计问题。在 LS-SVR 算法中，优化指标采用平方项，将不等式约束转变为等式约束，从而将二次规划问题转变成线性方程组的求解问题。与标准 SVR 算法相比，减少了一个调整参数，减少了多个优化变量，因此简化了计算的复杂性，提高了收敛速度。

同样假设回归估计函数为：

$$f(x) = w \cdot \varphi(x) + b \quad (4.4.1)$$

LS-SVR 中的最优化问题为:

$$\min \quad \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma \sum_{i=1}^l \xi_i^2 \quad (4.4.2)$$

约束条件:

$$\begin{cases} y_i - [w \cdot \varphi(x_i) + b] = \xi_i, & i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \quad (4.4.3)$$

构造如下拉格朗日函数:

$$L(w, b, \xi) = \frac{1}{2}\|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i [y_i - w \cdot \varphi(x_i) - b - \xi_i] \quad (4.4.4)$$

$$\text{令 } \frac{\partial L}{\partial w} = 0 \quad \Rightarrow \quad w + \sum_{i=1}^l \alpha_i \varphi(x_i) = 0 \quad (4.4.5)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^l \alpha_i = 0 \quad (4.4.6)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad \alpha_i + \gamma \xi_i = 0 \quad (4.4.7)$$

又由(4.4.4)式的约束条件, 有:

$$y_i - [w \cdot \varphi(x_i) + b] = \xi_i,$$

由以上四式得到线性方程组:

$$\begin{pmatrix} I & 0 & 0 & Z^T \\ 0 & 0 & 0 & I \\ 0 & 0 & \gamma I & I \\ Z^T & I & I & 0 \end{pmatrix} \cdot \begin{pmatrix} w \\ b \\ \xi \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ Y \end{pmatrix} \quad (4.4.8)$$

其中 $Z = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_l)]$,

$$I = [1, 1, \dots, 1],$$

$$Y = [y_1, y_2, \dots, y_l],$$

则通过学习得到的回归估计函数为：

$$f(x) = \sum_{x_i \in SV} \alpha_i K(x_i \cdot x) + b \quad (4.4.9)$$

其中 α , b 为(4.4.8)的解。

§ 4.5 W-支持向量回归机

在支持向量回归中，如果样本来自于同一分布且是独立的，也就是说随机误差项具有相同的方差并且不相关，此时将具有很好的建模效果。但由于实际问题是错综复杂的，因而在建立实际问题的回归模型时，经常会出现某一因素随着解释变量观测值的变化而对被解释变量产生不同的影响，导致随机误差项产生不同的方差。当一个回归问题存在异方差性时，如果仍然用标准的支持向量回归算法来建立模型，往往达不到好的效果。为了解决这个问题，本节提出一种加权支持向量回归算法^[53, 112]，仿真实验表明，当数据中存在异方差性时，加权回归方法能够获得更好的预测效果。

下面介绍加权支持向量回归机的算法理论。

假设回归估计函数为：

$$f(x) = w \cdot \varphi(x) + b \quad (4.5.1)$$

在加权支持向量机算法中，参数 C 和 ε 是针对各个样本来选择的，其最优化问题为：

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l s_i (\xi_i + \xi_i^*) \quad (4.5.2)$$

约束条件：

$$y_i - w \cdot \varphi(x_i) - b \leq \varepsilon \cdot t_i + \xi_i \quad (4.5.3)$$

$$w \cdot \varphi(x_i) + b - y_i \leq \varepsilon \cdot t_i + \xi_i^* \quad (4.5.4)$$

$$\xi_i, \xi_i^* \geq 0, \quad i=1,2,\dots,l \quad (4.5.5)$$

其中 s_i, t_i 分别表示第 i 个学习样本对参数 C 和 ε 的加权系数。

构造如下拉格朗日函数

$$L(w, b, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l s_i (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i [\varepsilon \cdot t_i + \xi_i - y_i + w \cdot \varphi(x_i) + b] - \sum_{i=1}^l \alpha_i^* [\varepsilon \cdot t_i + \xi_i^* + y_i - w \cdot \varphi(x_i) - b] - \sum_{i=1}^l (\eta_i \xi_i + \mu_i^* \xi_i^*) \quad (4.5.6)$$

$$\text{令 } \frac{\partial L}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (4.5.7)$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.5.8)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C \cdot s_i - \alpha_i - \eta_i = 0 \quad (4.5.9)$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \quad \Rightarrow \quad C \cdot s_i - \alpha_i^* - \eta_i^* = 0 \quad (4.5.10)$$

将(4.5.7)—(4.5.10)代入(4.5.6),得到其对偶形式(4.5.11):

$$\max \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i \cdot x_j) - \varepsilon \sum_{i=1}^l t_i (\alpha_i + \alpha_i^*) \quad (4.5.11)$$

约束条件:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.5.12)$$

$$0 \leq \alpha_i \leq C s_i \quad (4.5.13)$$

$$0 \leq \alpha_i^* \leq Cs_i \quad (4.5.14)$$

最后得到的回归估计函数为：

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i \cdot x) + b \quad (4.5.15)$$

类似标准 SVR 算法，我们可以求出：

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_i < Cs_i} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) - \varepsilon \sum_{x_j \in SV} t_j] \right. \\ \left. + \sum_{0 < \alpha_i^* < Cs_i} [y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) + \varepsilon \sum_{x_j \in SV} t_j] \right\} \quad (4.5.16)$$

上式中 N_{NSV} 为标准支持向量数量。

§ 4.6 基于线性规划的支持向量回归机

在标准 SVR 算法中，优化目标是一个二次范数，也正因为这个二次范数，才将最优化问题转化成了二次规划问题来求解。为此 Kecman 和 Hadzic^[107-109] 提出将优化目标用一次范数替代二次范数，从而将最优化问题转化成线性规划问题来求解，简化了计算的复杂性。

其最优化问题为：

$$\min \sum_{i=1}^l (\alpha_i + \alpha_i^*) + C \sum_{i=1}^l \xi_i \quad (4.6.1)$$

约束条件：

$$y_i \left[\sum_{j=1}^l (\alpha_j - \alpha_j^*) K(x_j, x_i) + b \right] \geq 1 - \xi_i, \quad (4.6.2)$$

$$\alpha_i, \alpha_i^*, \xi_i \geq 0, i = 1, \dots, l \quad (4.6.3)$$

与 ν -支持向量分类机相对应的线性规划问题为^[117]：

$$\min \quad \frac{1}{l} \sum_{i=1}^l \xi_i - \nu \rho \quad (4.6.4)$$

约束条件:

$$\frac{1}{l} \sum_{i=1}^l (\alpha_i + \alpha_i^*) = 1 \quad (4.6.5)$$

$$y_i \left[\sum_{j=1}^l (\alpha_j - \alpha_j^*) K(x_j, x_i) + b \right] \geq \rho - \xi_i \quad (4.6.6)$$

$$\alpha_i, \alpha_i^*, \xi_i \geq 0, i = 1, \dots, l; \rho \geq 0 \quad (4.6.7)$$

求解这些优化问题, 可得相应的决策函数。

§ 4.7 单参数约束下的支持向量回归机

4.7.1 单参数回归模型

标准的支持向量回归算法(4.2.1)、(4.2.2)引入两个参数 ξ, ξ^* 来控制误差的大小, 如优化方程(4.2.3)。本节采用一个参数来控制误差, 进而给出单参数约束下的支持向量回归算法^[144]。

设给定的训练样本为:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subset R^n \times R$$

同样采用一个非线性映射 φ 把数据映射到一个高维特征空间, 然后在高维特征空间中进行线性回归。设回归函数为:

$$f(x) = w \cdot \varphi(x) + b \quad (4.7.1)$$

在 ε -不敏感损失函数下, 采用一个参数来控制误差项, 可得下面的优化问题:

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (4.7.2)$$

约束条件:

$$|f(x) - y_i| \leq \varepsilon + \xi_i, \quad i = 1, \dots, l \quad (4.7.3)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (4.7.4)$$

将优化(4.2.3)也看作是经过 φ 映射以后在高维空间中的回归优化, 则有下面的定理。

定理 4.1 优化问题(4.7.2)与优化问题(4.2.3)是等价的。

证明: 令 $d_i = w \cdot \varphi(x_i) + b - y_i$, 并用符号 x_+ 表示 $\max\{x, 0\}$ 。当优化问题(4.2.3)取最优解时, 有

$$\xi_i = \max\{-d_i - \varepsilon, 0\} = (-d_i - \varepsilon)_+,$$

和

$$\xi_i^* = \max\{d_i - \varepsilon, 0\} = (d_i - \varepsilon)_+,$$

则优化问题(4.2.3)等价于

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l ((-d_i - \varepsilon)_+ + (d_i - \varepsilon)_+)$$

从而可进一步写为:

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (|d_i| - \varepsilon)_+,$$

另一方面, 当优化问题(4.7.2)取最优解时, 有

$$\xi_i = \max\{d_i - \varepsilon, -d_i - \varepsilon, 0\}$$

$$= \max\{|d_i| - \varepsilon, 0\}$$

$$= (|d_i| - \varepsilon)_+,$$

因此, 优化问题(4.7.2)等价与

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (|d_i| - \varepsilon)_+$$

于是证明了这两个模型的等价性。

下面给出优化问题(4.7.2)的对偶形式, 引入 Lagrange 函数:

$$\begin{aligned} L(w, b, \xi, \alpha, \alpha^*, \gamma) = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [\xi_i + \varepsilon - y_i + f(x_i)] \\ & - \sum_{i=1}^l \alpha_i^* [\xi_i + \varepsilon + y_i - f(x_i)] - \sum_{i=1}^l \gamma_i \xi_i \end{aligned} \quad (4.7.5)$$

其中 $\alpha_i, \alpha_i^*, \gamma_i \geq 0, i=1, \dots, l$ 。

函数 L 的极值应满足条件:

$$\frac{\partial L}{\partial w} = 0, \quad \frac{\partial L}{\partial b} = 0, \quad \frac{\partial L}{\partial \xi_i} = 0 \quad (4.7.6)$$

得到下面的式子:

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (4.7.7)$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (4.7.8)$$

$$C - \alpha_i - \alpha_i^* - \gamma_i = 0 \quad (4.7.9)$$

将(4.7.7)-(4.7.9)代入到 Lagrange 函数(4.7.5)中, 得到优化问题的对偶形式:

$$\begin{aligned} \max \quad W(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ & + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (4.7.10)$$

约束条件为:

$$\begin{aligned}
\sum_{i=1}^l (\alpha_i - \alpha_i^*) &= 0 \\
\alpha_i + \alpha_i^* &\leq C, i=1, \dots, l \\
\alpha_i, \alpha_i^* &\geq 0, i=1, \dots, l
\end{aligned} \tag{4.7.11}$$

上面的优化函数同支持向量回归函数是一样的，不同的就是约束条件。解出 α 的值，可得 $f(x)$ 的表达式如下：

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \tag{4.7.12}$$

按照 KKT 条件，在鞍点有：

$$\alpha_i [\xi_i + \varepsilon - y_i + f(x_i)] = 0, \quad i=1, \dots, l \tag{4.7.13}$$

$$\alpha_i^* [\xi_i + \varepsilon + y_i - f(x_i)] = 0, \quad i=1, \dots, l \tag{4.7.14}$$

$$\gamma_i \xi_i = 0, \quad i=1, \dots, l \tag{4.7.15}$$

由(4.7.13)和(4.7.14)式知： $\alpha_i \alpha_i^* = 0$ ，即任何一组 α_i 和 α_i^* 都不会同时为非零，再由(4.7.9)和(4.7.15)得到

$$(C - \alpha_i - \alpha_i^*) \xi_i = 0, \quad i=1, \dots, l \tag{4.7.16}$$

由(4.7.16)可知，只有对应于 $\alpha_i + \alpha_i^* = C$ 的 $f(x_i)$ 与 y_i 的误差可能小于 ε ，因此有

$$\varepsilon - y_i + f(x_i) = 0, \text{ 对于 } \alpha_i + \alpha_i^* < C \text{ 且 } \alpha_i > 0 \tag{4.7.17}$$

$$\varepsilon + y_i - f(x_i) = 0, \text{ 对于 } \alpha_i + \alpha_i^* < C \text{ 且 } \alpha_i^* > 0 \tag{4.7.18}$$

由以上两式可以求出 b 。

4.7.2 标准与单参数支持向量回归机的约束条件比较

由前面的推导可知，得到同样的优化式子，所不同的就是约束条件。单参

数回归下的约束条件为(4.7.11)，标准回归下的约束条件为(4.2.13)和(4.2.14)。为了说明方便，将单参数回归下的约束条件记为1-S.T，标准回归下的约束条件记为2-S.T。

在约束条件1-S.T中，用 $2C$ 代替 C ，那么显然满足2-S.T的条件一定满足1-S.T的约束条件，说明1-S.T的约束条件比2-S.T的条件宽，因此1-S.T约束下的最优解大于等于2-S.T约束下的最优解，后面的实验也说明了这一点。

如果两种约束条件取同样的 C 值，那么满足1-S.T的条件一定满足2-S.T的条件，说明2-S.T的条件比1-S.T的条件宽，此时有下面的定理^[127]。

定理 4.2 优化问题(4.7.10)与优化问题(4.1.12)的最优解是等价的。

该定理进一步说明，在参数 C 和 ε 取同样值的条件下，单参数约束下的回归模型同标准回归模型的最优解是等价的。优化问题(4.7.10)与优化问题(4.1.12)的优化函数式相同，只是约束条件不同。后面的仿真实验将表明单参数回归算法能够在一定程度上提高程序的运行速度。

4.7.3 算法实现与计算实例

为了能用标准的二次规划工具箱实现上述算法，将优化问题(4.7.10)化为标准的二次规划形式：

$$\min \quad \frac{1}{2} x^T H x + c^T x \quad (4.7.19)$$

约束条件：

$$\begin{aligned} a^T x &= 0 \\ x &\geq 0 \\ Ax &\leq C \end{aligned} \quad (4.7.20)$$

其中参数取下面的值：

$$x = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}, \quad c = \begin{bmatrix} \varepsilon - y \\ \varepsilon + y \end{bmatrix}, \quad H = \begin{bmatrix} D & -D \\ -D & D \end{bmatrix},$$

$$a = [1 \ 1 \ \cdots \ 1 \ -1 \ -1 \ \cdots \ -1]^T,$$

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

其中, $D = K(x_i, x_j), i, j = 1, \dots, l$, a 为 $2l$ 维向量, A 为 $l \times 2l$ 维矩阵。

下面以 Henon 混沌时间序列预测为例来比较两种算法的运算结果。Henon 混沌时间序列由下式产生:

$$y_t = 1 - 1.4y_{t-1}^2 + 0.3y_{t-2}, t \in Z \quad (4.7.21)$$

取初值[0,0], 取其中的一维得到 288 个样本, 前 280 个作为训练样本, 后 8 个作为测试样本。为了对该序列进行预测, 首先要重构相空间, 这里取嵌入维数为 2。另外, 为了提高训练样本的质量, 每次预测都选择距离输入样本较近的一些样本作为训练集, 选择标准为欧氏距离小于 $\delta(>0)$, 这里取 $\delta=0.7$, 向量回归中的参数取为 $\varepsilon = 0.0001, C = 500$, 核函数取高斯函数, 其中, $\sigma^2 = 1.5$ 。

为了和标准支持向量回归算法比较, 每预测一个样本, 重新训练支持向量机预测下一个样本, 并给出相应的支持向量数及运行时间(m)。

第一个实验在单参数回归算法中取 $2C$ 代替 C , 其它参数同标准的回归算法中一样。在 PIII50 上采用 Matlab6.5 编程实现, 两种方法比较结果如表 4.1 所示。由表可见, 单参数回归方法在获得与标准回归方法几乎同样的预测精度下, 支持向量数少, 运行时间短。

第二个实验采用同样的 C 值, 用同样数据进行上面实验, 单参数回归和标准回归所得预测值相同, 支持向量数相同, 如表 4.1 中的标准回归结果, 而运行时间不同。为了节省篇幅, 这里只给出平均运行时间, 单参数回归的平均运行时间为 1.8511(m), 标准回归平均运行时间为 3.9256(m)。

表 4.1 标准回归算法与单参数回归算法的比较结果

实测值	标准回归 预测值	单参数回归 预测值	标准回归 支持向量数	单参数回归 支持向量数	标准回归 运行时间	单参数回归 运行时间
1.2110	1.2108	1.2108	30	19	2.5245	1.5607
-1.0020	-1.0020	-1.0020	29	23	2.7182	1.5699
-0.0423	-0.0420	-0.00421	28	17	0.3188	0.2102
0.6969	0.6967	0.6968	26	22	0.1025	0.0995
0.3074	0.3075	0.3073	35	26	10.5993	4.8077
1.0768	1.0767	1.0767	31	20	2.7927	1.7615
-0.5311	-0.5311	-0.5311	29	21	9.8230	2.0808
0.9282	0.9281	0.9281	27	21	2.5260	0.4328

§ 4.8 支持向量回归算法中的预测信任度

支持向量机一个最有吸引力的特征就是不需要对样本数据分布提出更多的假设，这一优点也同时给支持向量机带来一定的不足，就是它只有点输出，没有概率输出^[144]。比如，用支持向量回归算法建立起的预测模型，通常把待预测样本输入，只能获得一个点输出，至于这个预测值到底有多大的可信度并不知道。在对风险敏感有较高要求的领域中，如医疗诊断，自信信息显得尤为重要。于是人们设法将贝叶斯方法引入到支持向量机中，赋予支持向量机一定的概率特性。这些方法的给出也同时提出新的问题，比如需要事先给出数据的先验概率，如果先验假设不能真实地反映实际分布，将会导致无效的结果。

本节利用支持向量回归算法较好的稳健性，并根据局部预测的思想，对每个待预测输入样本都找到它的一个邻近集，然后以这个邻近集作为训练集，并且根据这个邻近集定义待预测值的预测信任度，这样就使训练样本质量高而且数量少，从而能提高程序的运行速度，而且对每个待预测值都能提供一个预测

信任度, 预测信任度能够在一定程度上反映数据中的噪声 (即误差) 含量。

设总的训练集为 $\{x_i, y_i\}_{i=1}^M$, M 表示训练数据的个数, $x_i \in R^m$ 是输入数据, $y_i \in R$ 是输出数据。回归的目的就是根据训练样本建立回归模型, 从而对新样本 x_{M+1} , 获得其预测值 \hat{y}_{M+1} 。当样本数 M 很大时, 用整个训练样本集建立预测模型, 不但使模型复杂, 而且往往不能获得好的预测效果。此时, 采用局部预测方法不但能够简化运算, 而且还能够提高预测精度。

首先给定一个适当的正数 δ , 选择到 x_{M+1} 的距离小于等于 δ 的样本作为训练集 (设数量为 l), 然后根据支持向量回归算法得到回归函数 $f(x)$ 。给出能够接受的误差 $e > 0$, 预测值 \hat{y}_{M+1} 的 e -预测信任度定义为:

$$R_e(\hat{y}_{M+1}) = \frac{\sum_{i=1}^l I\{|y_i - f(x_i)| \leq e\}}{l} \quad (4.8.1)$$

其中 $I(A) = \begin{cases} 1, & \text{若 } A \text{ 为真} \\ 0, & \text{否则} \end{cases}$ 。

该定义可以被理解为预测值 $\hat{y}_{M+1} \in [y_{M+1} - \varepsilon, y_{M+1} + \varepsilon]$ 的可信度为 $R_e(\hat{y}_{M+1})$, 这样对预测值的可信性可以有一定的了解。

可信度随着 e 的变化而变化, 如果当 e 相对较小时, 可信度还很高, 说明数据中几乎没有噪声, 若可信度比较小, 说明数据中可能含有一定的噪声。从直观上看, e 和可信度之间的关系能够反映数据中所含噪声的多少。当然, 可信度的大小还和支持向量回归中 ε -不敏感损失函数中的 ε 值有关, ε 值越小, 可信度应该越大。而 e 值越小, 可信度就越小, 由此可见 ε 和 e 对可信度的影响成相反关系。若取 $e = \varepsilon$, 当 ε 值逐渐减小时, 可信度的变化相对比较平稳, 则说明数据中噪声很小或没有噪声, 若可信度变化较快, 则说明数据中含有一定的

噪声。总之，通过对各参变量的调整，能够在一定程度上辨别出数据中的噪声含量^[144]。

§ 4.9 本章小结

本章首先介绍了支持向量回归机的基本原理，然后详细地阐述并分析了各种支持向量回归机模型，包括 ε -支持向量回归机、 ν -支持向量回归机、 LS -支持向量回归机、 W -支持向量回归机、基于线性规划的支持向量回归机，同时，建立了在 ε -不敏感函数下的单参数约束下的支持向量回归模型，并证明了单参数回归模型与标准回归模型的等价性。最后探讨并建立了支持向量回归中的预测信任度。

第五章 支持向量机多值分类方法研究

支持向量机的模式分类方法最初是为两值分类问题而设计的，而在实际问题的研究与应用中，多值分类问题更为普遍^[154, 150, 163, 166]，如何将支持向量机的优良性能推广到多值分类当中去，已成为目前支持向量机研究的另一个热点问题。特别是对于已具有客观评价标准（即已有客观分类标准）的多类别分类方法。

目前，支持向量机多值分类方法可以分为两类：一种是通过构造多个 SVM 二值分类器并将它们组合起来实现多类分类^[145]；另一种途径是直接在一个优化公式中来同时考虑所有类别的参数优化。后者尽管看起来简洁，但是在最优化问题求解过程中的变量远远多于第一类方法，训练速度远不及第一类方法，而且在分类精度上也不占优^[128]。当训练样本数非常大时，这一问题更加突出。

本章在对现有主要的 SVM 多值分类算法进行研究基础上，分析各种方法的特点与缺陷，进而提出新的多值分类方法，并通过数值实验来分析、比较各种算法的特点。

§ 5.1 基于两值分类的 SVM 多值分类方法

设训练样本为 $\{(x_1, y_1), \dots, (x_l, y_l)\} \in R^n \times R$ ，其中， n 为输入向量维数， $y_i \in \{1, \dots, M\}$ ， M 为类别数。将样本分成 M 类，其中 $\{(x_i^{(s)}, y_i^{(s)}), i = 1, \dots, l\}$ 代表第 s 类训练样本， $l_1 + \dots + l_M = l$ 。

目前主要有以下一些基于二值分类的多值分类方法实现支持向量机的多类别分类。

5.1.1 “一对多”方法

“一对多”方法(one-against-rest 方法)是构造 k 个支持向量机子分类器^[141]。在构造第 j 个支持向量机子分类器时, 将属于第 j 类别的样本数据标记为正类, 不属于 j 类别的样本数据标记为负类。测试时, 对测试数据分别计算各个子分类器的决策函数值, 并选取函数值最大所对应的类别为测试数据的类别。

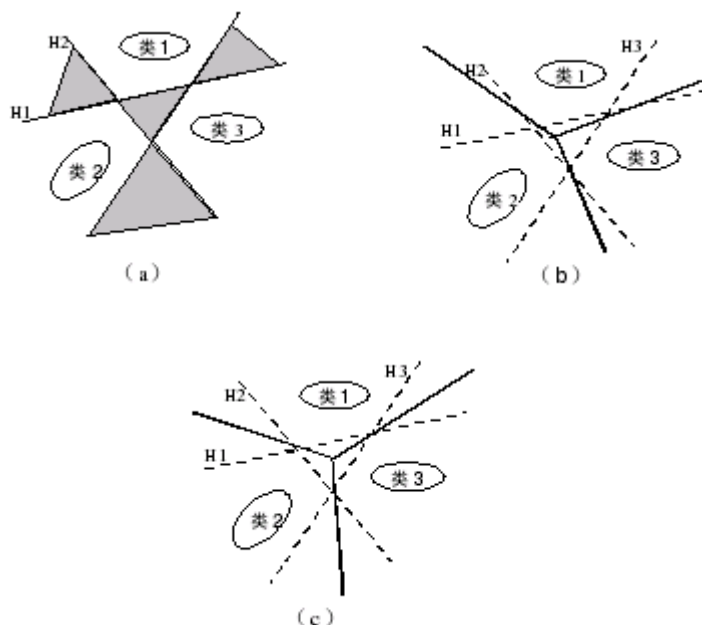


图 5.1 “一对多”方法分类图示

这种方法的优点是: 只需要训练 k 个两类分类支持向量机, 故其所得到的分类函数的个数(k 个)较少, 其分类速度相对较快。

这种方法的第一个缺点为每个分类器的训练都是将全部的样本作为训练样本, 这样需要求解 k 个含 l 个变量的二次规划问题, 因为支持向量机的训练速度随着训练样本的数量的增加急剧减慢, 因此, 这种方法训练时间较长。

第二个缺点是: 如果以两类分类器的输出取符号函数, 如图 5.1(a) 的阴影部分所示。如果最后的输出是两类分类器输出为最大的一类, 则有可能存在测

试样本同时属于多类或不属于任何一类的区域。如果最后的输出是两类分类器输出为最大的那一类，则可能人为地将分类超平面偏转，如图 5.1(b)所示，理想分类超平面如图 5.1(c)所示。之所以发生这种情况的原因是因为分类器的输出是一个相对距离，同一分类器的输出具有可比性，而不同分类器由于其相对的标准不同，其输出不具有可比性。

5.1.2 “一对一”方法

“一对一”方法(one-against-one method)^[137, 144]是在 k 类训练样本中构造所有可能的两类分类器，每个分类器仅仅在 k 类中的 2 类训练样本上训练，分别构造 $\frac{1}{2}k(k-1)$ 个 SVM 子分类器。在构造类别 i 和类别 j 的 SVM 子分类器时，在样本数据集选取属于类别 i 、类别 j 的样本数据作为训练样本数据，并将属于类别 i 的数据标记为正，将属于类别 j 的数据标记为负。测试时，将测试数据对 $\frac{1}{2}k(k-1)$ 个 SVM 子分类器分别进行测试，并累计各类别的得分，选择得分最高者所对应的类别为测试数据的类别。

“一对一”方法的优点是：其训练速度较“一对多”方法快^[134]。“一对一”方法的缺点是：1) 如果单个两类分类器不规范化，则整个 k 类分类器将趋向于“过学习”；2) 泛化误差无界；3) 分类器的数目 $k(k-1)/2$ 会随着类数 k 的增大而急剧增加，导致在决策时速度很慢^[136]。

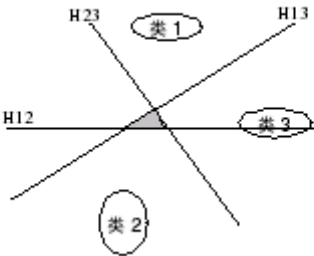


图 5.2 “一对一”方法分类图示

由于在决策时，此方法采用了投票法，有可能存在多个类投票相同的情况，即有可能存在一个样本同时属于多个类的情况，而使得此方法无法进行很好的决策。如图 5.2 所示，阴影部分的样本在 1、2、3 类的投票均为 1，从而无法决策此区域的点属于哪一类。“一对一”方法存在不可区分区域。

5.1.3 DDAG 多值分类方法

DDAG-SVM 方法是 Platt 等提出的决策导向非循环图(Decision Directed Acyclic Graph, DDAG)方法^[104]，将多个两类分类器组合成多类分类器。在训练阶段，其与“一对一”方法相同，对于 k 类问题，DDAG 含有 $\frac{1}{2}k(k-1)$ 个二类分类器。然而在决策阶段，使用从根节点开始的导向非循环图 (DDAG)，具有 $\frac{1}{2}k(k-1)$ 个内部节点以及 k 个叶子节点，每个内部节点都是一个二类分类器，叶子节点为最终的类值。给定一个测试样本，从根点开始根据分类器的输出值决定其走左侧或右侧路径，如此做下去，一直到叶子节点为止得到样本所属的类值。如图 5.3 所示。

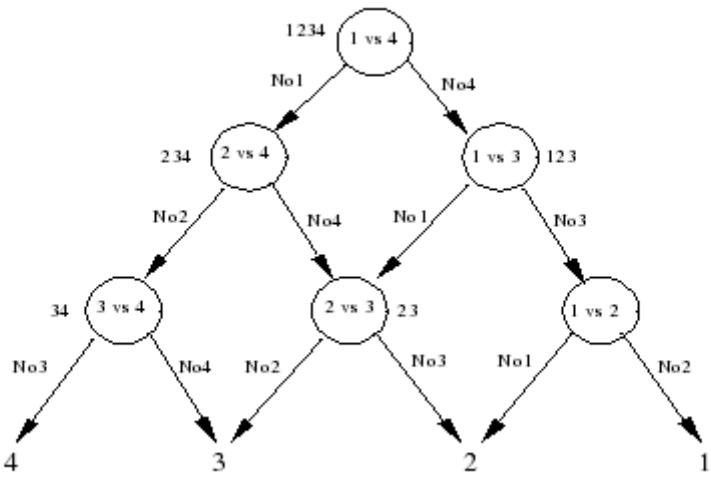


图 5.3 DDAG-SVM 决策图示

其优点是决策速度显著比“一对多”方法或“一对一”的投票方法快。其缺点是根节点的选择直接影响着分类的结果，不同的分类器作为根节点，其分类结果可能会不同，从而产生分类结果的不确定性。

比如对于“一对一”方法中同时属于多类的例子，当选择第 1 类与第 3 类的分类器 SVM13 作为根节点的话，如图 5.4(a)所示，图 5.2 中阴影部分样本的分类结果为类 2。而如果选择第 1 类与第 2 类的分类器 SVM12 作为根节点的话，如图 5.4(b)所示，图 5.2 中阴影部分样本的分类结果为类 3。

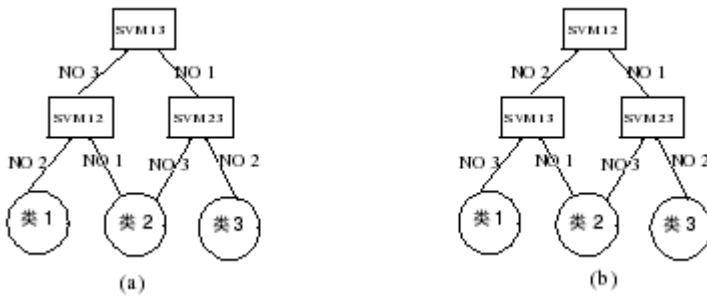


图 5.4 DDAG 法分类

§ 5.2 基于超球面的 SVM 多值分类方法^[144, 145]

5.2.1 分类算法

由于通常构造多值分类的方法具有很高的计算复杂性，因此提出简单实用的多值分类方法显得非常必要。在一类分类思想的启发下，本节提出一种多类分类算法。该方法是在高维特征空间中对每一类样本求出一个超球体中心，然后计算待测试样本到每类中心的距离，最后根据最小距离来判断该点所属的类，具体步骤如下：

设训练样本为 $\{(x_1, y_1), \dots, (x_l, y_l)\} \in R^n \times R$ ，其中， n 为输入向量维数， $y_i \in \{1, \dots, M\}$ ， M 为类别数。将样本分成 M 类，其中， $\{(x_i^{(s)}, y_i^{(s)}), i = 1, \dots, l\}$ 代

表第 s 类训练样本, $l_1 + \dots + l_M = l$ 。首先给出原空间中的优化算法, 为了求包含每类样本的最小超球体, 同时允许一定的误差存在, 构造下面的二次优化:

$$\min \sum_{s=1}^M R_s^2 + C \sum_{s=1}^M \sum_{i=1}^{l_s} \xi_{si} \quad (5.2.1)$$

约束条件

$$(x_i^{(s)} - a_s)^T (x_i^{(s)} - a_s) \leq R_s^2 + \xi_{si}, s = 1, \dots, M, i = 1, \dots, l_s \quad (5.2.2)$$

$$\xi_{si} \geq 0, s = 1, \dots, M, i = 1, \dots, l_s \quad (5.2.3)$$

该优化问题的对偶形式为:

$$\max \sum_{s=1}^M \sum_{i=1}^{l_s} \alpha_i^{(s)} (x_i^{(s)} \cdot x_i^{(s)}) - \sum_{s=1}^M \sum_{i=1}^{l_s} \sum_{j=1}^{l_s} \alpha_i^{(s)} \alpha_j^{(s)} (x_i^{(s)} \cdot x_j^{(s)}) \quad (5.2.4)$$

约束条件

$$0 \leq \alpha_i^{(s)} \leq C, s = 1, \dots, M, i = 1, \dots, l_s \quad (5.2.5)$$

$$\sum_{i=1}^{l_s} \alpha_i^{(s)} = 1, s = 1, \dots, M \quad (5.2.6)$$

借助核映射思想, 首先通过映射 φ 将原空间映射到高维特征空间, 然后在高维特征空间中进行上面的优化, 并通过引入核函数 $K(x, y)$ 代替高维特征空间中的内积运算, 于是可以得到核映射下的优化方程为:

$$\max \sum_{s=1}^M \sum_{i=1}^{l_s} \alpha_i^{(s)} K(x_i^{(s)}, x_i^{(s)}) - \sum_{s=1}^M \sum_{i=1}^{l_s} \sum_{j=1}^{l_s} \alpha_i^{(s)} \alpha_j^{(s)} K(x_i^{(s)}, x_j^{(s)}) \quad (5.2.7)$$

约束条件

$$0 \leq \alpha_i^{(s)} \leq C, s = 1, \dots, M, i = 1, \dots, l_s \quad (5.2.8)$$

$$\sum_{i=1}^{l_s} \alpha_i^{(s)} = 1, s = 1, \dots, M \quad (5.2.9)$$

上面优化式是多值分类问题最终的优化方程，待优化的参数个数是样本总数 l 。因此，该优化方程的计算复杂性只与总的样本数量有关，而与样本的分类数无关。由此可知，该算法在处理多值分类问题时比用 SVM 构造一系列二值优化器要简单得多。

根据 KKT 条件，对应于 $0 \leq \alpha_i^{(s)} \leq C, s=1, \dots, M, i=1, \dots, l_s$ 的样本满足：

$$R_s^2 - (K(x_i^{(s)}, x_i^{(s)}) - 2 \sum_{j=1}^{l_s} \alpha_j^{(s)} K(x_j^{(s)}, x_j^{(s)}) + a_s^2) = 0 \quad (5.2.10)$$

利用上式分别计算出 $R_s, s=1, \dots, M$ 的值。

给定待识别样本 z ，计算它到各个中心点的距离：

$$f_s(z) = K(z, z) - 2 \sum_{i=1}^{l_s} \alpha_i^{(s)} K(z, x_i^{(s)}) + \sum_{i=1}^{l_s} \sum_{j=1}^{l_s} \alpha_i^{(s)} \alpha_j^{(s)} K(x_i^{(s)}, x_j^{(s)}), s=1, \dots, M$$

比较大小，找到最小的 $f_k(z)$ ，则 z 就属于第 k 类。同时可定义该分类结果的信任度如下：

$$B_k = \begin{cases} 1, & \text{当 } R_k \geq f_k(z) \\ \frac{R_k}{f_k(z)}, & \text{否则} \end{cases} \quad (5.2.11)$$

上式表明当所得的值 $f_k(z)$ 位于超球体内部时，此时的信任度为 1，否则，信任度小于 1，并且距离超球体中心越远，信任度越小。

该算法的关键是找到各类的中心点，因此还可以通过适当调整参数 C 的取值来抑制噪声的影响。

另外，考虑到各类别中含有样本数的不同，可能对以上分类原则有一定的影响。比如两类样本数相差悬殊，如图 5.5 所示，设小圆代表样本数少的第一类样本，大圆代表样本数多的第二类样本，那么根据 $f_k(z)$ 的大小可判定新样本（图中由矩形表示）属于第一类，但由于新样本处在第一类样本区域外，而位于

第二类样本区域内，此时将新样本判为第二类更合理。

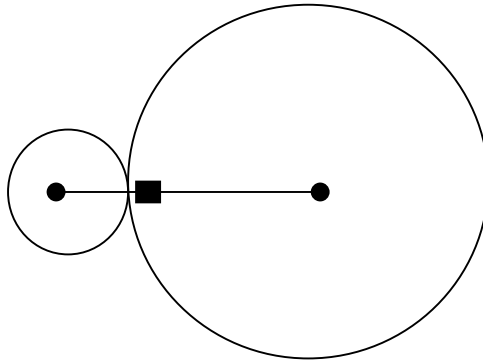


图 5.5 样本数目相差悬殊时的分类

为了在各类样本数不同的情况下仍能保持合理的分类结果，可以将原来分类原则中“找出最小的 $f_k(z)$ ”改为“找出最小的 $\frac{f_z(z)}{R_k}$ ”，这样就可以克服原分类原则中样本数相差悬殊时的不合理分类情况。

文中采用的都是各类样本数比较均衡的分类数据，因此，本文的分类结果都是通过直接比较 $f_k(z)$ 的大小得到的。

5.2.2 分解算法

上节所给算法用一个式子将所有类的中心求出，这在处理较小规模数据时具有一定的简洁性。从算法中可以发现，每类的优化过程实际上是独立的，因此，完全可以将每类的优化过程分开，这在处理大规模数据时是必要的。另外，根据所给算法的特点，当每一类的样本数仍然很大时，还可以进一步将每一类的样本分解为一些较小规模的样本集合，然后分别对小样本集合实行优化操作，进而获得各个小样本集的中心。

设整个训练样本为 M 类，分别将每一类的样本分解为一些较小规模的样本

集合。设各类的分解数量分别为： N_1, N_2, \dots, N_M 。于是整个样本数据可分开写为：

$$\{(x_1^{(st)}, y_1^{(st)}), \dots, (x_{l_{st}}^{(st)}, y_{l_{st}}^{(st)}), s=1, \dots, M, t=1, \dots, N_s\}$$

分别对各个小规模的数据集合实行一类分类运算，从而获得各个小规模数据集合的决策中心 $a_{st}, s=1, \dots, M, t=1, \dots, N_s$ 。

给定待识别样本： z ，计算其到各个小规模数据集合的中心距离：

$$f_{st}(z) = K(z, z) - 2 \sum_{i=1}^{l_{st}} \alpha_i^{(st)} K(z, x_i^{(st)}) + \sum_{i=1}^{l_{st}} \sum_{j=1}^{l_{st}} \alpha_i^{(st)} \alpha_j^{(st)} K(x_i^{(st)}, x_j^{(st)}),$$

$$s=1, \dots, M; t=1, \dots, N_s$$

可以通过两种方法进行类别判定：

①求平均距离：

$$\gamma_i = (f_{i1}(z) + \dots + f_{iN_s}(z)) / N_i, i=1, \dots, M$$

比较大小，找出最小的 γ_k ，则 z 属于第 k 类。

②直接比较所有 $f_{st}(z)$ 的大小，然后根据最小值判断新样本所属类别。

由于所给分解算法的各个优化过程是独立的，因此可以进行并行运算。下面以一些实际的分类数据为例来说明分解算法的有效性。

5.2.3 数值实验

选择数据为：German 和 banana 数据^[144]。German 有 740 个训练样本，300 个测试样本，20 维输入，输出维数是 1；banana 有 400 个训练样本，4900 个测试样本，2 维输入，输出维数是 1。

在分解算法中，将每类样本分成两组小样本集，然后采用平均距离判断新样本所属类别。另外，对同样的数据，分别用未分解算法、分解算法和 SVM 算法进行分类，计算结果参见表 5.1 所示，括号中的数据表示运行时间（单位：

秒 s)，它们的识别精度和运行时间比较结果如表 5.1 所示。

表 5.1 分解算法、未分解算法与 SVM 的结果比较

方法	SVM 算法	未分解算法	分解算法
German	90% (135.4250)	90% (125.5200)	90% (62.8800)
banana	88.82% (2.4737e+003)	87.29% (726.2550)	89.33% (262.8780)

各个算法中取同样的参数，其中， $C = 5, \sigma^2 = 0.3$ ，核函数为：

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right).$$

§ 5.3 基于线性规划的 SVM 多值分类方法

下面将线性规划下的分类方法推广到多值分类情况^[744]。采用的方法是对每一类样本实行一类分类运算，进而对每类样本都获得一个决策函数 $f_i(x)$ ，然后将待测试样本输入每一个决策函数中，根据决策函数的最大值来判断该点所属的类，具体算法如下所述：

设训练样本为 $\{(x_1, y_1), \dots, (x_l, y_l)\} \in R^n \times R$ ，其中， n 为输入向量维数， $y_i \in \{1, \dots, M\}$ ， M 为类别数。将样本分成 M 类，其中， $\{(x_i^{(s)}, y_i^{(s)}), i = 1, \dots, l\}$ 代表第 s 类训练样本， $l_1 + \dots + l_M = l$ 。

构造下面的线性规划：

$$\min \quad -\rho + C \sum_{s=1}^M \sum_{i=1}^{l_s} \xi_{si} \tag{5.3.1}$$

约束条件：

$$\sum_{j=1}^{l_s} \alpha_j^{(s)} K(x_j^{(s)}, x_i^{(s)}) \geq \rho - \xi_{si}, s=1, \dots, M; i=1, \dots, l_s \quad (5.3.2)$$

$$\sum_{j=1}^{l_s} \alpha_j^{(s)} = 1, s=1, \dots, M \quad (5.3.3)$$

$$\alpha_i^{(s)}, \xi_{si} \geq 0, s=1, \dots, M; i=1, \dots, l_s \quad (5.3.4)$$

求解，即可得到 M 个决策函数：

$$f_s(x) = \sum_{j=1}^{l_s} \alpha_j^{(s)} K(x_j^{(s)}, x), s=1, \dots, M \quad (5.3.5)$$

对于给定待识别样本 z ，计算

$$\gamma_s = f_s(z), s=1, \dots, M$$

比较大小，找到最大的 $\gamma_k(z)$ ，则 z 就属于第 k 类。同时可定义该分类结果的信任度如下：

$$B_k = \begin{cases} 1, & \text{当 } \gamma_k \geq \rho \\ \frac{\gamma_k}{\rho}, & \text{否则} \end{cases} \quad (5.3.6)$$

当各类样本数相差悬殊时，可以在优化过程中调整 ρ 的不同取值实现决策分类。

§ 5.4 二叉树的 SVM 多值分类方法

5.4.1 训练样本的选择方案

对 k 类样本根据各类样本的数量进行排序，训练样本最多的类为第 1 类，以此类推，一直到第 k 类^[140]。对于 k 类的训练样本，训练 $k-1$ 个支持向量机，第 1 个支持向量机以第 1 类样本为正的训练样本，将第 2, 3, ..., k 类训练样本作

为负的训练样本训练 SVM1, 第 i 个支持向量机以第 i 类样本为正的训练样本, 将第 $i+1, i+2, \dots, k$ 类训练样本作为负的训练样本训练 SVM i , 直到第 $k-1$ 个支持向量机将以第 $k-1$ 类样本作为正样本, 以第 k 类样本为负样本训练 SVM $k-1$ 。

这样的样本选择方案解决了这样三个问题: (1) 所需要训练的两类支持向量机的数量少, 对于一个 k 类分类问题, 只需要训练 $k-1$ 个两类支持向量机即可。 (2) 消除了决策时存在同时属于多类或不属于任何一类的区域。 (3) 第一个两类支持向量机在全部样本上进行训练, 第二个两类支持向量机在除第一类样本外的其他样本上进行, 直到第 $k-1$ 个两类支持向量机在第 $k-1$ 类样本和第 k 类样本上进行训练。因此随着训练的进行, 总共训练样本和“一对多”方法相比减少了很多。

5.4.2 算法实现

假设有 1 个训练样本 $(x_1, y_1), \dots, (x_l, y_l)$, 其中 $x_i \in R^n, i=1, \dots, l, y_i \in \{1, \dots, m\}$ 是样本 x_i 所属的类别, m 为训练样本的类数目。设第 k 个支持向量机的训练样本集合为: $X = \{(x_i, y_i) | y_i \geq k\}$, 这样的支持向量机, 其数学描述如下:

$$\min_{w^k, b^k, \xi_j^k} \frac{1}{2} \|w\|^2 + C \sum_{j=1}^l \xi_j^k \quad (5.4.1)$$

约束条件

$$w^k \cdot \varphi(x_j) + b^k \geq 1 - \xi_j^k, \quad \text{if } y_j = k \quad (5.4.2)$$

$$w^k \cdot \varphi(x_j) + b^k \leq 1 - \xi_j^k, \quad \text{if } y_j > k \quad (5.4.3)$$

$$\xi_j^k \geq 0; \quad j=1, 2, \dots, l, m \in \{1, \dots, k\} \quad (5.4.4)$$

此问题最终可转化为其对偶问题进行求解:

$$\max W(\alpha^k) = \sum_i \alpha_i^k - \frac{1}{2} \sum_{i,j} \alpha_i^k \alpha_j^k y_i^k y_j^k (x_i^k \cdot x_j^k) \quad (5.4.5)$$

其对应的决策函数为：

$$f(x) = \text{sgn}(w^k \cdot \varphi(x_j) + b^k) \text{。} \quad (5.4.6)$$

5.4.3 决策方案

为了吸取 DDAGSVM 方法中决策时需要的分类器少以及决策速度快的优点，本方案中提出了一种将决策二叉树与支持向量机相结合的决策方法。

首先将 SVM1 作为决策二叉树的根节点，将属于第一类的测试样本决策出来，不属于第一类的样本通过 SVM2 进行分类，如此直到 SVM $k-1$ 将第 k 类样本决策出来。在 DDAGSVM 方法当中，所有的测试样本都必须经过 $k-1$ 次分类才能得到其所属的类。

而在本方案当中，第 1 类的样本只要经过一次分类就可以得到其所属的类，第 2 类的样本只须经过两次分类得到其所属的类，如此直到第 $k-1$ 和 k 类样本须经过 $k-1$ 次分类得到其所属的类，因此从总体效果上看，其决策时所需要经过分类器的数量要大大少于 DDAGSVM 方法，实验结果表明，在决策阶段，其速度要快于 DDAGSVM 方法。

5.4.4 数值试验

本方案在 UCI 数据库的 wine, vehicle, iris, glass, vowel 五个数据上进行了试验。DDAGSVM 方法的训练阶段与“一对一”方法完全相同，因此其训练和决策速度都是最快的，文献[142]的实验结果已经证明了这一点，因此本实验只将本文所提出的基于二叉树的多类分类方法与 DDAGSVM 进行了比较。实验结果如表 5.2。

5.2 各种训练样本的比较

	样本数	样本维数	样本类数	DDAGSVM			二叉树分类方法		
				训练时间(s)	训练精度	测试时间(s)	训练时间(s)	训练精度	测试时间(s)
wine	178	13	3	0.88	100%	0.66	0.66	100%	0.62
vehicle	150	4	3	0.55	100%	0.55	0.55	100%	0.40
iris	846	18	4	69.26	100%	9.17	9.17	100%	6.36
glass	214	10	7	2.5	100%	1.81	1.81	100%	0.99
vowel	990	13	11	46	100%	20.05	20.05	100%	14.90

从以上试验的结果可以看出，基于决策二叉树的多类分类方法在保留了“一对多”方法所需要训练的支持向量机数量少的优点，其训练速度有了明显的提高（“一对多”的训练速度远远慢于“一对一”或 DDAGSVM 方法^[154]），可以达到并超过了“一对多”及 DDAGSVM 方法。

§ 5.5 本章小结

本章分析与总结了目前存在的支持向量机的多值分类算法，包括“一对多”方法、“一对一”方法、超球面分类法、二叉树的多类分类方法以及其它求解多类别的方法，分析与讨论了它们的优缺点及性能。

第六章 水质预测的 SVR 模型研究

基于数据的机器学习是继专家系统之后人工智能应用的又一重要研究领域，也是人工智能核心研究课题之一，研究从观测数据出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测。现有的机器学习方法的共同理论基础之一就是统计学。

传统的统计学是在样本数目趋于充分大的前提条件下进行研究的，所提出的各种方法都只有在样本数目趋于无穷大时才能有理论上的保证。但是在实际问题中，样本数目往往是有限的，这时很多方法都难以取得理想的效果。一方面，经验风险 R_{emp} 只有在大数定理的保证下才能与期望风险 R 在概率上保持一致；另一方面，即使我们得到在样本数目无穷大时趋近 R 的 R_{emp} ，也不能保证在有限样本情况下仍然得到好的效果^[46]。

水质参数的地面监测方法是在研究的水域布置大量的观测点，经过大量的采样，室内理化检测、人工分析得到水质参数浓度的方法。在水质参数浓度数据确定以后，如何将参数数据转化为确定水质状况信息，获得水环境现状及其水质分布状况，即水质评价是水质预测部门主要任务之一。要想得到精度很高的水质的预测值，就要建立精确的数学模型。上世纪 90 年代中期，Vapnik 和他的 At&Bell 实验室小组提出了 SVM 算法，进一步丰富和发展了统计学习理论，使它不仅是一种理论分析工具，还是一种能构造具有多维预测功能的预测学习算法的工具，使抽象的学习理论能够转化为通用的实际预测算法。

本章在第四章综合研究几种 SVR 算法的基础上，针对水环境中某些污染物浓度的预测问题，提出了一种新的支持向量回归机的优化算法。然后，利用改

进的支持向量回归机对实验数据进行回归估计,并与BP网络方法进行比较分析。

§ 6.1 预测学习方法

预测学习方法是机器学习的一个重要研究领域,绝大部分机器学习算法的应用都涉及到对数据的预测性学习,目的是从已知数据中估计相关性以达到准确预测将来数据变化。经典的模式识别方法,各种统计方法以及近些年来出现的各种神经网络算法都被应用于估计这种数据中固有的相关性,试图建立一个能预测未来数据的模型。虽然,这些应用在一些特殊的领域取得了一定的进展^[172],并且提出了一些新的概念和方法^[175],但是,这些方法有其固有的算法缺陷:由于不知道数据的分布密度以及有限的训练数据,常常导致了病态的训练结果,预测效果往往不尽人意。人们迫切需要关于预测学习算法共同的概念和理论框架以及更高效的预测学习方法。

到目前为止,几乎所有的预测学习算法都可以归结为以下三类^[211]:

(1) 传统的统计预测方法

这类方法是在参数结构形式已知的前提下,通过训练数据,预测各参数的值。例如:最小二乘法。应用这些预测方法除了需要很强的先验知识外,还需预先知道模型的结构形式。但是,在处理大量的实际预测问题时,常常不知道背景知识,面对一大堆采样来的数据,也不知道模型的结构形式。由于传统统计学研究的前提是样本数目趋于无穷大时的渐进理论,而参数预测方法几乎都是建立在这一前提基础之上的。因此只有当采样数据趋于无穷时,参数方法的训练结果才趋于真实的模型。由于实际样本数目是有限的,很难满足这一前提。

(2) 经验非线性预测方法

近些年发展起来的人工神经网络和柔性统计方法就属于这一类。虽然,这些新的方法克服了参数预测方法的部分弱点,能够依照需要,假设数据内在相关性而构造非线性模型。然而,这些非线性方法缺乏统一的数学理论基础,通

常是从生物学的理论和一些学术流派中得到灵感，对于诸如神经网络中的结构选择和权重的初值设定，仍需要借助于经验，得到的模型通常是局部最优解，而非全局最优。

(3) 统计学习理论

早期的统计学习理论是从上世纪 60 年代发展起来的^[46]。直到 90 年代，它一直是作为一种针对有限样本的函数预测问题的纯理论分析工具^[40]。虽然，早期的统计学习理论提出了 VC 维(Vapnik-Cervonenkis)理论^[34]，为衡量预测模型的复杂度，提出了有效的理论框架。但是，它仍然是建立在经验风险最小化原则基础之上，即：以训练的平均误差为最小的模型作为期望的最终模型。所以，早期的统计学习理论一直停留在抽象的理论和概念的探索之中，直到 90 年代初期，VC 维理论还没有得到很好的应用^[211]。

近年来，人们发展了支持向量机函数拟合方法^[35]，即支持向量回归机。由于支持向量机是建立在结构风险最小化的原理而非经验风险最小原理基础之上，因此具有很好的泛化能力，同时其最终转化为求解一个凸二次规划问题，它可以按任意精度逼近非线性函数，具有全局极小值点、收敛速度快的优点。至 1995 年之后，受到了广大科技工作者的关注，成为了研究的热点。目前国内外对此方面的研究工作还很少。

§ 6.2 支持向量回归机函数拟合的基本原理

对于支持向量回归机算法(SVR)函数拟合，可以分为线性和非线性两种情况。目前函数拟合多为非线性函数拟合。在非线性情况下，可以把样本 x 通过非线性映射 $\varphi(x)$ 映射到高维特征空间 H ，并在 H 中求解最优回归函数。这样，在高维空间中实际上只需进行内积运算，而这种内积运算是可以用原空间的函数来实现的，我们甚至没有必要知道变换的具体形式。根据泛函的有关理论，只要有一种核函数 $K(x_i, x)$ 满足 Mercer 定理^[79]，它就对应某一空间中的内积。因此，

在最优回归函数中采用适当的核函数 $K(x_i, x)$ 代替高维空间中的向量内积 $\varphi(x_i) \cdot \varphi(x)$ ，就可以实现某一非线性变换后的线性拟合，而计算复杂度却没有增加。

此时式(2.3.28)及式(2.3.29)的最优化问题最终可转化为：

$$\max \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i \cdot x_j) \quad (6.2.1)$$

此时回归函数的法向量和回归函数为：

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i \cdot x) \quad (6.2.2)$$

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i \cdot x) + b \quad (6.2.3)$$

§ 6.3 支持向量回归机的拟合特性分析

一个模型质量的好坏不仅取决于它的对样本的学习能力，而且取决于对未知样本的预测能力，通常称为学习能力和泛化能力。目前用于函数拟合的支持向量机通常采用径向基核函数，因为径向基核函数与多项式核函数相比，具有更好的拟合能力。径向基核函数是局域核函数，同时加上支持向量机具有结构风险最小化以及全局最优的特点，其在训练样本的范围内，采用径向基核函数的支持向量机的函数拟合能够以很高的精度逼近任何非线性函数，具有很强的拟合和预测能力。

SVM 用来估计回归函数时，有三个主要的特色：1)利用在高维空间中定义的线性函数集来回归估计；2)利用线性最小化来实现回归估计，风险是用 Vapnik 的 ε -不敏感损失函数来度量的；3)采用的风险函数是由经验误差和一个由结构风险最小化原则导出的正则化部分组成。

§ 6.4 改进的函数拟合与预测方法的研究思路

在支持向量回归机中, 如果样本来自于同一分布且是独立的, 也就是说随机误差项具有相同方差并且不相关, 此时将具有很好的建模效果。但由于实际问题是错综复杂的, 因而在建立实际问题的回归模型时, 经常会出现某一因素随着自变量观测值的变化而对因变量产生不同的影响, 导致随机误差产生不同方差^[145]。当一个回归问题存在异方差性时, 如果仍然用标准的支持向量回归算法来建立模型, 往往达不到好的效果。为了解决这个问题, 下面提出一种改进的加权支持向量回归算法。

6.4.1 改进的加权支持向量回归机

这里仍然直接给出核函数方法下的回归模型。设给定的训练样本为:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subset R^n \times R$$

同样采用一个非线性映射 φ 把数据映射到一个高维特征空间, 然后在高维特征空间中进行线性回归。设回归函数为:

$$f(x) = w \cdot \varphi(x) + b \quad (6.4.1)$$

设随机误差项的方差具有下面的性质:

$$\begin{cases} E(\varepsilon_i) = 0, i = 1, \dots, l. \\ \text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma_i^2, i = j, \\ 0, i \neq j \end{cases} \end{cases} \quad (6.4.2)$$

在 σ_i^2 不相同的情况下, 标准支持向量回归模型的效果往往不理想, 因为优化过程中各个参数 ξ_i, ξ_i^* 的地位是相同的, 也就是说对误差项的考虑是相同的惩罚尺度, 这就往往导致回归线被拉向方差大的项, 而方差小的项拟合程度就差。为了解决这个问题, 通过引入一个适当的权数 λ_i , 以调整各项在回归中的作用,

也就是在标准支持向量回归模型中加入权系数来控制异方差的影响，于是可得下面的优化形式：

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \lambda_i (\xi_i + \xi_i^*) \quad (6.4.3)$$

约束条件：

$$w \cdot \varphi(x_i) + b - y_i \leq \xi_i^* + \varepsilon, \quad i = 1, \dots, l \quad (6.4.4)$$

$$y_i - w \cdot \varphi(x_i) - b \leq \xi_i + \varepsilon, \quad i = 1, \dots, l \quad (6.4.5)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \quad (6.4.6)$$

在实际应用中，各个权数可根据实际情况按上面式子成比例选取。

优化问题(6.4.3)的对偶形式为：

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ & + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (6.4.7)$$

约束条件：

$$\begin{aligned} \sum_{i=1}^l (\alpha_i - \alpha_i^*) &= 0 \\ \alpha_i, \alpha_i^* &\leq \lambda_i C, i = 1, \dots, l \\ \alpha_i, \alpha_i^* &\geq 0, i = 1, \dots, l \end{aligned} \quad (6.4.8)$$

解出 α 和 α^* 的值，可得 $f(x)$ 的表达式如下：

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (6.4.9)$$

按照 KKT 条件，故有

$$\varepsilon - y_i + f(x_i) = 0, \text{ 对于 } 0 < \alpha_i < \lambda_i C \quad (6.4.10)$$

$$\varepsilon + y_i - f(x_i) = 0, \text{ 对于 } 0 < \alpha_i^* < \lambda_i C \quad (6.4.11)$$

由以上两式可以求出 b 。

6.4.2 算法实现

将上面的优化算法化为下面的标准二次规划形式：

$$\min \quad \frac{1}{2} x^T H x + c^T x \quad (6.4.12)$$

约束条件：

$$a^T x = 0 \quad (6.4.13)$$

$$0 \leq x \leq C\lambda \quad (6.4.14)$$

其中的参数取下面的值：

$$x = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}, \quad c = \begin{bmatrix} \varepsilon - y \\ \varepsilon + y \end{bmatrix}, \quad H = \begin{bmatrix} D & -D \\ -D & D \end{bmatrix}, \quad (6.4.15)$$

而 $D = K(x_i, x_j), i, j = 1, \dots, l;$

$$a = [1 \quad 1 \quad \dots \quad 1 \quad -1 \quad -1 \quad \dots \quad -1]^T,$$

$$\lambda = [\lambda_1 \quad \dots \quad \lambda_l \quad \lambda_1 \quad \dots \quad \lambda_l]^T,$$

在实际问题的研究中，误差项的方差 σ_i^2 通常是未知的，这就要根据实际情况来确定。关于异方差性的检验，统计学家们进行了大量的研究，提出的诊断方法已有多种，常用的有残差图分析法与等级相关系数检验法等，但目前还没有一个公认的最优方法。

6.4.3 数值实验

以一个经常被采用的例子(toy)来说明所给方法的合理性。取自变量 $x \in [-1, +1]$ ，其中间隔为 0.1，因变量 $y = \sin c(x)$ ，然后在因变量 y 的前 10 个样

本中加入噪声 $N(0,0.8)$ ，其余样本中加入噪声 $N(0,0.1)$ 。分别用标准的支持向量回归和改进的加权支持向量回归建立预测模型，两类支持向量回归机中的参数都取为 $\varepsilon = 0.0001, C = 5$ ，核函数仍取高斯函数：

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

其中 $\sigma^2 = 1.5$ 。改进的加权支持向量回归中的加权值，前 10 个取为 1/8，其余为 1，然后在 -1 到 +1 之间选 10 个测试样本(不同于训练样本)，测试指标采用均方误差：

$$MSE = \sqrt{\frac{1}{K} \sum_{i=1}^K (y_i - \hat{y}_i)^2} \quad (6.4.16)$$

其中， y_i 为实际值， \hat{y}_i 为预测值， K 为测试样本的数量。这里 y_i 取未加入噪声的真实数据。用 Matlab 编程实现，标准支持向量回归算法的预测结果为 $MSE = 0.1319$ ，而改进的加权回归算法的预测结果为 $MSE = 0.0303$ ，这就说明了当数据中存在异方差性时，采用改进的加权支持向量回归算法所获得预测值更接近真实值。

§ 6.5 水质预测的 SVR 模型研究

随着经济活动的发展，环境污染问题日益突出。水环境和水资源与人们的生产和生活息息相关，如何对水域水质进行有效的监测，以便及时发现问题、解决问题，是环保部门的重要任务。水质参数的确定是水质监测的重要目的和任务，也是水质评价与水污染防治的主要依据^[213]。为加强水资源流域统一管理和保护，根据《中华人民共和国水法》第三十二条和《中华人民共和国水污染防治法》第十八条要求，各流域水资源保护局每月都要对流域水体组织水资源进行预测，并发布通报。

确定水质参数的常规方法是在研究水域布置大量观测点^[9-11]，经过实地采样，人工进行物理、化学等分析方法得到，这个过程受人力、物力和气候、地形、水文等条件的限制，难以实现连续、快速地跟踪调查与分析^[6, 7, 192-208]。为缩短预测时间，能够提前预知水质状况从而能够更好的做出水质预测，采用新的间接监测手段越来越受到环保部门的重视。

间接的水质监测方法也有很多，如通过建立数学预测模型进行预测。本节将每月的水质参数平均浓度值的预测看成一个时间序列预测问题，通过建立适当的预测模型对每月的水质参数浓度值进行预测。但是由于水质参数的复杂性和多变性，难以用简单的线性的数学预测模型来描述，所以寻求一种较好的非线性数学预测模型是进行水质参数预测首先要解决的问题。常见的预测方法有单耗法、趋势外推法、弹性系数法、回归分析法、时间序列法、灰色模型法、专家系统法以及近年来应用较多的神经网络预测法、模糊逻辑方法等^[4, 219, 210]。

这里利用支持向量机非线性回归估计的优势，利用改进的加权支持向量回归模型，建立基于支持向量回归机的水质预测模型。

设给定的训练样本为：

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subset R^n \times R \quad (6.4.17)$$

首先采用一个非线性映射 φ 把数据映射到一个高维特征空间，然后在高维特征空间中进行线性回归，同时引入松弛变量 ξ_i ， ξ_i^* ，构造容错惩罚系数 C 和常数 ε 。设回归函数为：

$$f(x) = w \cdot \varphi(x) + b \quad (6.4.18)$$

引入一个适当的权数 λ_i ，以调整各项在回归中的作用，也就是在标准 SVR 模型中加入权系数来控制异方差的影响。本文选取观测误差项方差的倒数，即

$$\lambda_i = \frac{1}{\sigma_i^2} \quad (6.4.19)$$

于是可得下面的优化形式:

$$\min \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \lambda_i (\xi_i + \xi_i^*) \quad (6.4.20)$$

约束条件:

$$(w \cdot \varphi(x_i)) + b - y_i \leq \xi_i^* + \varepsilon, i = 1, \dots, l \quad (6.4.21)$$

$$y_i - (w \cdot \varphi(x_i)) - b \leq \xi_i + \varepsilon, i = 1, \dots, l \quad (6.4.22)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \quad (6.4.23)$$

优化问题(6.4.20)的对偶形式为:

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ & + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (6.4.24)$$

约束条件:

$$\begin{aligned} \sum_{i=1}^l (\alpha_i - \alpha_i^*) &= 0 \\ \alpha_i, \alpha_i^* &\leq \lambda_i C, i = 1, \dots, l \\ \alpha_i, \alpha_i^* &\geq 0, i = 1, \dots, l \end{aligned} \quad (6.4.25)$$

解出 α 和 α^* 的值, 可得 $f(x)$ 的表达式如下:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (6.4.26)$$

按照 KKT 条件, 故有

$$\varepsilon - y_i + f(x_i) = 0, \text{ 对于 } 0 < \alpha_i < \lambda_i C$$

$$\varepsilon + y_i - f(x_i) = 0, \text{ 对于 } 0 < \alpha_i^* < \lambda_i C$$

由以上两式可以求出 b 。

利用支持向量机进行水环境的水质预测。首先要确定影响水质的主要影响因素；其次选择具有实测数据的主要因素构成样本数据集；然后利用 SVM 进行学习训练；最后根据训练后获得的参数进行预测。

§ 6.6 应用实例

表 6.1 太湖流域 2004 年全年及 2005 年 1-8 月份的高锰酸盐指数平均浓度值

高锰酸盐指数平均		高锰酸盐指数平均	
时间（月）	浓度(mg/L)	时间（月）	浓度(mg/L)
2004 年 1 月	4.75	2004 年 11 月	5.00
2004 年 2 月	4.49	2004 年 12 月	4.78
2004 年 3 月	5.36	2005 年 1 月	5.67
2004 年 4 月	5.80	2005 年 2 月	4.99
2004 年 5 月	5.03	2005 年 3 月	4.98
2004 年 6 月	4.11	2005 年 4 月	5.10
2004 年 7 月	5.19	2005 年 5 月	5.25
2004 年 8 月	5.43	2005 年 6 月	4.71
2004 年 9 月	5.12	2005 年 7 月	4.49
2004 年 10 月	5.47	2005 年 8 月	4.91

本文以太湖流域水资源保护局发布的太湖流域水环境监测数据为研究对象，选用太湖 2004 年全年及 2005 年部分月份的高锰酸盐指数月平均浓度作为网络的回归训练数据^[227]，通过建立支持向量机的水质参数评价模型，来预测评价下个月份的高锰酸盐指数月平均浓度值。太湖流域 2004 年全年及 2005 年 1-8 月份的高锰酸盐指数平均浓度值如表 6.1 所示。将太湖流域每月的高锰酸盐指数

平均浓度值的监测过程看成一个时间序列。将前 20 组作为训练样本，其余 2 组作为预测样本，利用新构造的支持向量回归机进行训练与预测。

6.6.1 预测模型的参数选择与训练

本文采用 LIBSVM^[233] 库文件构建 SVM 多值分类模型。LIBSVM 是使用 MATLAB 和 C 语言作为平台的一个专门针对 SVM 算法的软件包，可以通过改变其中的各个参数来实现核函数及其相关系数的选择优化，并进行回归预测。核函数及训练参数的选择对支持向量机的性能有至关重要的影响。因此，选择合适的核函数类型和参数对提高预测精度有重要的作用。由于多项式核函数运算速度较慢，而 sigmoid 核函数在有些情况下会导致结果发散，若采用径向基核函数就可以在运算时间和预测效果之间获得较好的平衡，因此选用径向基核函数进行所有的预测工作。

为获得满意的预测效果，消除实测数据之间大小的差异对计算结果的影响，首先对实测数据进行归一化处理，本文采用极差化处理方法，即

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

其中 x_{\max}, x_{\min} 分别为实测数据中的最大与最小值。

选择 20 个实测点数据组成训练样本，2 个实测点数据作为验证样本，选择不同惩罚系数 C 和参数 σ 、 ε ，用 5 次交叉验证方法进行回归预测，以确定最优参数组合。

根据实验结果，当 $C=500$ ， $\sigma^2 = 128$ ， $\varepsilon = 0.25$ 时预测精度较高。

6.6.2 预测结果分析

用本节构建的 SVR 预测模型对 2005 年 9 月、10 月两个月份的水质参数高锰酸盐指数平均浓度值进行预测，结果如表 6.2 所示：

表 6.2 SVR 模型与 BP 网络模型预测值

时间（月）	实测值 (mg/L)	SVR 预测值(mg/L)	RBFN 预测值(mg/L)	平均相对误差	
				SVR	RBFN
2005.9	5.08	5.01	4.97	-0.07	-0.11
2005.10	5.17	5.09	5.01	-0.08	-0.16

表 6.3 SVR 模型与 BP 网络模型的预测结果与误差比较

性能比较	预测最大误差（%）	平均相对误差（%）
支持向量回归机	1.76	0.76
神经网络	3.24	1.46

由表 6.3 和图 6.1 可以看出，在样本很少的情况下，SVR 的预测能力比 BP 神经网络要强。这是因为神经网络在训练过程中出现“过学习”现象，即把各个训练样本的细节都保存下来，难以区分噪声和真实信号，而 SVR 模型则通过常数的调整，一方面使误差尽可能小，另一方面使回归函数尽可能平滑，从而提高了泛化能力。

经过与人工神经网络方法比较分析可知，支持向量机方法能充分利用训练样本的分布特性。根据部分训练样本即可构建判别函数，不需要过多的先验信息和使用技巧，而 BP 神经网络方法受到可利用样本数较少情况的影响，其泛化预测能力受到较大的削弱，因此预测效果不佳。SVM 算法最终转化为二次寻优问题。从理论上说，得到的将是全局最优解，有效地避免了神经网络易陷入的局部极值问题。同时通过非线性变换和核函数巧妙地解决了高维数问题，使得其算法复杂度与样本维数无关，加速了训练学习速度。此外。它能根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中。保证其有较好的泛化性能。

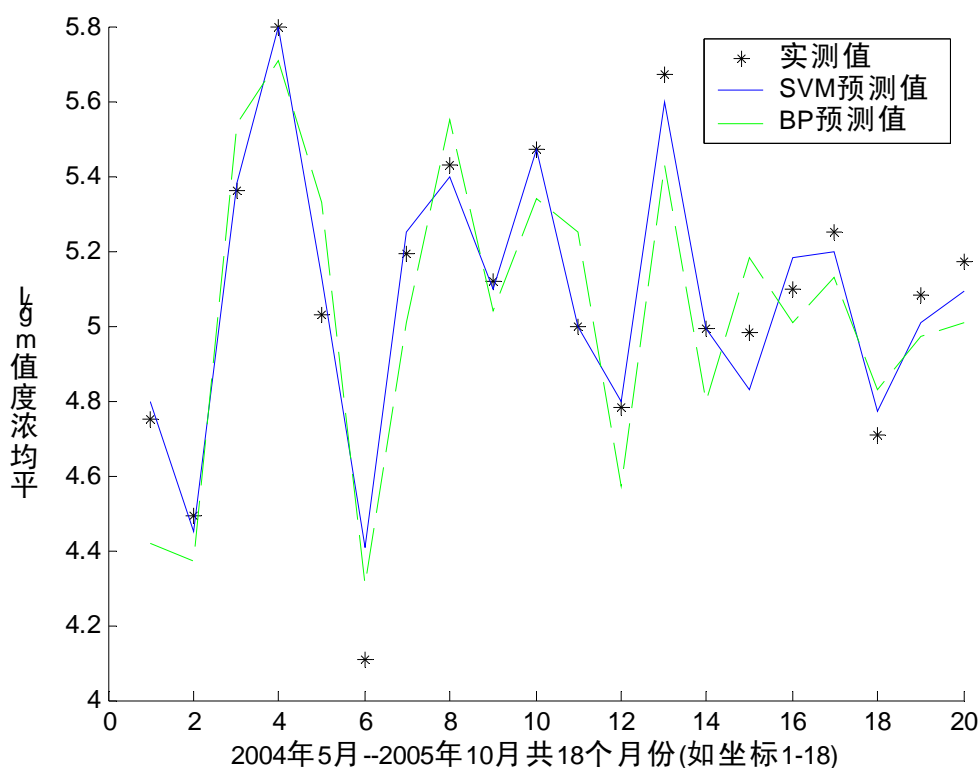


图 6.1 SVR 模型及 BP 网络模型的预测结果曲线

§ 6.7 本章小结

SVM 的方法最早是针对模式识别问题提出的,随着 Vapnik 对 ε -不敏感损失函数的引入,已将其推广应用到非线性回归估计和曲线拟合中,得到了用于曲线拟合的支持向量回归机方法,并且表现出很好的学习效果。由于支持向量回归机同样是建立在结构风险最小化的原理而非经验风险最小原理基础之上,因此具有很好的泛化能力,同时其最终转化为求解一个凸二次规划问题,它可以按任意精度逼近非线性函数,具有全局极小值点、收敛速度快的优点。它不仅是一种理论分析工具,还是一种能构造具有多维预测功能的预测学习算法的工具,使抽象的学习理论能够转化为通用的实际预测算法。至 1995 年之后,受到

了广大科技工作者的关注，成为了研究的热点。目前国内外对此方面的研究工作还很少。

本章在对支持向量回归机的拟合特性分析的基础上，进行改进的函数拟合与预测方法的研究。针对水环境的特点，首先构造改进的加权支持向量回归模型，然后将改进的 **SVR** 模型应用于水质浓度预测，并将预测结果与 **BP** 神经网络方法进行了比较。结果表明 **SVR** 在整体性能上优于 **BP** 神经网络方法，有很强的学习能力、泛化能力，将其用于水质参数预测和水质评价具有很大的应用价值。

第七章 基于 SVM 的水环境质量综合评价方法研究

针对水环境系统水质综合评价的特点和要求，以及常用的 SVM 多值分类方法存在的缺陷，提出改进的支持向量机多类分类算法——基于循环决策二叉树的 SVM 多值分类算法，构建水环境质量综合评价模型，并进行应用性的研究与分析。

§ 7.1 改进 SVM 多值分类方法的提出

由于 one-against-rest 和 one-against-one 的分类方法会存在大量的不可分区域，如图 7.1 中的阴影部分。虽然 one-against-rest 方法，可取最大值来得到类别，但其推广性却受到很大影响。对于图 7.1(b)的阴影，由于总为负值，所以 one-against-one 方法无法识别位于此区域的样本。

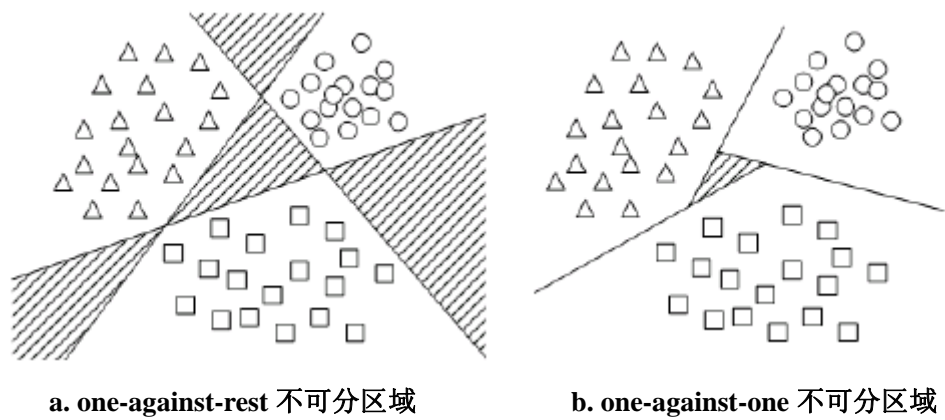


图 7.1 线性情况下的不可分区域示意图

本文在这里建立改进的 SVM 多值分类算法，其基本思想是：结合预抽取支持向量以及循环迭代算法，提出了一种基于循环决策二叉树的 SVM 多值分类算

法，力求克服其他方法存在的问题及缺点，提高支持向量机的训练和决策速度，解决了目前支持向量机多值分类方法的不可分区域的分类问题。

具体做法是：首先将所有类别分成两个子类，再将子类进一步划分成两个次级子类，如此循环下去，直到所有的节点都只包含一个单独的类别为止，此节点也是决策二叉树中的叶子，这样就得到一个倒立的二叉树分类树。该方法将原有的多类问题同样分解成了一系列的两类分类问题，其中的两个子类间的分类函数采用 SVM。决策二叉树方法可以避免传统方法的不可分情况，并且只需构造 $k-1$ 个 SVM 分类器，测试时并不一定需要计算所有的分类器判别函数，从而可节省测试时间。

7.1.1 训练样本的选择方案

首先对 k 类样本根据各类样本的数量进行排序，训练样本最多的类为第 1 类，以此类推，一直到第 k 类^[145]。

对于 k 类的训练样本，训练 $k-1$ 个支持向量机，第 1 个支持向量机以第 1 类样本为正的训练样本，将第 2, 3, ..., k 类训练样本作为负的训练样本训练 SVM1, 第 i 个支持向量机以第 i 类样本为正的训练样本，将第 $i+1, i+2, \dots, k$ 类训练样本作为负的训练样本训练 SVM $_i$, 直到第 $k-1$ 个支持向量机将以第 $k-1$ 类样本作为正样本，以第 k 类样本为负样本训练 SVM $_{k-1}$ 。

这样的样本选择方案解决了三个问题：1) 所需要训练的两类支持向量机的数量少，对于一个 k 类分类问题，只要求训练 $k-1$ 个两类支持向量机即可；2) 消除了决策时存在同时属于多类或不属于任何一类的区域；3) 第一个两类支持向量机在全部样本上进行训练，第二个两类支持向量机在除第一类的样本的其它样本上进行，自到第 $k-1$ 个两类支持向量机在第 $k-1$ 类样本和第 k 类样本上进行训练。因此随着训练的进行，总共训练样本和“一对多”方法相比减少了许多。假如总共为 k 类样本，每类样本数为 n ，“一对多”方法的总的训练样本

数为 $k^2 n$ ，“一对一”方法的总的训练样本数为 $k(k-1) \times n$ ，而本训练样本方案总的训练样本数为 $[k(k+1)/2-1] \times n$ ，因此如果类数较多时，本方案总共训练样本数近似为“一对多”方法及“一对一”方法的一半。

7.1.2 算法实现

假设有 l 个训练样本 $(x_1, y_1), \dots, (x_l, y_l)$ ，其中 $x_i \in R^n, i=1, \dots, l, y_i \in \{1, \dots, m\}$ 是样本 x_i 所属的类别， m 为训练样本的类数目。设第 k 个支持向量机的训练样本集合为： $X = \{(x_i, y_i) | y_i \geq k\}$ 。这样的支持向量机，其数学描述如下：

$$\min_{w^k, b^k, \xi_j^k} \frac{1}{2} \|w\|^2 + C \sum_{j=1}^l \xi_j^k \quad (7.1.1)$$

约束条件：

$$w^k \cdot \varphi(x_j) + b^k \geq 1 - \xi_j^k, \quad \text{if} \quad y_j = k \quad (7.1.2)$$

$$w^k \cdot \varphi(x_j) + b^k \leq 1 - \xi_j^k, \quad \text{if} \quad y_j > k \quad (7.1.3)$$

$$\xi_j^k \geq 0; \quad j=1, 2, \dots, l, m \in \{1, \dots, k\} \quad (7.1.4)$$

此问题最终可转化为其对偶问题进行求解：

$$\max W(\alpha^k) = \sum_i \alpha_i^k - \frac{1}{2} \sum_{i,j} \alpha_i^k \alpha_j^k y_i^k y_j^k (x_i^k \cdot x_j^k) \quad (7.1.5)$$

其对应的决策函数为：

$$f(x) = \text{sgn}(w^k \cdot \varphi(x_j) + b^k)。 \quad (7.1.6)$$

7.1.3 支持向量机的训练方案

为了吸取“一对一”方法的每个两类支持向量机的训练样本少、训练速度快的优点，提高本方案中每一个二类支持向量机的训练速度，本文采取预抽取

支持向量与循环迭代方法训练支持向量机。

在 SVM 训练当中，并不是所有的样本在训练中都起同样重要的作用，而只是少数几个被称作为支持向量的样本对支持向量机的分类起决定的作用。由于支持向量机最终可转化为一个求解非线性二次规划问题，因此，在普通的 SVM 学习当中，花费了大量时间去优化非支持向量，而支持向量在整个样本中点的比例相当小，这样影响了 SVM 方法进行机器学习的效率。如果我们能够在 SVM 学习之前预抽取支持向量集合以减小训练样本的数量，这样就会大大提高 SVM 的训练速度。在有些情况下，由于样本的分布十分复杂，有时一次无法将全部的支持向量预抽取出来，因此为了保证支持向量机的训练精度，需要将预抽取支持向量与循环迭代方法结合起来，从而在保证分类精度的同时，提高支持向量机的训练速度。

7.1.4 预抽取两类样本的相对边界向量

假设训练样本集合为 A，工作集合为 B。

定义 1 样本间距离

给定两个样本 $x_1, x_2 \in R^N$ ，则两样本之间的距离可以表示为 $d(x_1, x_2)$ ，（其中 x_i^j 表示样本 x_i 的第 i 个分量）。

在线性的情况下，两样本 $x_1, x_2 \in R^N$ 之间的距离定义如下：

$$d(x_1, x_2) = \|x_1 - x_2\|_2 = \sqrt{\sum (x_1^i - x_2^i)^2} \quad (7.1.7)$$

在非线性情况下，两样本 $x_1, x_2 \in R^N$ 之间的距离定义如下：

$$d(x_1, x_2) = \|\varphi(x_1) - \varphi(x_2)\| = \sqrt{K(x_1, x_1) - 2K(x_1, x_2) + K(x_2, x_2)} \quad (7.1.8)$$

其中 $\varphi(x)$ 为将原空间中向量 x 映射到高维向量空间中所对应的向量，

$K(x_1, x_2) = \varphi(x_1) \cdot \varphi(x_2)$ 为核函数。

假设一类样本为 $x_{1i}, i=1, \dots, l$, 另一类样本为 $x_{2j}, j=1, \dots, m$, $d(x_{1i}, x_{2j})$ 表示第一类中的第 i 个样本到第二类中的第 j 个样本的距离, 则对每一个 i 值, $\min d(x_{1i}, x_{2j}) (j=1, 2, \dots, m)$ 所对应的向量 x_{2j} 就是第二类样本的一个边界向量。这样 i 取遍所有值后, 就可得到第二类样本的所有和第一类样本相对边界的边界向量集合。同理, 可得到第一类样本的相对边界向量。这样, 我们就得到了以相当大的概率包括支持向量的相对边界向量, 而相对边界向量的数量远远小于样本的数量。以此集合作为初始的工作集合 B 。

7.1.5 训练支持向量机的循环迭代方案

定义 2 样本与最优超平面间的距离 $d = d(x_i) \times y_i$ 。其中 $d(x_i) = 2/\|w\|$, y_i 为样本 x_i 的类标, w 是最有超平面的法向量。

定义 3 分类正确率 $P = n_1/n$ 。其中 n_1 为距离 $d \geq 1$ 的样本数量, n 为训练样本集合 A 中样本的数量。

循环迭代算法的步骤描述如下:

①以工作集 B 为训练样本, 获得分类超平面。

②利用所得到的分类超平面对训练样本集合 A 进行测试, 然后根据定义 2 和定义 3 计算 A 中样本与分类超平面间距离 d 以及分类正确率 P 。按照下面方框图 7.1 的方案将违反 KKT 条件最严重的数量为 β 一部分训练样本以及上一次循环的支持向量样本放入工作集合 B 当中, 取代原先的工作集 B , 作为下一次循环迭代的训练样本。图中 L 为上一次所获得的分类器的支持向量数, M 为样本规模控制的上界。

③当分类正确率 P 小于 1 时, 重复步骤①, 当分类正确率 P 等于 1 时跳出循环, 循环迭代结束。

这种循环迭代算法，即控制了每次迭代所要求解的二次规划的规模，同时兼顾了循环迭代次数，从而达到提高训练速度的目的。

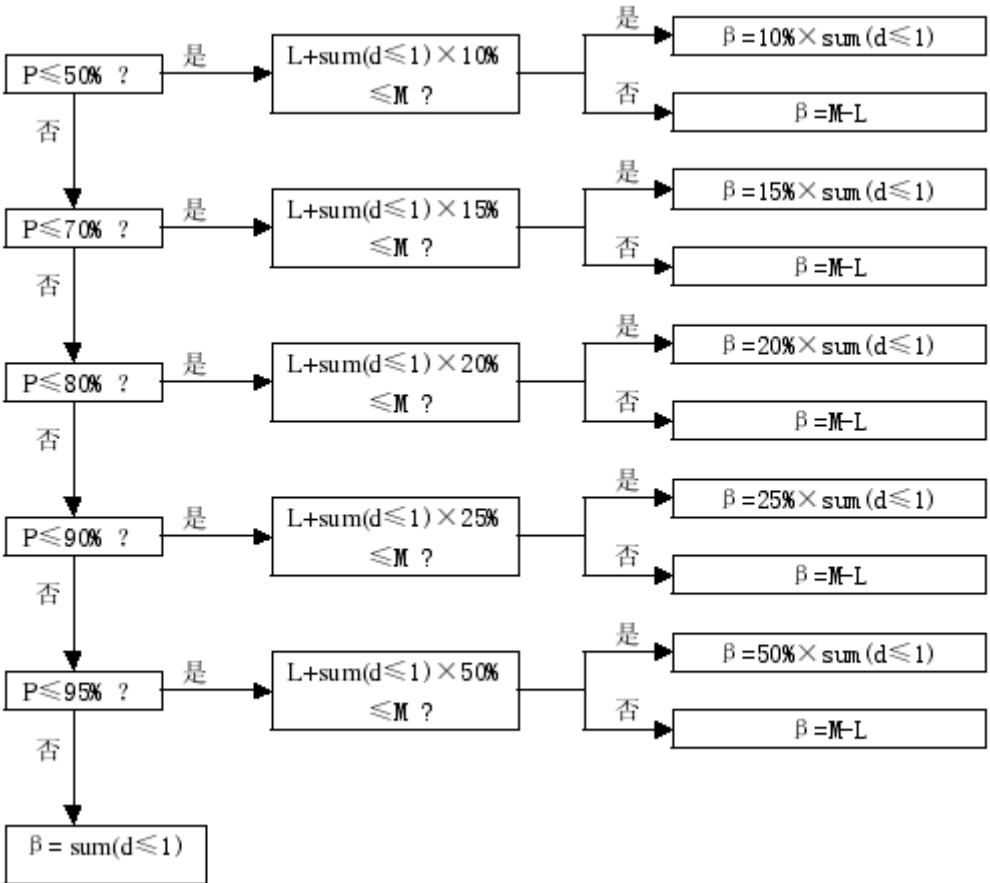


图 7.2 循环迭代算法程序产生框图

7.1.6 支持向量机多值分类的决策方案

为了吸取 DDAG-SVM 方法中决策时需要的分类器少以及决策速度快的优点，本方案中提出了一种将决策二叉树与支持向量机相结合的决策方法。

首先将 SVM1 作为决策二叉树的根节点，将属于第一类的测试样本决策出来，不属于第一类的样本通过 SVM2 进行分类，如此自到 SVM_{k-1} 将第 k 类样本决策出来。决策树如图 7.2 所示。在 DDAGSVM 方法当中，所有的测试样本

都必须经过 $k-1$ 次分类才能得到其所属的类。

而在本方案当中，第 1 类的样本只要经过一次分类就可以得到其所属的类，第 2 类的样本只须经过两次分类得到其所属的类，如此直到第 $k-1$ 和 k 类样本须经过 $k-1$ 次分类得到其所属的类，因此从总体效果上看，其决策时所需要经过分类器的数量要大大少于 DDAGSVM 方法，实验结果表明，在决策阶段，其速度要快于 DDAGSVM 方法。

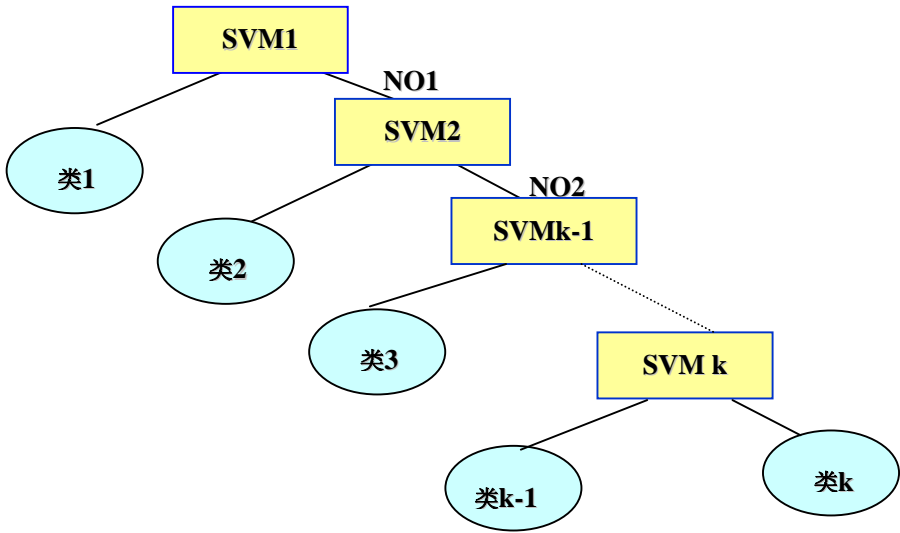


图 7.3 决策二叉树的分类

7.1.7 实施步骤

第一步：对训练样本集合按每类中样本的数量进行排序，样本数量最多的为第一类，以此类推；

第二步：构造支持向量机的训练样本集合，如果训练的是第 k 个支持向量机，则其训练样本集合为 $X_k = \{(x_i, y_i) | y_i \geq k, y_i \in (1, m), k \in (1, m)\}$ ；

第三步：预抽取支持向量集合作为工作集合；

- 第四步：以循环迭代算法训练第 k 个支持向量机；
- 第五步：返回到第一步，直到训练完所有 m-1 个支持向量机为止；
- 第六步：利用本文所提出的决策方法对未知样本进行预测。

§ 7.2 北京市湖泊水库的 SVM 综合评价

7.2.1 北京市水资源概况

北京位于华北平原北端，面积 1.68 万 km²，北部、西部分别为燕山山脉和太行山脉。北京境内有 5 条河流，从东到西分布有洵河、潮白河、北运河、永定河、拒马河，分属于海河流域的蓟运河、潮白河、北运河、永定河和大清河五大水系，共有大小河流 100 余条，全长 2700km。它们均由北山、西山流入东南平原区，是北京平原区地下水重要补给来源。北京有大小湖泊、水库 120 余座，其中大中型以上的湖泊、水库 24 座，蓄水量 45 亿立方米，密云水库、官厅水库是两座最大的水库，蓄水量 37.6 亿立方米，承担着供北京市工业及城市生活用水的主要任务^[208-215]。

表 7.1 1996-2001 年北京地区水资源量统计表

年份	1996	1997	1998	1999	2000	2001	多年平均
地表水资源量 (亿 m ³)	25.95 偏丰水年	10.61 偏枯水年	17.83 平水年	5.16 特枯水年	6.80 枯水年	6.90 枯水年	21.78 ---
入境水量 (亿 m ³)	25.12	8.50	14.45	5.79	6.80	5.34	17.72
出境水量 (亿 m ³)	39.41	19.44	21.43	10.73	9.83	7.23	17.35
地下水资源量 (亿 m ³)	30.26	16.40	29.21	12.81	16.13	15.75	25.21

资料来源：北京市环境保护局，北京市环境质量报告书

北京属温带大陆性季风气候，四季分明，年平均气温 12℃左右。多年平均降水量 626mm，年均降水总量 99.96 亿立方米。受季风气候及地形影响，降雨时空分布极不均匀，年际间变化悬殊^[235]。丰枯水年降水量相差 3 倍^[236]。年降水量集中于 6-9 月份，占全年降水量的 80%以上。地表水水量受降雨的影响，遇到干旱天气，水库蓄水量会大量减少。

在过去的 50 年中北京经历过四次水危机^[238]。上世纪 60 年代第一次水危机通过引密云水库的水而解决；70 年代第二次危机通过过量开采地下水度过；80 年代初第三次水危机，密云水库停止向天津、河北供水，只保北京供水，同时在全市开展计划用水、节约用水，才度过了水危机；从 1999 年开始北京已连续六年干旱，地表资源量开始急剧减少(表 7.1)，北京又面临着第四次供水危机。由于降雨量减少和上游工农业的发展，水库年均入库水量锐减，需求水量的较大差距，使得北京不得不连年超采地下水来保障供水。与 1960 年相比，北京市平原地区地下水量超采了 45 亿 m³，与 1980 年相比超采了 25 亿 m³，形成以朝阳区红庙一带为中心的 2000km²的地下水超采区，其中严重超采区约 1250km²，部分地区地下水含水层已近疏干。地下水严重超采带来了生态环境的恶化、地面沉降、水质超标等一系列严重问题。

7.2.2 水质评价指标

在进行水质评价之前，首先需要进行两方面的工作：一是对自然环境特征、水文特征、经济结构特征进行调查；二是必须掌握污染源和受评水体的污染特征，用以描述污染特征参数即为评价指标。评价指标的选择随评价水体的性质和所采用方法的不同而不同，选择的指标要具有代表性，一般应根据污染源调查分析结果、水质检测情况、以及水体水文特征和所采用的评价方法等确定。对水质进行检测，应根据监测参数的要求，进行采样和采样点的合理优化布置，以便监测数据具有代表性，并实施多次监测，以获得较全面、准确的资料，这是水质评价的前提。对于污染源复杂、超标污染物的种类多、水质监测资料多

的大河可多选一些评价指标；对于污染源少、超标污染物种类相对较少的地区可少选用一些评价指标。

7.2.3 水质评价标准

水体有许多功能，其中最重要的是作为水资源的载体、饮用水水源和野生动物的生境。在国家地表水标准中，水体按照其功能划分为自然保护区和源头(I类)、生活饮用水水源区(II类及生活饮用水卫生标准)、水产养殖区(II类及渔业水质标准)、旅游区(III类及海水质量标准)、工业用水(IV类)、农业灌溉用水(V类及农灌用水标准)。上述标准是衡量具有特定用途或目的水体水质是否污染的人为设定指标。随着人类对水体污染认识的不断深入和对水质要求的不断提高，不同社会的文化和经济条件和不同的生产工艺，决定了不同国家或社区在不同社会发展阶段对水质标准要求有所不同，进而在衡量水体是否污染时所采用的尺度并不相同。在长远的观点上，具有合理的生物多样性和种群结构、健康的营养结构和有着正常能量代谢的水生生态系统，是衡量水体水质良好的唯一标准。

水质的优劣需要进行评价，水质评价是根据水体的用途，按照一定的评价参数、要求水体质量的标准和所选的评价方法，对水域的水质或水域的综合体的质量进行定性或定量的评定。为了防治水体污染，并合理的应用水资源，人们从19世纪开始了对水体质量的评价。1902-1909年，德国学者曾采用生物学方法综合评价水质；1909-1911年，英国学者根据某些化学指标对河流水质评价分类。在水质评价过程中，使用国家法定的标准方法对所选定的监测项目进行监测，并对所取得数据用适当的方法进行统计处理。常见的评价参数有30多项，包括：

一般水质参数：水温、色度、透明度、悬浮固体、电导率、硬度、碱度、总矿化度、总盐量等；

氧平衡参数：溶解氧、化学需氧量、生化需氧量等；

重金属参数：汞、镉、铬、铅、铜、锌、铁、锰等；

有机污染参数：酚、油类及其他特殊有毒有害污染物等；

无机污染参数：氨氮、硝酸盐氮、硫酸盐、磷酸盐、磷化物、氰化物、氯化物等；

生物学参数：细菌、大肠杆菌、无脊椎动物、藻类等以及放射性核素等。

表 7.2 水库湖泊环境质量标准基本项目限值 GB3838—2002

序号	指标	I 类	II 类	III 类	IV 类	V 类
1	水温 $^{\circ}\text{C}$	人为造成的环境水温变化应限制在： 周平均最大温升 ≤ 1 ；周平均最大温降 ≤ 2				
2	PH 值	6---9				
3	DO(饱和率%) \geq	9(7.5)	6	5	3	2
4	高锰酸盐指数 \leq	2	4	6	10	15
5	化学需氧量 \leq	15	15	20	30	40
6	BOD ₅ \leq	3	3	4	6	10
7	$\text{NH}_3 - \text{N}$ \leq	0.15	0.5	1.0	1.5	2.0
8	TP(总磷) \leq	0.01	0.025	0.05	0.1	0.2
9	TN(总氮) \leq	0.2	0.5	1.0	1.5	2.0
10	铜 \leq	0.01	1.0	1.0	1.0	1.0
11	锌 \leq	0.05	1.0	1.0	2.0	2.0
12	氟化物(F^-) \leq	1.0	1.0	1.0	1.5	1.5
13	硒 \leq	0.01	0.01	0.01	0.02	0.02
14	砷 \leq	0.05	0.05	0.05	0.1	0.1
15	汞 \leq	0.00005	0.00005	0.0001	0.001	0.001
16	镉 \leq	0.001	0.005	0.005	0.005	0.01
17	铬(6价) \leq	0.01	0.05	0.05	0.05	0.1
18	铅 \leq	0.01	0.01	0.05	0.05	0.1
19	氟化物 \leq	0.005	0.05	0.2	0.2	0.2
20	挥发酚 \leq	0.002	0.002	0.005	0.01	0.1
21	石油类 \leq	0.05	0.05	0.05	0.5	1.0
22	阴离子表面活性剂 \leq	0.2	0.2	0.2	0.3	0.3
23	硫化物 \leq	0.05	0.1	0.2	0.5	1.0
24	粪大肠菌群(个/L) \leq	200	2000	10000	20000	40000

在水质评价中应当依据评价目的不同和水质特点,选用不同水质参数^[249-233]。

本文采用国家环境保护总局制定的《地表水环境质量标准》(GB3838-2002)^[219]。依据《北京市海河流域水污染防治规划》中水体功能类别的划分,不同功能水体执行其相应类别的标准。监测项目包括:PH、温度、电导率、溶解氧、五日生化需氧量、化学需氧量、总氮、总磷、氨氮、锌、铜、铅、镉、铬、铁、锰。

7.2.4 评价方法的选择

现在实行的地表水环境质量标准(GB3838-2002)(参见表 7.2)中要求,地表水环境质量评价应根据应实现的水域功能类别,选取相应类别标准,进行单因子评价,评价结果应说明水质达标情况,超标的应说明超标项目和超标倍数。实际上国标中对地表水综合评价均已确定悲观的原则,即以单项指标最差所属级别确定其综合水质级别,这是符合实际的,因为若某项指标浓度严重超标,已不适于饮用,却仍评为可作为饮用水源地的三类水,将给人民生命健康带来威胁,因此选取单因子评价法对北京市地表水体进行评价。但是单因子评价法各评价参数之间互不联系,没有可比性,因此选择了支持向量机,可以全面反映了河流水体污染的综合情况,可以与国家地表水质量标准统一起来,具有可比性,至于计算过程复杂问题可通过计算机编程来解决。单因子评价法可以列出超标污染物及其超标倍数,而改进的循环决策二叉树的支持向量机多值分类方法,可以全面反映河流水体污染的综合情况,两种方法结合起来可以很好地评价水质。

7.2.5 SVM 综合评价

在进行湖泊水库的水质评价时,需采用国家地表水质量标准中对湖泊、水库的特殊要求标准,主要体现在对总磷和总氮两个指标比一般的地表水体要求更为严格。

表 7.3 北京市十大湖泊水库监测结果

采样地点	珍珠湖	昆明湖	玉渊潭	雁栖湖	密云水库	怀柔水库	官厅水库	海子水库	霞云岭水库	沙河水库
水温	4	15.5	15	20	9	18	3	7	14	5
NH ₃ -N	0.195	3.15	0.570	0.492	0.370	0.327	0.504	0.346	0.265	20.77
NO ₃ -N	1.93	0.70	1.16	0.78	1.22	1.45	4.71	1.57	3.00	5.88
总磷	0.00	0.035	0.03	0.00	0.00	0.00	0.14	0.01	0.01	2.26
总氮	2.3	11.5	2.0	1.68	2.0	2.01	70	2.1	3.5	34
COD	12	12	16	6	4	11	44	5	6	42
BOD	1	2.5	4	1	3	3	15	0	2	15
DO	6	10	9	9	11	8	11	13	10	12
PH	8	8	8	8	8	8	8	8	8	8
Ec	491	433	687	424	354	476	465	368	422	1105
Fe	0.31	0.29	0.15	0.03	0.05	0.31	0.12	0.14	0.07	0.04
Mn	0.01	0.00	0.00	0.00	0.02	0.06	0.01	0.00	0.04	0.13
Zn	0.10	0.08	0.01	0.91	0.01	0.10	0.01	0.00	0.10	0.01
Cu	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01
Cr	0.00	0.00	0.00	0.05	0.01	0.05	0.00	0.00	0.10	0.01
Pb	0.05	0.02	0.02	0.04	0.01	0.04	0.05	0.03	0.00	0.01
Cd	0.001	0.000	0.002	0.002	0.003	0.002	0.000	0.005	0.005	0.00

本文采用在 2003 年 3 月 24 日至 4 月 22 日的枯水期以及 7 月 15 日至 9 月 28 日的丰水期的采样数据,选择湖泊 4 个,水库 6 个,具体的实测结果见表 7.2。

具体步骤如下:

(1)选择训练数据

本文参照地表水环境质量标准(GB3838-2002), 对应于五类评价结果, 试验性地构造随机数据, 最后选择合乎要求的数据作为训练样本。

(2)确定属性指标

从监测的 16 项指标中, 选择溶解氧 DO、五日生化需氧量 BOD₅、化学需氧量 COD_{Cr}、总磷 TP、氨氮 NH₃-N、锌 Zn、铜 Cu、铅 Pb、锡、铬 Cr 共 10 项作为评价指标。

(3)数据的归一化无量纲处理

由于水质标准中各个指标的量级不同, 必须在评价之前进行数据的归一化无量纲处理。本文采用极差化处理方法, 即 $x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$, 其中 x_{\max}, x_{\min} 分别为监测数据中的最大与最小值。

(4)评价方法与结果

本次评价遵循地表水综合评价的悲观原则, 所以选择单因子评价法和改进的循环决策二叉树的支持向量机多值分类方法, 其中单因子评价法可以列出超标污染物及其超标倍数, 支持向量机可以与国家地表水质量标准统一起来, 全面反映了河流水体污染的综合情况, 同时具有可比性。两种方法结合起来可以更好地评价水质。

对北京市十大湖泊水库的单因子评价结果表明(表 7.3), III 类水体功能要求的湖泊符合水质功能要求, II 类湖泊不符合水质功能要求, 但只有一种污染物 Pb 超标, 说明北京市湖泊水质比较好。北京市水库的评价结果表明, 除了密云水库、海子水库符合水质功能要求外, 其余的水库均有污染物超标, 但超标倍数不大, 其中官厅水库、沙河水库水质最差, 为劣 V 类水质。

北京市湖泊、水库 SVM 评价结果表明(表 7.4), 监测的 4 个湖泊有 2 个达标, 水库除密云水库和海子水库外, 其余的水库均不达标, 这与单指标评价的结果是一致的。

表 7.4 北京市十大湖泊水库水质的单因子评价结果

采样地点	珍珠湖	昆明湖	玉渊潭	雁栖湖	密云水库	怀柔水库	官厅水库	海子水库	霞云岭水库	沙河水库
水质分类	II	II	III	III	II	II	IV	III	III	IV
超标污染物及其超标倍数	Pb (3.90)	Pb (2.39)				Cr (0.18) Pb (2.72)	COD _{Cr} (1.94) BOD ₅ (4.12) Pb(3.98)		Cr (0.90)	NH ₃ -N (12.85) 总磷 (7.53) BOD ₅ (4.12)
评价结果	III	III	III	III	II	IV	>V	III	V	>V

综上所述，湖泊水质比较好，基本符合水质功能的要求。除了密云水库、海子水库达标外，其余的水库均受到不同程度的污染，其中官厅水库、沙河水库水质最差，为劣 V 类水质。

表 7.5 北京市十大湖泊水库水质的 SVM 评价结果

采样地点	珍珠湖	昆明湖	玉渊潭	雁栖湖	密云水库	怀柔水库	官厅水库	海子水库	霞云岭水库	沙河水库
水质分类	II	II	III	III	II	II	II	III	III	IV
模糊数学综合评价结果	III	III	II-III	II	I-II	IV	V	II	V	V
SVM 评价结果	III	III	III	III	II	IV	V	II	IV	V

采用单因子评价法、模糊数学综合评价法、SVM 法的评价结果差异性及与河流水质功能进行比较参见(表 7.4)(表 7.5)。

饮用水源问题关系到整个北京的稳定和发展，加强饮用水源的保护，保证北京饮用水安全、优质是当前的主要任务。经过上面的分析可看出，官厅、密云两大水库上游的来水污染日趋严重，尤其是官厅水库水质目前大于Ⅴ类，水体呈富营养化状态。密云水库水质目前虽符合饮用水标准，但随着总磷、总氮等营养盐类的增加，水体有向富营养化发展的趋势。因此，遏制官厅、密云水库水质恶化，对水源地必须采取有力的保护防治措施。

§ 7.3 本章小结

本章根据水质评价的原则与要求，结合单因子评价方法，建立了改进的循环决策二叉树支持向量机多值分类的北京市湖泊水库的水质评价模型，并对其结果进行了比较分析。改进的评价方法意义更清楚，它既突出了单项超标的效果，体现了水质评价的基本原则，也考虑了各项指标的综合作用效果；同时该方法具有以下特点：所需训练样本少，计算简便，误差较小，符合研究对象所具有的客观特征。该方法不仅适用于水环境质量评价，它可以拓广到一类定量的评价问题处理。其评价结果可以为北京市湖泊水库进一步规划、管理与利用提供一定的理论依据。

第八章 总结与展望

§ 8.1 全文总结

支持向量机(Support Vector Machines, SVM)是复杂非线性科学和人工智能科学的研究前沿,其在水资源的水质预测与评价领域的应用研究在国内外尚处于初创阶段。本文在较全面分析评述了水质预测与评价研究现状,在分析阐述了 SVM 的基本原理与算法基础上,在国内首次将 SVM 方法引入水环境质量预测及其质量综合评价中的应用领域,主要在水质预测和水质评价的 SVM 建模方法以及算法方面进行了一些创新性的研究工作,为提高水质评价和水质模拟的智能化水平做出了努力。

水质预测与质量评价是认识和研究水环境的一项重要内容,其目的是准确反映水环境的质量和污染状况,预测未来的发展趋势,是水环境管理保护和治理的一项重要基础性工作。目前在进行水质预测与评价时,存在的主要问题有:迄今没有一个被大家公认通用的具有可比性的水质监测和评价数学模型,各部门进行评价时,选用数学模型的任意性很大。如传统的统计分析方法以及神经网络方法等。这些方法虽然在实际应用中占据主导地位,但人们也发现它们还存在着许多不足之处。传统的一些水质评价方法如多因子综合评价中确定环境因子权重存在任意性,缺乏比较客观可靠的确定环境因子权重的方法,因而常常造成不能反映本地区污染状态,同时也不便于与其他地区水质状况进行比较等一些情况。而且传统的统计方法一般需要事先知道样本的先验分布,并要求有足够多的样本数据,但是在实际水质数据监测与评价中,难以布设足够多的水质监测点以满足网络训练数据量的要求,而这些要求在实际应用中往往难以达到,这就使其在实际应用中效果往往并不理想。神经网络方法虽然很好地解

决了非线性问题，但由于其自身存在着结构不易确定、易陷入局部极小等固有的缺陷，从而限制了其实际应用。另外，神经网络的学习算法仅仅试图使经验风险最小化，并没有使期望风险最小化，在原理上缺乏实质性的突破，这也是神经网络“过拟合”现象产生的原因，从而导致了其推广能力的下降。

支持向量机是在统计学习理论上开发出来的一种新的、非常有效的机器学习新方法，是建立在统计学的 VC 维 (Vapnik-Chervonenks Dimension) 理论和结构化风险最小化原则 (Structural Risk Minimization SRM) 的基础上。它较好地解决了以往困扰很多学习方法的小样本、非线性、过学习、高维数、局部极小点等实际问题，具有很强的推广能力。目前，理论和实验表明，SVM 不但结构简单，而且具有较好的泛化能力，尤其是对于小样本问题，成功地克服了神经网络学习过程中的“过学习”现象和可能陷入局部极小等问题。统计学习理论和支持向量机作为小样本学习的最佳理论，开始受到越来越广泛的重视，正在成为人工智能和机器学习领域新的研究热点。

水质的预测与评价对应于支持向量机多值分类和回归估计算法。水质评价和水质预测问题，是水体分类和水质参数回归的问题，因此新兴的机器学习支持向量机能够解决水质评价和水质预测问题。本文基于这一出发点，详细研究了支持向量机的理论与方法，探讨将支持向量机算法应用于水环境的水质预测及其质量综合评价的可行性，并运用于实际问题的研究。其结果表明，应用支持向量机理论与算法对水质进行监测与评价，结果是可行和合理的，同时也拓展了支持向量机理论的应用范围。同时这也将为水资源的合理规划和可持续地开发利用提供决策依据和技术支持。

本文所做的具体工作概括如下：

1) 研究分析了水环境质量预测和评价方法的产生背景、发展过程和国内外的研究现状，并从水环境的水质监测与评价的客观性、公正性的角度，论述了支持向量机应用于水质监测与评价研究的合理性和可行性。

2) 详细地阐述了支持向量机的理论基础——统计学习理论的若干原理, 并对支持向量机算法的发展历史和现状及其主要研究内容作了概述。

3) 详细讨论了现有两类分类问题的支持向量机算法及其变形算法。通过对这些变形的支持向量机分类算法进行客观、全面地讨论与研究, 探讨了各种算法的优缺点和适用范围等情况, 为构造基于支持向量机分类理论的水环境质量综合评价提供了理论准备。

4) 重点阐述用于曲线拟合的支持向量回归机的基本原理, 然后详细地阐述并分析了各种支持向量回归机模型, 包括 ε -支持向量回归机、 ν -支持向量回归机、LS-支持向量回归机、W-支持向量回归机、基于线性规划的支持向量回归机, 同时, 建立了在 ε -不敏感函数下的单参数约束下的支持向量回归模型, 证明了单参数回归模型与标准回归模型的等价性, 并进行数值实验; 最后探讨了支持向量回归算法优劣的评价依据, 从而完善了支持向量回归机的优化理论基础和决策依据。

5) 全面总结了目前存在的基于支持向量机的多类别分类方法, 包括“一对多”方法、“一对一”方法、超球面分类算法、二叉树的支持向量机多值分类方法等等, 比较了它们的优缺点及性能, 并通过实验对其中的几种常用的方法进行了验证与比较。

6) 在综合研究 SVR 算法的基础上, 针对水环境中污染物浓度的监测问题, 将水质参数平均浓度值的监测看成一个时间序列预测问题, 提出了一种新的支持向量回归机的优化算法。然后, 利用改进的支持向量回归机对实验数据进行回归估计分析, 并与 BP 网络方法进行了比较。从评价结果来看, 利用 SVM 方法建立的评价模型, 能充分利用训练样本的分布特性, 评价结果与实际情况更为相符, 为湖泊富营养化评价及其他环境质量评价工作提供了一种新的有效的方法。

7) 运用改进的循环决策二叉树的支持向量机多值分类方法, 建立了北京市

湖泊水库的水质评价模型，并对其结果进行了比较分析。改进的评价方法意义更清楚，体现了水质评价的基本原则，也考虑了各项指标的综合作用效果；同时该方法具有以下特点：所需训练样本少，计算简便，误差较小，符合研究对象所具有的客观特征。其评价结果可以为北京市湖泊水库进一步规划、管理与利用提供一定的理论依据。该方法不仅适用于水环境质量评价，它可以拓广到一类定量的评价问题处理。

本文研究的创新点，主要有以下几个方面：

1. 首次将 **SVM** 方法引入水环境质量预测及其质量综合评价中的应用领域，主要在水质预测和水质评价的 **SVM** 建模方法以及算法方面进行了一些创新性的研究工作，提高了水质综合评价和水质预测的智能化水平。
2. 建立了支持向量回归算法优劣的评价依据，从而完善支持向量回归机的优化理论基础和决策依据。
3. 拓展 **SVM** 的应用范围，提出改进的 **SVM** 多值分类模型，构建一种新的水库湖泊水质质量综合评价方法（第六章）。
4. 提出改进的 **SVR** 算法并建立一种新的水质预测模型（第七章）。

本文将支持向量机理论与算法应用到水质预测与评价还是一个新的尝试，取得了一些有价值的研究成果。与传统评价方法相比，该方法所需训练样本少、计算简便、快捷、精度高，具有较大的实用性，且评价结果更具有客观性。同时还能根据有限的样本信息在模型的复杂型和学习能力之间寻求最佳折衷，具有很强的推广能力。

§ 8.2 存在的问题和研究前景展望

支持向量机是一种非常有发展前景的融合算法，它建立在统计学习理论的 VC 维理论和结构风险最小化原理之上，根据有限的样本信息在模型的复杂性和

学习能力之间寻求最佳折中，以期获得最好的泛化能力。同时，它在实际中的一些应用也说明了其实用价值。但是由于支持向量机尚处在发展阶段，很多方面尚不完善，很多关于支持向量机和 VC 维的理论和应用问题都亟待研究。如：核函数和参数的选择缺乏理论指导；支持向量机的训练集的规模同训练速度之间的矛盾；多类 SVM 的优化设计等等。

将支持向量机算法应用到水质监测与评价还是一个较新的尝试。在本文的 SVM 水质预测与评价模型构建中，还有以下几个方面的工作值得以后进一步深入研究：

1) 在支持向量机的训练过程中，如何找到结构简单而且推广能力好的算法模型，如何选择适合于水质监测与评价各模型的更好的 SVM 核函数及参数，是今后需要研究的方向。

2) 本文主要是进行了实验研究，如何将这些方法有效地应用到实际研究工作当中去，与实际问题相结合，即应用研究，也是今后研究工作的一部分。

3) 在 SVM 的应用研究方面，其应用领域越来越广泛，如建模、参数辨识和自适应控制等问题，但与理论发展相比，其他方面如时间序列分析、回归估计、聚类分析等方面的研究，还有待进一步的完善，应用领域还有待拓展。

4) 由于河流水质综合评价自身存在的特殊性，迄今没有一个被大家公认通用的、具有可比性的水质综合评价数学模型。各地各部门在进行水质评价时，选用模型的任意性很强，评价的结果不便于与其它地区水质状况进行比较，国家难以全面掌握水环境污染的时空分布态势。许多实际工作者已经注意到，数学模型的有效性并不同数学模型本身的繁杂程度来决定。有些数学模型模式很复杂，模型本身看来很完善，却忽视了水质评价的特性，难以在实际工作中发挥作用。因此水质综合评价工作应从实际出发，要从简练明了入手，达到评价目的即可，应该发展和建立标准化的计算方法和评价方法，使评价结果有代表性、可比性、实用性。

支持向量机技术是一种新兴的技术，随着各种算法研究以及计算机硬件发展的日益成熟及其本身具有的优良特性，必将会应用到社会的各个方面，大大推动科学技术的进步，因此，今后支持向量机在各个领域的应用研究，将是一个很重要的研究方向。

参考文献

1. 石虹. 浅谈全球水资源态势和中国水资源环境问题[J]. 水土保持研究, 2002, 9 (1): 145-150.
2. 陈静生. 河流水质全球变化研究若干问题[J]. 环境化学, 1992, 11(2): 43-51.
3. 陈静生. 陆地水水质变化研究国内外进展[J]. 环境科学学报, 2000, 20 (1): 10-15.
4. 叶常明. 水环境数学模型的研究进展[J]. 环境科学进展, 1993, 1(1): 74-80.
5. 邓聚龙著. 灰色理论系统[M]. 武汉: 华中科技大学出版社, 2002.
6. 王开章, 刘福胜, 孙鸣. 灰色模型在大武水源地水质预测中的应用[J]. 山东农业大学学报 (自然科学版), 2002, 33(1): 66-71.
7. 吕海涛, 赵林明. 应用灰色模型时注意的一些问题[J]. 华北水利水电学院学报, 2000, 21(1): 76-78.
8. 王季震, 韩宏大, 赵洪宾, 何文杰. 天津市城市用水量模拟方法的研究[J]. 给水排水, 2001, 27(10): 33-33.
9. Holger R. Maier and Graeme C. Dandy. The use of artificial neural networks for the prediction of water quality parameters[J]. Water Resources Research, 1996, 32(3): 54-61.
10. 徐立中, 王建颖, 黄凤辰, 孙少华. 基于神经网络的水环境监测数据融合处理与评价模型[J]. 测试技术学报, 2003, 25: 103-104.
11. 蒋林, 谢晓燕. 基于人工神经网络的非线性组合预测方法研究[J]. 西安邮电学院学报, 1999, (1): 32.
12. 于磊, 赵君明. 统计学[M]. 上海: 同济大学出版社, 2003.
13. 赵志坚. 灰色聚类法在水环境质量评价中的应用[J]. 四川环境, 1997, 16(3): 49-51.
14. 朱雷, 陈威. 模糊综合指数法在水质评价中的应用[J]. 武汉理工大学学报, 2001, 23(8): 61-65.
15. 费忠华, 罗定贵, 周利麟. 地下水环境质量评价模糊数学模型新探[J]. 工程勘察, 1995(6): 15-20.
16. 陈昌彦, 相桂生. 人工神经网络在地下水水质评价中的应用[J]. 水文地质工程地质, 1996, (6): 39-42.
17. 韩波, 孙利. 水质评价模式识别的 BP 神经网络方法[J]. 广州环境科学, 2005, 20(4): 30-33.
18. 朱静平. 几种水环境质量综合评价方法的探讨[J]. 西南科技大学学报, 2002, 17(4): 62-67.
19. 陆洲, 夏秋颖, 周琳, 马涛. 等斜率灰色聚类法在地面水环境质量评价中的应用[J]. 环

20. 肖晓柏, 许学工. 地表水环境质量灰关联评价方法探讨[J]. 环境科学与技术, 2003 26 (3): 34-36.
21. Lee H K, Oh K D, Park D H, et al. Fuzzy expert system to determine stream water quality classification from ecological information [J]. Water Science and Technology, 1997, 36 (12):199-206.
22. 朱雷, 陈威. 模糊综合指数法在水质评价中的应用[J]. 武汉理工大学学报, 2001, 23(8):61-65.
23. Yi-Ming Kuo, Chen-Wuing Liu, Kao-Hung Lin, Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan [J]. Water Research, 2004 (38)148-158.
24. 张乃尧, 阎平凡. 神经网络与模糊控制[M]. 北京: 清华大学出版社, 1999.
25. Loke E, Warnars E A, Jacobsen P, et al. Artificial neural networks as a tool in urban storm damage [J]. Water Science and Technology, 1997, 36(8-9): 100-110.
26. Karul C, Soyupak S, Cilesiz A F, et al. Case studies on the use of neural networks in eutrophication modeling [J]. Ecological Modeling, 2000, 134(2-3): 145-152.
27. Nello Cristianini, John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods [M]. New York: Springer, 2000.
28. Filip Mulier, Vapnik-Chervonenkis (VC) Learning Theory and Its Applications, IEEE Translation On Neural Networks, 1999, 10(5).
29. Vladimir N. Vapnik 著, 张学工译, 统计学习理论的本质[M]. 北京: 清华大学出版社, 2000.
30. Vapnik V. and Chervoknenkis A. Y, The necessary and sufficient conditions for the uniform convergence of averages to their expected values. Teoriya Veroyatnostei I Ee Primeneniya, 1981, 26(3): 543-564.
31. Vapnik V. Estimation of dependencies based on empirical data (in Russian) [M]. Nauka, Moscew. 1979(英译本): Vapnik V. Estimation of dependencies based on empirical data. Springer, New York, 1982).
32. Boser B., Guyon L., and Vapnik V. A training algorithm for optimal margin classifier, In Fifth annual Workshop on Computational Learning Theory, Baltimore, MD: ACM Press, 1992:144-152.
33. Cortes C., Vapnik V, The soft margin classifier, Technical memorandum 11359-931209-18TM, AT&T Bell Labs, 1993.
34. Vapnik V., The nature of statistical learning theory. Springer-Verlag, New York, 1995.

35. Vapnik V, Gokowich S. and Smola A., Support vector method for function approximation, regression estimation, and signal processing. In: Advances in Neural Information processing Systems 9, Cambridge, MA, MIT Press, 1997:281-287.
36. Suykens.J.A.K, Vandewalle J., De Moor B., Optimal control by least support vector machines[J]. Neural Networks, 2001, 14(1):23-35.
37. Bureges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121-167.
38. Mukhejee S, Osuma E, Girosi F. Nonlinear prediction of chaotic time series using support vector machines[J]. Neural Networks for Signal Processing, 1997, 7:511-520.
39. Vapnik V, Lerner A. Pattern recognition using generalized portrait method[J]. Automation and Remote Control, (24), 1963.
40. Vapnik V, Chervoknenkis A. On the uniform convergence of relative frequencies of events to their probabilities[J]. Doklady Akademii Nauk USSR, 181(4), 1968.
41. Vapnik V, Chervoknenkis A. On the uniform convergence of relative frequencies of events to their probabilities[J]. Theory of Probability and its Application, 1971, 16(2): 264-280.
42. Vapnik V. Estimation of dependence based on empirical data[M]. New York., Springer-Verlag, 1982.
43. Vapnik V, Estimation of dependence based on empirical data. New York, Springer-Verlag, 1982.
44. Vapnik V. and Chervoknenkis A. Y. The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognition and Image Analysis, 1991,1(3): 283-305.
45. Vapnik V., Levin E., Cun Y Le, Measuring the VC-dimension of a learning machine. Neural Computation, 1994, 6.
46. Vapnik V, Statistical learning theory[M]. New York:John Wiley& Sons, 1998.
47. Vapnik V., An overview of statistical learning theory, IEEE Trans. On Neural Networks, 1999, 10(5).
48. Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995, (20): 1-25.
49. Schölkopf B., Burges C. J. C and Vapnik V, Extracting support data for a given task. In: Fayyad U M, Uthurusamy R, eds. Proceedings of First International Conference on Knowledge Discovery & Data Mining. German: AAAI Press, 1995: 262-267.
50. Schölkopf B., Burges C., Vapnik V. Incorporating invariances in support vector learning machines, In: von der Malsburg C, von Seelen W, Vorbruggen J C et al (eds). Artificial

- Neural Networks ICANN'96, Spingers Lecture Notes in Computer Science, Berlin, 1996, 11(12): 47-52.
51. Schölkopf B., Comparing support vector machines with Gaussian kernels to radial basis function classifier, IEEE Transactions on signal processing, November 1997, 45(11).
 52. Schölkopf B., Smola A, Muller K R., Kernel principal component analysis. In: Proc. of ICANN'97, 1997: 583-589.
 53. Schölkopf B. Smola A., A tutorial on support vector regression[R]. NeuroCOLT2 Technical Report Series NC2-TR-1998-030, October. 1998
 54. Schölkopf B., Smola. A, and Vapnik V., Prior knowledge in support vector kernels, in Advances in Neural Information Proceeding Systems 10, M.Jordan, M. Kearns, and S. Solla, Eds., Cambridge, MA. MIT Press, 1998: 640-646.
 55. Schölkopf B., Williamson R., and Shawe-Taylor J., Single-class support vector machines. In J. Buhmann. W. Maass, H. Ritter and N. Tishby, editors, Unsupervised Learning, Dagstuhl-Seminar-Report 235, 1999: 19-20.
 56. Schölkopf B., Burges C. J. C., and Smola A. J., Advances in kernel methods support vector learning, MIT Press, Cambridge, MA, 1999.
 57. Schölkopf B., Smola A., and Muller K-R., Kernel principal component analysis. Advances in Kernel Methods- Support Vector Learning. MIT Press, Cambridge, MA, 1999:327-352.
 58. Schölkopf B., Smola A., Williamson R. C. et al, New support vector algorithms [J]. Neural Computation, 2000, 12(5): 1207-1245.
 59. Schölkopf, B., J. C. Plat, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. Neural Computation, 2001, 13 (7): 1443-1471.
 60. Smola A. Generalization bounds for convex combinations of kernel functions. Alex J. Smola, GMD. Neu-roCOLT2 Technical Report series, NC2-TR-1998-020, July, 1998.
 61. Weston J. Extensions to the support vector method. Ph.D thesis, Royal Holloway University of London, 1999.
 62. M. Anthony, P. Bartlett. Learning in Neural Networks: Theoretical Foundations. Cambridge University Press, 1999.
 63. J. Shawe-Taylor, N. Cristianini. Margin Distribution and Soft Margin. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, Advances in Large Margin Classifiers, MIT Press, Cambridge, MA, 2000: 349-358.
 64. J. Weston. Leave-One-Out Support Vector Machines. IJCAI 1999: 727-733
 65. A. Smola, B. Schölkopf, K.-R. Muller. Convex Cost Functions for Support Vector

- Regression. In L. Niklasson, M. Bodtn, and T. Ziemke, editors, Proceedings of the Eighth International Conference on Artificial Neural Networks, Perspectives in Neural Computing, Berlin, Springer-Verlag.1998.
66. Burges C.J.C A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 1998.
 67. Smola A. Learning with kernels. PH.D thesis, 1998.
 68. Yang. Y, Expert network: Effective and efficient learning from human decisions in text categorization and retrieval. In Proceedings of the Seventeenth International ACM SIGIR Conference on Research and Development in Information Retrieval, 1994: 13-22.
 69. Yang. Y, Jan O. Pederson. A comparative study on feature selection in text categorization. Proceedings of the 14th International Conference on Machine Learning, Nashville, TN, Morgan Kaufmann, 1997: 412-420.
 70. Osuna E., Freund R, Girosi F. An improved training algorithm for support vector machine[A]. Proceedings of 1997 IEEE workshop on neural networks for signal processing[C]. Amelée Island, FL: IEEE, 1997: 276-285.
 71. Osuna E., Freund R., Girosi F., Training support vector machines: An application to face detection. In: Proceedings of CVPR'97, New York, NY, IEEE,1997: 130-136.
 72. Joachims T., Text categorization with support vector machines[R].Technical report, LS VIII Number 23, University of Dortmund, German, 1997.
 73. Joachims T., Making Large-Scale SVM Learning Practical. In:Scholkopf B, Burges C Smola J C, A eds. Advances in Kernel Methods Support Vector Learning, Cambridge, MA:MIT Press, 1998: 169-184.
 74. Joachims T., Transductive inference for text classification using support vector machine. In: Proceedings of the Sixteenth International Conference on Machine Learning. Morgan Kaufmann, 1999: 148-156.
 75. U.Kreßel. Pairwise classification and support vector machines, In Advances in Kernel Methods-Support Vector Learning, Cambridge, MA:MIT Press, 1999:255-268.
 76. Cai Yu-Dong, Liu Xiao-Jun, Xu Xue-biao et al. Prediction of protein structural classes by support vector machines[J]. Computers and Chemistry,2002, 26(3): 293-296.
 77. Kwok J T Y, Support vector mixture for classification and regression problems, ICPR'98, 1998.
 78. Eosuna E., Freund R. and Girosi F., Support vector machines: training and applications, A. I. Memo 1602, MIT Artificial Intelligence Laboratory, 1997.

79. 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 1988.
80. 黄凤岗, 宁克欧, 模式识别[M], 哈尔滨: 哈尔滨工业大学出版社, 1998.
81. Cortes C, Vapnik V. Support vector Networks. Machine Learning[J], 1995 (20): 273-297.
82. Burges C J C, Schölkopf B, Burges C. and Smola A. Geometry and invariances in kernel based methods, in advances in kernel methods-support vector learning, Eds., Cambridge, MA, MIT Press, 1999: 89-116.
83. Yi-Ming Kuo, Chen-Wuing Liu, Kao-Hung Lin, Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Research, 2004 (38) 148-158.
84. 张学工, 关于统计学习理论与支持向量机[J], 自动化学报, 2000, 26 (1): 32-42.
85. 萧嵘, 王继成, 张福炎, 支持向量机理论综述[J], 计算机科学, 2000, 26 (3) 1-3.
86. Zhang X G. Using class-center vectors to build support vector machines, NNSP'99, 1999.
87. Suykens J. and Vandewalle J., Least square support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293-300.
88. Suykens J., Branbanter J D, Lukas L, et al. Weighted least squares support vector machines: robustness and sparse approximation[J]. Neurocomputing, 2002, 48(1): 85-105.
89. Le C Y, Jackel L D, Bottou L, et al. Comparison of learning algorithms for handwritten digit recognition. In F. Fogelman-Soulie and P. Gallinari, editors, Proceedings ICANN'95 - International Conference on Artificial Neural Networks, volume II, 1995, pages 53-60.
90. Joachims T. Text categorization with support vector machines. In Proceedings of European Conference on Machine Learning (ECML), 1998.
91. Joachims T. Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms[M]. Kluwer, 2002.
92. Ancona N, Cicielli G, Stella E, et al. Ball detection in static images with Support Vector Machines for classification[J]. Image and Vision Computing 21(2003): 675-692.
93. 邓乃扬, 田英杰. 数据挖掘中的最优化方法---支持向量机[M]. 北京: 科学出版社, 2004.
94. Chih-Chung Chang, Chih-Wei Hsu, et al. The Analysis of Decomposition methods for Support Vector Machines[J]. IEEE Transactions on Neural Networks, 2000(11), 4: 1003-1008.
95. Chih-Jen Lin, et al. On the convergence of the decomposition method for support vector machines. IEEE Transactions on Neural Networks, 2001(12), 6: 1288-1298.
96. D. Tax, R. Duin. Support Vector Domain Description. Pattern Recognition Letters, 1999, 20(11-13): 1191-1199.

97. Edgar Osuna, Robert Freund, Federico Girosi. Training support vector machines: an application to face detection[A]. In: IEEE Conference on Computer Vision and Pattern Recognition[C], Puerto Rico, USA, 1997: 130-136.
98. Friess,T.T., Cristianimi, N.,&Campbell, C. The kernel adatron algorithm: a fast and simple learning procedure for support vector machines. In Proceeding of 15th Intl. Conf. Machine Learning. Morgan Kaufman Publishers, 1998.
99. Mangasarian, O. L.,& Musicant, D. R. Successive overrelaxation for support vector machines. IEEE Trans. Neural Networks, 1999,10(5): 1032-1037.
100. Mangasarian O. L., Generalized support vector machines. In B. S. A. J. Smola, P. Bartelett and D. Schuurmans, editors, Advances in large margin classifiers, MIT Press, 2000: 135-146.
101. Mangasarian O. L. and Musicant D. R., Lagrangian support vector machines. Journal of Machine Learning Research, 2001,1:161-177.
102. Chang Chih-Chung, Lin Chih-Jen. Training ν -support vector classifiers: theory and algorithms[J], Neural Computation, 2001,13(9):2119-2147.
103. Lee Y-J,Mangasarian O. L. SVM: a smooth support vector machine, Technical Report 99-03,Data Mining Institute, and A. J. Smole, editors, Advances in Kernel Methods—Support Vector Learning, MIT Press, 1999, 185-208.
104. Platt J C. Fast training of support vector machines using sequential minimal optimization[A]. Scholkoph B, Burges C J C, Smola A J. Advances in kernel method-support vector learning[C]. Cambridge, MA: MIT Press, 1999:185-208.
105. Platt J. C., Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In Advances in Large Margin Classifiers. MIT Press, 1999.
106. Platt J. C., Cristianini N., and Shawe-Taylor J., Large margin DAGs for multiclass classification. In Advances in Neural Information Processing Systems, MIT Press, 2000, 12: 547-553.
107. Keoman V, Hadzic I. Support vectors selection by linear programming [A].Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. Como, Italy, 2000,5: 193-198.
108. Kreßel, U. Pairwise classification and support vector machines. In B. Scholkopf, C. J. C. Burges, and A. J. Smola (Eds.), Advances in Kernel methods Support Vector Learning, Cambridge, MA, MIT Press, 1999: 255-268.
109. Laskov P. Feasible direction decomposition algorithms for training support vector machines[J]. Machine Learning, 2002, 46(1): 315-349.

110. Lee Y. J. and Mangasarian O. L., RSVM: Reduced support vector machines. In Proceedings of the First SIAM International Conference on Data Mining, 2001.
111. Lin Chun-Fu, Wang Sheng-De. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 464-471.
112. 朱家元, 陈开陶, 张恒喜. 最小二乘支持向量机算法研究[J]. 计算机科学. 2003, 30(7): 157-159.
113. Lin C-J. On the convergence of the decomposition method for support vector machines[J]. IEEE Trans. on Neural Networks, 2001, 12(6): 1288-1298.
114. 杜树新, 吴铁军. 用于回归估计的支持向量机方法[J]. 系统仿真学报. 2003, 15(11): 1580-1585.
115. 杜树新, 吴铁军. 回归型加权支持向量机方法及其应用[J]. 浙江大学学报. 2004, 38(3): 302-306.
116. Chew Hong-Gunn, Bogner Robert E., Lim Cheng-Chew. Dual nu-support vector machine with error rate and training size biasing[J]. Proceedings of 26th IEEE ICASSP 2001, Salt Lake City, USA, 2001: 1269-1272.
117. Keoman V, Hadzic I. Support vectors selection by linear programming [A]. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. Como, Italy, 2000, 5: 193-198.
118. Tax D. and Duin R., Data domain description by support vectors. In Proceedings of ESANN99, ed. M Verleysen, D. Facto Press, Brussels, Baldonado, M., Chang, 1999: 251-256
119. Tax D. and Duin R, Support vector domain description. Pattern Recognition Letters, 1999(20): 119-1199
120. Xing Dong, Wu Zhao-Hui and Pan Yun-He, A new multi-class support vector machines, Proceeding of 2001 IEEE International Conference on Systems, Man and Cybernetics, October 2001: 1673-1676.
121. J. P. Pedroso, N. Murata. Support vector machines for linear programming: motivation and formulations.
<http://citeseer.nj.nec.com/pedroso99support.html>.
122. P. S. Bradley, O. L. Mangasarian. Massive Data Discrimination via Linear Support Vector Machines.
<http://citeseer.nj.nec.com/bradley98massive.html>.
123. O. L. Mangasarian. Arbitrary-norm separating plane. Operation Research Letters. 1999, 1(24): 15-23.

124. A. J. Smola, T. T. Frie13, B. Schölkopf. Semiparametric Support Vector and Linear Programming Machines.
<http://citeseer.nj.nec.com/smola98semiparametric.html>.
125. 周伟达, 张莉, 焦李成. 线性规划支持向量机[J]. 电子学报, 2001, No. 11(29) 1507-1511.
126. 孙德山, 支持向量机分类与回归方法研究, [D]. 杭州, 浙江大学, 2004, 49-56.
127. Weston J. and Watkins C. Multi-class support vector machines, Proc. ESANN99, Brussels, Belgium, 1999.
128. Crammer K and Singer Y On the learnability and design of output codes for multiclass problems, Computer Learning Theory, 2000:35-46.
129. Chih-Wei Hsu, Chih-Jen Lin. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):415-425.
130. Hsu Chih-Wei and Lin C. J., A simple decomposition method for support vector machines. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei. Taiwan. Submitted to Machine learning. 1999.
131. Hsu Chih-Wei, Lin Chih-Jen. A comparison of methods for multi-class support vector machines[J]. IEEE Transactions on Neural Networks 2002, 13(2): 415-425.
132. Knerr, S., L. Personnaz, and G. Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network. In J. Pögelman (Ed.), Neurocomputing: Algorithms, Architectures and Applications. Springer-Verlag, 1999.
133. Chih-Wei Hsu, Chih-Jen Lin. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):415-425.
134. 刘江华, 程君实, 陈佳品. 支持向量机训练算法综述[J], 信息与控制. 2002, (31)1:45-50.
135. Platt J. C., Cristianini N., and Shawe-Taylor J., Large margin DAGs for multiclass classification. In Advances in Neural Information Processing Systems, MIT Press, 2000, 12: 547-553.
136. 范昕炜, 支持向量机算法的研究及其应用[D], 杭州, 浙江大学, 2003, 5, 39-50.
137. Ronan C. Samy B, Support Vector Machines for Large-Scale Regression Problems, IDIAP-RR00-17. [Http://www.idiap.ch](http://www.idiap.ch), 2000.
138. Pavlov D, Mao J, Dom B. Scaling-up support vector machines using boosting algorithm, Proceedings on Pattern Recognition, 2000, 15th International Conference, 2000(12):219-222.
139. Kai-Min Chung, Wei-Chun Kao, et al. Decomposition methods for linear support vector machines, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003(4):868-871.

140. 孙德山, 支持向量机分类与回归方法研究[D], 广州, 中南大学, 2004,49-54.
141. Bennett K, Campbell C. Support vector machines: hype or hallelujah, ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), Expolarations, 2000, 2(2):1-13.
142. Keerthi S, Shevade S. Bhattacharyya C, A Fast Iterative Nearest Point Algorithm for Support Vector Machine Classifier design, IEEE Transactions on Neural Network, 2000,1(1):124-136.
143. Roobaert D. DirectSVM: a fast and simple support vector machine perception, Proceedings of the 2000 IEEE Signal Processing Society Workshop, 2000(1),356-365.
144. 安金龙, 支持向量机若干问题的研究[D], 天津, 天津大学,2004, 6,65-69.
145. Amari S, Wu.Improving support vector machine classifier by modifying kernel function. Neural Networks, 1999,12:783-789.
146. Chen J.H., Chen, C. S. Reducing SVM Classification Time Using Multiple Mirror Classifiers IEEE Transactions on Systems, Man and Cybernetics, 2004(34),2:1173-1183.
147. Maydt J, Lienhart R. A fast method for training support vector machines with a very large set of linear features, IEEE International Conference on Multimedia and Expo, 2002(1),309-312.
148. Balcazar J.L., Yang Dai, Watanabe O. Provably fast training algorithms for support vector machines, IEEE International Conference on Data Mining, 2001:43-50.
149. Chun-Hong Zheng, Li-cheng Jiao. Fuzzy Pre-extracting Method For Support Vector Machine, Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 2002.
150. Glenn Fung and Olvi L. Mangasarian Proximal Support Vector Machine Classifiers. Proceedings KDD-2001 on Knowledge Discovery and Data Mining, 2001,26-29.
151. Zhang Li, Zhou Weida, Jiao Licheng. Pre-extracting Support Vectors for Support vector Machine, Signal Processing Proceedings,2000(3):1432-1435.
152. Yang M H, Ahuja N. A Geometric Approach to Train Support vector Machines. In Proceedings of CVPR 2000, Hilton Head Island, 2000,430-437.
153. Mao, K.Z. Feature Subset Selection for Support Vector Machines Through Discriminative Function Pruning Analysis, IEEE Transactions on Systems, Man and Cybernetics-PART B: CYBERNETICS, 2004 (34) ,1:60-67.
154. Weston . and Watkins C. Muli-class support vector machines,roc. ANN99, ussels,Belgium, 1999.
155. Bredensteiner E J. Bennett K.P., Multi-category classification by support vector machines,

Computer Optimization Application. 1999:53-79.

156. Y.Guermeur. Combining Discriminant Models with New Multi-class SVMs, Neuro COLT tech. Rep, NC-TR-00-086,2000.
157. Crammer K and singer Y On the learn ability and design of output codes for multiclass problems, Computer Learning Theory, 2000:35-46.
158. Yann Guermeur. et al. A new multi-class SVM based on a uniform convergence result, In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000(4):183-188.
159. Kindermann J, Leopold E and Paass G. Multi-Class Classification with Error correcting Codes, In Treffen der G I-Fachgruppe 1.1.3, Maschinelles Lernen,Eds.,2000,GMD Rep.114.
160. E.Mayoraz and E.Alpaydin. Support vector machines for multi-class classification. In IWANN, 1999(2):833-842.
161. Bottou L., cortes C., et al. Comparison of classifier methods: A case study in handwriting digit recognition, In Proc. international conference in Pattern Recognition,1994:77-87.
162. Knerr S, Personnaz L, and Dreyfus G. Single-layer learning revisited: A stepwise procedure for building and training a neural network, In Neurocomputing: Algorithm, Architectures and Applications, New York: Springer-Veerlag, 1990.
163. Suykens J A K, Vandewalle J. Multiclass least squares support vector machines. In IJCNN'99 International Joint Conference on Neural Networks, Washington,DC,1999.
164. Kun-Lun Li, Hou-Kuan Huang. A novel multi-class SVM classifier based on DDAG. Proceedings on Machine Learning and Cybernetics, 2002(3):, 1203-1207.
165. .Hong-Jie Xing, Xi-Zhao Wang, Qiang He, Hong-Wei Yang. The multistage support vector machine, 2002 International Conference on Machine Learning and Cybernetics, 2002(4):1815-1818.
166. Guermeur Y, Elisseeff A, Paugam-Moisy H. A new multi-class SVM based on a uniform convergence result Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000(4):183-188.
167. Schwenker F. Hierarchical support vector machines for multi-class pattern recognition, Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000(2):561-565.
168. Muller K R, Smola AJ, et al. Predicting Time Series with Support Vector Machines. In Proc of the International Conference on Artificial Neural Networks, Lausanne, Switzerland, Springer, 1997,999-1004.

169. Drucker H, Burges C J C, Kaufman L, Smola A, Vapnik V N. Support Vector Regression Machines, *Advances in Neural Information Processing Systems*,9, Cambridge, MA, MIT Press, 1997, 155-161.
170. Gunn S. Support Vector Machines Classification and Regression, ISIS Technical Report, Image Speech & Intelligent Systems Group, University of Southampton, 1988.
171. Tin-Yau Kwok J. Support Vector Mixture for Classification and Regression Problems, In *Proc of 14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998,1:255-258.
172. 陶卿, 曹进德, 孙德敏. 基于支持向量机分类的回归方法[J]. *软件学报*, 2002,3(5): 1024-1028.
173. Tao Qing, Sun De-min, Fan Jin-song, et al. The maximal margin linear classifier based on the contraction of the closed convex hull, *Journal of Software*, 2002,13(3):404-409 (in Chinese)
174. Casdagli M. Nonlinear prediction of chaotic time-series, *Physica D*, 1989, 35:355-356.
175. Müller K R, Smola A J, Rätsch G, et al. Predicting Time Series with Support Vector Machines, In *Proceedings of ICANN'97*, 1997,999-1004.
176. Frontzek T, Navin Lal T, Eekmiller R. Predicting the Dynamics of Biological Neurons Using Support Vector Machines with different Kernels, In *Proc of the International Joint Conference on Neural Networks*, Washington, D C, USA, 2001,2:1492-1497.
177. Sayan Mukherjee, Edgar Osuna, and Frederico Girosi. Nonlinear prediction of chaotic time series using support vector machines, In *Proceedings of IEEE NNSP 97*, Sep. 1997.
178. Baesens B, Viaene S, Van Gestel, et al. An empirical Assessment of Kernel Type Performance for Least Squares Support Vector Machine Classifiers, *Fourth international conference on knowledge-based intelligent engineering systems & allied technologies*, Brighton, UK, 2000:313-316.
179. 卢虎, 李彦, 肖颖. 支持向量机理论及其应用[J], *空军工程大学学报*, 2003,4(4):89-91.
180. Brown M, Lewis H G, Gunn S R. Linear spectral mixture models and support vector machines for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 1999.38(5):2346--2360.
181. Xin Zhao, Kuh A. Adaptive kernel least square support vector machines applied to recover DS-CDMA signals. *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, 2002,1:943-947.
182. Chih-Wei Hsu, Chih-Jen Lin. A comparison of methods for multiclass support vector machines[J]. *IEEE Transactions on Neural Networks*, 2002,13(2):415-425.

183. Zhang Xuegong. Using class-center vectors to build support vector machines, Proceedings of the 1999 IEEE Signal Processing Society Workshop. Neural Networks for Signal Processing IX, 1999. Madison, WI, USA. 1999:3-11.
184. Trafalis T B, Ince, H. Support vector machine for regression and applications to financial forecasting, In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000(6), 348-353.
185. Friedman J H, Turkey J W. A projection pursuit algorithm for exploratory data analysis[J]. IEEE Trans On Computer, 1974, 23(9):881-890.
186. Dillon P J and Rigler F H. A test of a simple nutrient budget model predicting the phosphorus lake water[J]. Journal of Fisheries Research Board of Canada, 1975, 31(1): 1771-1778.
187. Aizaki M, Otsuki A, Fukushima T et al. Application Carlson's trophic state index to Japanese lakes and relationships between the index and other parameters[J]. Verh. Internat. Verein. Limnol. 1981, 21: 675-681.
188. Carlson R E. A trophic state index for lakes[J]. Limn and Oceanogr, 1977, 22(2): 361-369.
189. Dinius S H . Design of an index of water quality [J]. Water Resources Bulletin, 1987, 23 (5): 833-843
190. Norgaard M, Ravn O, Poulsen N K. NNSYSID and NNCTRL tools for system identification and control with neural networks, Computing & Control Engineering Journal, 2001, 12(1):29-36.
191. 张文鸽, 李会安, 蔡大应, 水质评价的人工神经网络方法[J]. 东北水利水电, 2004, 2(10):42-45.
192. 蒋火华, 等. 综合污染指数评价与水质类别判定的关系[J]. 中国环境监测, 1999. 15(6):46-48.
193. Hornik K. some new results on neural network approximation[J]. Neural networks 1993, 6:1069-1072.
194. Alexander GP, Bentito.F.A. An accelerated learning algorithm for multilayer perception network[J]. IEEE Trans. on neural network, 1995, 5(3), 493-497.
195. 陈晓宏, 江涛, 陈俊合. 水环境评价与规划[M]. 广州: 中山大学出版社, 2001, 105-107.
196. 黄兴. 湖流域(苏州片)水质同步监测调查与评价[J]. 环境监测管理与技术. 1999. 11(4).
197. Aguilera P A et al. Application of the kohonen neural network in coastal water management: methodological development for the assessment and prediction of water quality[J]. Water Research, 2001, 35(17):4053-4062.
198. Bhuyan S J, et al. An integrated approach for water quality assessment of a Kansas

- watershed[J]. Environmental Modelling and Software, 2003,18(5): 473-484.
199. Camusso M et al. Ecotoxicological assessment in the rivers Rhine (The Netherlands) and Po (Italy) [J]. Aquatic Ecosystem Health&Management, 2000, 3(3):335-345.
 200. Demuyne C, et al. Evaluation of pollution reduction scenarios in a river basin: application of long term water quality simulations[J]. Water Science and Technology, 1997,35(9): 65-75.
 201. Eilis K V. Surface water pollution and its control[M]. London: Macmillan Press LTD, 1989:15.
 202. Gueguen C, et al. Water toxicity and metal contamination assessment of a polluted river: the Upper Vistula River (Poland) [J].Applied Geochemistry, 2004,19(1):153-162.
 203. Hatan S, Kumar A. Environmental management of effluents: a case study[J]. Fuel and Energy Abstracts, 1997,38 (6): 440.
 204. 胡明星, 郭达志. 湖泊水质富营养化评价的模糊神经网络方法[J]. 环境科学研究,1998,11(4):40-42.
 205. James I D. Modeling pollution dispersion, the ecosystem and water quality in coastal waters: a review. Environmental Modeling&Software, 2002, 17 (4): 363-385.
 206. Jeffrey P, et al. Evaluation methods for the design of adaptive water supply systems in urban environments[J]. Water Science and Technology, 1997, 35 (9): 45-51.
 207. 陈守煜, 陈晓冰. 水质模糊评价理论与模型[J]. 环境科学学报, 1991,11(1): 1-8.
 208. 刘金英, 灰色预测理论与评价方法在水环境中的应用研究[D], 长春, 吉林大学, 2004.10,77-88.
 209. Cherkassky, V., Mulier F., Guest editorial Vapnik---Cherkassky(VC) learning theory and its application, in Transaction on Neural Networks, 1999,10(5).
 210. 杜霞, 武佃伟. 对官厅水库水环境治理问题的思考[J]. 水利, 2004 (1): 30-32
 211. 鲍全盛, 曹利军. 王华东.密云水库非点源污染负荷评价研究[J]. 水资源保护,1997(1):8-11
 212. 高学民. 长江沿程河湖及城市内河水质评价与模拟研究:[D]. 北京:北京大学, 2000
 213. 国家环保总局国家质量监督检验检疫总局.GB3838-2002.地表水环境质量标准[M]. 北京:中国环境科学出版社出版.2002.04-28.
 214. 郝仲勇, 刘洪禄. 北京市水资源短缺及其对策浅析[J]. 北京水利,2000 (5):17-18.
 215. 刘晨阳. 2002 年北京市水情分析[J].北京水利,2003 (3): 29-30.
 216. 张敏. 模糊数学法在地面水质综合评价中的应用[J]. 山东环境,2000 年增刊:92-93.
 217. 潘峰, 付强, 梁川. 模糊综合评价在水环境质量综合评价中的应用研究[J]. 环境工程,2002,20(2):58-61.

218. 薛建军, 姚桂基. 人工神经网络在水质评价中的应用[J]. 水文.1997,13:37-39.
219. 陆雍森. 环境评价(第一版) [M]. 上海:同济大学出版社,1999. 138.
220. Herbrich, R, Graepel, T, Obermayer, K. Support vector learning for ordinal regression, In Ninth International Conference on Artificial Neural Networks, 1999(1), 97-102.
221. 林智仁等, 软件 LIBSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
222. 庞尔鸿. 改善北京城市河湖水环境的几个问题[J]. 北京规划建设,1999 (1): 35-38.
223. 杜桂森等. 密云水库水质现状及发展趋势[J]. 环境科学,1999 20 (2): 110-112.
224. 袁博宇. 官厅水库水质现状及趋势分析[J]. 北京水利,2000 (5): 29-31.
225. Kamal, et al. Assessment of pollution of the river Buriganga, Bangladesh, using a water quality model. [J] Water Science and Technology, 1999, 40(2):129-136.
226. Lu Ruei-Shan, Lo Shang-Lien. Diagnosing reservoir water quality using self-organizing maps and fuzzy theory [J]. Water Research, 2002, 36 (9):2265-2274.
227. 姚焕玫. 基于 GIS 技术的湖泊水质污染综合评价的研究[D]. 武汉, 武汉大学: 2005.489-95.
228. Mailhot Alain, et al. A watershed-based system for the integrated management of surface water quality: the GIBSI system[J]. Water Science and Technology, 1997, 36(5): 381-387.
229. McAvoy D C, et al. Risk assessment approach for untreated wastewater using the QUAL2E water quality model[J]. Chemosphere, 2003, 52(1):55-66.
230. Meybeck M, Helmer R. The quality of rivers: from pristine stage to global pollution. palaeogeography, palaeoclimatology, palaeoecology(Global and Planetary Change), 1989, 75: 283-309.
231. Novotny V. integrated water quality management[J]. Water Science and Technology,1996, 33(4-5): 112-126.
232. Qian. Song S. Water quality model structure identification using dynamic linear modeling: river cam case study revisited[J]. Water Science and Technology, 1997,36 (5):27-34.
233. Roux D J, Kleynhans C J, Thirion C. Biological monitoring and assessment of rivers as a basis for identifying and prioritizing river management options[J]. Water Science and Technology, 1999, 39 (10-11):207-10.

攻读博士期间发表的学术论文及参加的主要科研项目

一、发表的学术论文

1. 徐红敏, 加权支持向量回归机及其在水质预测中的应用[J] 世界地质 2007. vol 26(1):58-61.
2. 徐红敏、王来生、李明, 矛盾方程组的改进正交化算法及其应用[J] 东北师范大学学报 2003. vol 35(10):23-25.
3. 徐红敏、杨天行, 基于支持向量机分类理论的湖泊水环境评价研究[J] 吉林大学学报(地球科学版) 2006. vol 36(4):570-573.
4. 徐红敏、刘亚禄, 改进的支持向量回归机及其在水质预测中的应用[J] 统计与决策 2006 .12(总 227):12-13.

二、参加的科研项目

1. 国家重点基础研究发展规划项目(973 项目)《首都北京及周边地区大气、水、土环境污染机理与调控原理》(项目编号: G199045705) 课题名称《密云水库及上游水环境污染形成、输送机理及其环境效应》.
2. 东北石油局委托项目《开发吉林新星液碳有限公司企业管理信息系统》(项目编号 01220101000269).
3. 国家统计局计算中心项目《国家普查数据处理管理系统的抽样调查设计》(项目编号: LX03-Y20).

致 谢

本论文是在导师杨天行教授的悉心指导下完成的，正是在他多年来的严格要求下，我才得以顺利完成博士论文。杨老师学识渊博、治学严谨，而且待人诚恳，诲人不倦。从导师那里，我不仅学到了丰富的专业知识，增强了科研能力，开拓了创新精神，更领悟到了许多治学的态度和做人的道理，老师的言传身教将使我终身受益。由衷感谢杨老师的培养和无私教诲！

在此，向我的导师表示最真挚的谢意并致以崇高的敬意！同时对师母陈玉珍女士多年来在生活上给予的关怀表示衷心的感谢！

向在工作中给予关照、学习上给予支持的吉林大学数学科学院数学教学与研究中心、地球探测科学与技术学院的各位领导、老师，特别是朝阳校区数学教研室的各位领导、老师表示衷心的感谢！

向一起学习的师姐妹师兄弟：刘金英、黄继国、马瑞杰、聂晶、谷照升、计有权、刘英敏、刘本华等博士表示感谢！他们的学习精神和开拓创新的能力使我获益匪浅。

感谢北京石油化工学院的领导和同事给予我学习和工作上的大力支持。

感谢文中引用过文献的所有作者们，他们的科学态度让我感受到最完美的学习体验。

感谢我的家人多年来的培养、帮助、关心和支持，他们付出的心血是我无论如何也难以补偿的。

最后，向所有其他在我读博期间帮助和支持过我的人表示衷心感谢！

摘 要

论文题目：基于支持向量机理论的水环境质量预测与评价方法研究

专 业：地球探测与信息技术

指导教师：杨天行 教授 吉林大学

支持向量机(Support Vector Machines, SVM)是复杂非线性科学和人工智能科学的研究前沿,其在水资源的水质预测与评价领域的应用研究在国内外尚处于初创阶段。本文在较全面分析评述了水质预测与评价研究现状,在分析阐述了 SVM 的基本原理与算法基础上,在国内首次将 SVM 方法引入水环境质量预测及其质量综合评价中的应用领域,主要在水质预测和水质评价的 SVM 建模方法以及算法方面进行了一些创新性的研究工作,为提高水质评价和水质模拟的智能化水平做出了努力。

水质评价和水质预测问题,实质上是水体分类和水质参数回归的问题。但是传统的统计学研究的是假定样本数目趋于无穷大时的渐近理论,现有的机器学习方法大多是基于这个假设。然而在实际的问题中,样本数目往往是有限的。现有的基于传统统计学的学习方法在有限样本的情况下难以取得理想的效果。统计学习理论是在有限样本情况下新建立起来的统计学理论体系,为人们系统地研究小样本情况下机器学习问题提供了有力的理论基础。

支持向量机是在统计学习理论基础上开发出来的一种新的、非常有效的机器学习新方法。它较好地解决了以往困扰很多学习方法的小样本、非线性、过学习、高维数、局部极小点等实际问题,具有很强的推广能力。目前,统计学

习理论和支持向量机作为小样本学习的最佳理论，开始受到越来越广泛的重视，正在成为人工智能和机器学习领域新的研究热点。

作为一门新兴的学科，有关支持向量机方面的很多研究在国外也是刚刚起步，国内的有关研究只是在近几年才逐渐引起人们的注意。虽然支持向量机方法有比较完善的理论基础，但有关其应用研究还有很多亟待解决的问题，并且随着其应用领域的不断扩大，其理论也有待进一步完善和发展。

本文的主要贡献如下：

1) 研究分析了水环境的水质预测和评价方法的产生背景、发展过程和国内外的研究现状，并从水环境的水质预测与质量评价的客观性、公正性的角度，论述了支持向量理论与方法应用于水质预测与质量评价研究的合理性和可行性。

2) 详细地阐述了统计学习理论研究的基本问题及主要内容。为了更好地说明统计学习理论在实际问题研究中的实现问题，首先回顾了 SVM 的基本概念及基本理论；然后着重介绍了线性与非线性标准支持向量、标准支持向量回归机的基本理论与方法。

3) 详细讨论了两类分类问题的支持向量机算法及其变形算法。通过对这些变形的支持向量机分类算法进行的分析与研究，探讨了各种算法的优缺点和适用范围等情况。通过对这些方法的客观、全面地认识，为构造基于支持向量机分类理论的水环境质量综合评价提供了理论基础。

4) 重点阐述用于曲线拟合的支持向量回归机的基本原理，然后详细地阐述并分析了各种支持向量回归机模型，包括 ε -支持向量回归机、 ν -支持向量回归机、LS-支持向量回归机、W-支持向量回归机、基于线性规划的支持向量回归机。同时，建立了在 ε -不敏感函数下的单参数约束下的支持向量回归模型，证明了单参数回归模型与标准回归模型的等价性，并进行数值实验；最后探讨了支持向量回归算法优劣的评价依据，从而完善了支持向量回归机的优化理论基础和决

策依据。

5) 总结了目前存在的基于支持向量机的多类别分类方法, 包括“一对多”方法、“一对一”方法、超球面多值分类算法、二叉树的支持向量机多值分类等若干种方法, 比较了它们的优缺点及性能; 并通过实验对其中的几种常用的方法进行了验证与比较。

6) 在综合研究 SVR 算法的基础上, 针对水环境中污染物浓度的预测的问题, 将水质参数平均浓度值的监测看成一个时间序列预测问题, 提出了一种新的支持向量回归机的优化算法。并利用支持向量回归机对实验数据进行回归估计分析, 并与 BP 网络方法进行了比较。从预测结果来看, 利用 SVR 方法建立的预测模型, 能充分利用训练样本的分布特性, 预测结果与实际情况更为相符。而且 SVR 在整体性能上优于 BP 网络方法, 有很强的学习能力、泛化能力。

7) 在进行了深入的研究与分析的基础上, 根据水质评价的原则与要求, 结合预抽取支持向量以及循环迭代算法, 提出了一种改进的循环决策二叉树的支持向量机多值分类方法, 成功地克服了其他方法存在的问题及缺点, 提高支持向量机的训练和决策速度, 并且解决了目前支持向量机多值分类方法的不可分区域的分类问题, 取得了较为理想的效果。将它应用于北京市湖泊水库的水质评价, 其研究结果表明: SVM 模型在相当多的评价指标情况下, 仍可很快地给出评价结果, 且可达到相当高的精度, 同时模型可表述定量和定性的评价指标, 增强了评价方法的通用性和适用性。

本文研究的创新点, 主要有以下几个方面:

1. 首次将 SVM 方法引入水环境质量预测及其质量综合评价中的应用领域, 主要在水质预测和水质评价的 SVM 建模方法以及算法方面进行了一些创新性的研究工作, 提高了水质综合评价和水质预测的智能化水平。
2. 建立了支持向量回归算法优劣的评价依据, 从而完善支持向量回归机的优化理论基础和决策依据。

3. 拓展 SVM 的应用范围, 提出改进的 SVM 多值分类模型, 构建一种新的水库湖泊水质质量综合评价方法 (第六章)。
4. 提出改进的 SVR 算法并建立一种新的水质预测模型 (第七章)。

本文将支持向量机理论与算法应用到水质预测与评价还是一个新的尝试, 取得了一些有价值的研究成果。与传统评价方法相比, 该方法所需训练样本少、计算简便、快捷、精度高, 具有较大的实用性, 且评价结果更具有客观性。同时还能根据有限的样本信息在模型的复杂型和学习能力之间寻求最佳折衷, 具有很强的推广能力。

该研究丰富和拓展了支持向量机理论及其应用范围, 不仅丰富了水环境质量预测与评价的理论与方法, 而且对水资源的合理规划和可持续地开发利用提供了科学的决策依据和技术支持, 具有实际应用价值。

关键词: 统计学习理论; 支持向量机(SVM); 支持向量回归机(SVR); 算法;
多值分类; 水环境; 水质预测; 水质评价

Abstract

Title: STUDY ON THE PREDICTION AND ASSESSMENT METHODS OF WATER ENVIRONMENT QUALITY BASED ON SUPPORT VECTOR MACHINES THEORY

Specialty: Geological exploration Science and Technology

Supervisor: Professor Yang Tianxing Jilin University

Innovative work is done on modeling and algorithm of Support Vector Machine (SVM) for the water environment quality prediction and quality evaluation in this dissertation. SVM plays a leading role in the sciences for complex non-linear phenomena and artificial intelligence. Researches on its application in the planning of the water environment quality prediction and quality evaluation are still in the preliminary stage in the world. On the basis of a comprehensive evaluation and analysis of the present situation of the researches in the water environment quality prediction and quality evaluation, and on the basis of a careful exposition of the basic principles, the algorithm and the varied pattern features of SVM, this dissertation gives an application of SVM approaches in the water environment quality prediction and quality evaluation, which, as the first attempt of its kind, can help to achieve a higher level in the application of artificial intelligence in this field.

In fact, the water environment quality prediction and quality evaluation is subject of classification and regression. Traditional statistics is based on assumption that samples are infinite, so are most of current machines learning methods. However, in many practical cases, samples are limited. Most of existing methods based on traditional statistical theory may not work well for the situation of limited samples. Statistical Learning Theory (SLT) is a new statistical theory framework established from finite samples. SLT provides a powerful theory fundament to solve machine

learning problems with small samples.

Support Vector Machine (SVM) is a novel powerful machine learning method developed in the framework of SLT. SVM solves practical problems such as small samples, nonlinearity, over learning, high dimension and local minima, which exist in most of learning methods, and has high generalization. Currently, being the optimal learning theory for small samples, SLT and SVM is attracting more and more researcher and becoming a new active area in the field of artificial intelligent and machine learning.

As a new subject, the research for SVMs only starts recently abroad. People didn't notice it until recent year in our country. Although SVM has a good system of theories, there are still many questions to be answered about its applications. Also there is much space to perfect and develop its theories.

The main contributions are following:

- 1) Introduce the background, development and the current research in water quality supervision and evaluation methods for water environment. Discuss the feasibility and innovation of the application of SVM in the water quality prediction and evaluation research from the angle of objectivity and fairness.

- 2) Elaborate the essential problems and main contents of statistical learning theory. To explain how to apply the statistical learning theory in practical problems, firstly introduce the basic concepts and basic theories, and then focus on the basic theories and methods about linear and nonlinear standard support vector and also the standard support vector regression machine.

- 3) Discuss two support vector machine algorithms and their variants in classification. The analysis and study on these algorithms shows the advantages, disadvantages and applicable range of each algorithm, which makes theoretical preparation for using SVM classification theory in water environment quality evaluation.

- 4) Focus on the introduction of the fundamental principles of SVR for curve

fitting, and then analyze several SVR models, including ε -SVM, ν -SVM, LS -SVM, W -SVM and SVM based on linear program. After that establish the SVM model constrained by single parameter of ε – non-sensitive function, which verifies the equivalence of single parameter regression model and standard regression model. Some numerical experiments are conducted. Finally discuss the criteria to estimate the goodness of SVM algorithm, which perfects the optimum theory and decision-making of SVR.

5) Summarize the existing multi-class classification methods, including one to many method, one to one method, hypersphere multi-classification method etc. Compare their advantages, disadvantages and performance and prove by experiment. A fuzzy iteration SVM multi-value classification algorithm based on decision binary tree is given combined with pre-draw SVM algorithm. The proposed method overcomes the problems of other existing methods, enhance the training and decision-making speed. Also it solves the problem of undivisible region for SVM multi-class classification method with satisfying results.

6) In this paper, a new SVR algorithm is put forward to solve the prediction problem of contamination concentration regarding the supervision of water quality parameter average concentration as a time series prediction question. The supervision data of pollution material concentration is performed regression estimate analysis using SVR. Compared with the result of BP network method, the prediction model established by SVR method can make full use of the distribution features and the prediction result fits the practical condition better. Moreover, SVR is better than BP network in integrating performance with great learning ability and generalization capacity.

7) According to the principles and requirement of water evaluation, a new method of iteration SVM multi-value classification algorithm based on decision binary tree is presented. Several simulations demonstrate that compared with the existing methods, the number of SVMs need to be trained is less by using the new

method, the speed of training and decision is fast and the region that can not be classified does not exist again. It is used to set up a water quality evaluation model for Beijing lake reservoir. The analysis result shows that with numerous assessment indexes, SVM model can quickly produce assessment results with a high accuracy, and, as it can work with both qualitative and quantitative indexes, has wider areas of application.

The innovated points are mainly listed as follows:

1. This dissertation gives an application of SVM approaches in the water environment quality prediction and quality evaluation. As the first attempt of its kind, can help to achieve a higher level in the application of artificial intelligence in this field.
2. Discuss the criteria to estimate the goodness of SVM algorithm, which perfects the optimum theory and decision-making of SVR.
3. Extend the applicable range of SVM, a new and effect method of iteration SVM multi-value classification algorithm based on decision binary tree is presented to the water environment quality evaluation (CH 6).
4. A new effective method is put forward to evaluate the water environment quality prediction based on the study of SVR algorithm (CH 7).

It is a new trial to apply SVM theory and algorithm in the water quality prediction and evaluation, and achive some satisfying results. This method is small sample needed, simple in computation, fast, accurate, practical and and coincidence with the objective features of the research objects interest. It can also find the optimum compromise between the model complexity and learning ability according to the limited sample information such that it has high generalization. This paper will extend the applicable range of SVM.

SVM technique will become mature as the development in the research method and hardware condition. SVM will be a very important research topic in every

research field due to its special characters and will stimulate great progress in scientific technique.

Key words: statistical learning theory; support vector machine; support vector regression machine; algorithm; multi-value classification; water environment; water quality prediction; water quality evaluation.