

国防科技大学学术著作专项经费资助出版

基于核方法的雷达高分辨 距离像目标识别理论与方法

肖怀铁 冯国瑜 郭雷 等编著



国防工业出版社
National Defense Industry Press

国防科技大学学术著作专项经费资助出版

基于核方法的雷达 高分辨距离像 目标识别理论与方法

肖怀铁 冯国瑜 郭雷
丛瑜 孙发圣 于新



国防工业出版社

·北京·

内 容 简 介

本书是作者多年来基于核方法的雷达高分辨距离像目标识别的研究成果的总结。全书共分10章。第1章简要概述了核方法的基本理论。第2章介绍了SVM的可分性问题以及非均衡数据目标识别SVM模型多参数优化选择。第3章介绍了基于核判别分析方法的雷达高分辨距离像识别。第4章介绍了基于核聚类的雷达高分辨距离像识别。第5章介绍了SVM多目标分类识别问题。第6章介绍了基于单空间SVDD的雷达高分辨距离像识别。第7章介绍了基于自适应SVDD的雷达高分辨距离像识别。第8章介绍了基于双空间SVDD的雷达高分辨距离像识别。第9章介绍了基于ISVDD的雷达高分辨距离像在线识别。第10章介绍了特征空间数据核矩阵收缩方法。

本书主要读者对象为信号与信息处理、人工智能与模式识别及相关专业的高年级大学生、研究生和教师、科研人员和工程技术人员。

图书在版编目(CIP)数据

基于核方法的雷达高分辨距离像目标识别理论与方法/肖怀铁等编著. —北京:国防工业出版社,2015.9
ISBN 978-7-118-10520-9

I. ①基... II. ①肖... III. ①雷达目标识别—研究
IV. ①TN959.1

中国版本图书馆CIP数据核字(2015)第234916号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路23号 邮政编码100048)

三河市众誉天成印务有限公司印刷

新华书店经售

*

开本 710×1000 1/16 印张 17 字数 322 千字

2015年9月第1版第1次印刷 印数 1—2000册 定价 55.00元

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

前 言

雷达目标识别的任务是从雷达接收到的目标散射回波中提取目标特征信息,并利用一定的判决规则,确定目标类别和属性。当雷达以足够宽的宽带信号来照射目标,可实现高的距离分辨率,获得反映目标散射中心沿距离方向分布的高分辨距离像。高分辨距离像包含了目标各散射中心间径向位置关系和目标精细结构特征等丰富的目标特征信息,可用于目标的特性分析和识别。高分辨距离像识别在雷达目标识别领域一直受到广泛的关注。距离像具有方位、平移、幅度敏感性等特点,对于雷达非合作目标识别,还存在样本非均衡的问题,因此,基于高分辨距离像的雷达目标识别是一个复杂的非线性分类问题,如何基于有限的构造最优分类器是识别的重点和难点之一。

核方法是近年来机器学习领域内的研究热点,作为一类可有效处理非线性问题的方法,已经被广泛应用于模式识别、信号处理、故障和图像检测等领域。核方法的基本思想是基于 Mercer 核展开定理,通过非线性映射,把样本由原始空间映射到一个高维乃至无穷维的特征空间(这个空间称为 Hilbert 空间)。Cover 定理表明,通过这种非线性变换将样本映射到足够高维的特征空间,输入空间处于非线性可分状态的样本将以极大的可能性变为线性可分。核方法通过引入核函数,使原始样本空间中的高度非线性分类、特征分析等问题,能够在高维特征空间中使用线性学习的方法解决,在处理非线性问题中具有高效性。

核方法被逐步应用于雷达高分辨距离像目标识别,在高维特征空间内对目标的相关数据进行特征提取和分类器设计,表现出了强大的优势。核方法的应用主要分为两种:一种是分类方法,如支持向量机(SVM)、支持向量数据描述(SVDD);另一种是特征提取方法,如核主分量分析、核 Fisher 判别分析等。因此,基于核方法的高分辨距离像识别主要包括两个重要内容,一是核特征提取,二是核分类器构建。

全书共分 10 章。第 1 章概述了核方法的基本理论。简要介绍了核映射方法基本概念、支持向量机和支向量数据描述分类原理。第 2 章研究了 SVM 的可分性问题以及非均衡数据目标识别 SVM 模型多参数优化选择。第 3 章研究了基于核判别分析方法的雷达高分辨距离像识别,主要有基于核主分量分析、核直接判别分析和核局部均值判别分析的特征提取和识别研究。第 4 章总结了基于核聚类的雷达高分辨距离像识别,首先研究了基于核 C - 均值聚类的高分辨

距离像识别,进一步研究了基于模糊核 C - 均值聚类的高分辨距离像识别。第 5 章研究了 SVM 多目标分类识别问题。总结了基于 K 最近邻的 SVM 快速训练算法、基于 K 最近邻的快速 SVM 增量学习算法以及基于核分级聚类的 SVM 多类分类算法。第 6 章研究了基于单空间 SVDD 的雷达高分辨距离像识别。介绍了 SVDD 超球空间分布特性和拒判域问题、SVDD 序贯最小相对距离多类目标识别以及基于 SVDD 的多目标模糊识别方法。第 7 章针对目标类样本充分、非目标类样本多样化的目标分类问题,建立了自适应超球半径模型,研究了基于自适应 SVDD 的雷达高分辨距离像识别。第 8 章将 SVDD 超球空间划分为内空间和延拓空间,使用隶属度模型、云模型、高斯混合模型分别对延拓空间的样本分布建模,研究了基于双空间 SVDD 的雷达高分辨距离像识别。第 9 章针对小规模训练样本下的在线识别问题,介绍了增量 SVDD 算法流程和适于在线学习的增量支持向量数据描述算法,研究了基于 ISVDD 的雷达高分辨距离像在线识别。第 10 章研究了特征空间数据核矩阵收缩方法。在特征空间对数据进行收缩,构造分类性能更优的核矩阵。

本书可作为信号与信息处理、人工智能与模式识别及相关专业的研究生和教师、特别是从事雷达信号处理和目标识别的科研人员和工程技术人员参考。

本书由肖怀铁教授、冯国瑜博士和郭雷博士执笔,单凯晶、丛瑜、孙发圣和于新硕士参与了部分内容的撰写。肖怀铁教授拟定全书内容并负责全书的统稿。

本书总结了 we 多年来基于核方法的雷达高分辨距离像目标识别的研究成果,这些研究工作得到了国家自然科学基金项目的资助。感谢国防科技大学学术著作出版资助专项经费对本书出版的资助。

由于时间仓促,水平有限,经验不足,有不当之处,希望读者批评指正。

作者

目 录

第 1 章 核方法基本理论	1
1.1 引言	1
1.2 核方法的基本概念	1
1.3 支持向量机(SVM)	3
1.3.1 最优分类超平面	3
1.3.2 构造最优超平面	4
1.3.3 广义最优分类面	5
1.3.4 高维空间中的最优分类面	6
1.3.5 构造 SVM	6
1.4 支持向量数据描述(SVDD)	8
参考文献	10
第 2 章 SVM 可分性与模型多参数优化选择	12
2.1 引言	12
2.2 SVM 可分性研究	12
2.2.1 线性可分的定义	13
2.2.2 SVM 线性可分充要条件	13
2.2.3 SVM 线性可分性的度量	19
2.2.4 惩罚因子 C 对分类性能的影响	23
2.3 SVM 模型多参数优化选择	25
2.3.1 SVM 模型单参数最优选择问题	26
2.3.2 非均衡数据目标识别 SVM 模型参数优化选择方法	30
2.3.3 实验结果与分析	34
参考文献	40
第 3 章 基于核判别分析的雷达高分辨距离像识别	42
3.1 引言	42
3.2 基于 KPCA 的特征提取和识别	43
3.2.1 主分量分析方法(PCA)	43
3.2.2 核主分量分析方法(KPCA)	44
3.2.3 KPCA 与 PCA 的比较	45

3.2.4	算法实现	47
3.2.5	实验结果与分析	48
3.3	基于 KDDA 的特征提取和识别	53
3.3.1	线性判别分析(LDA)方法	53
3.3.2	直接判别分析(D-LDA)方法	54
3.3.3	核直接判别分析(KDDA)方法	56
3.3.4	基于 KDDA 的特征提取和识别算法	59
3.3.5	实验结果与分析	59
3.4	基于核局部均值判别分析的特征提取和识别	61
3.4.1	核 Fisher 判别分析(KFDA)	61
3.4.2	局部均值判别分析(LMDA)	63
3.4.3	核局部均值判别分析(KLMDA)	64
3.4.4	KLMDA 与 KFDA、KDDA、CKFD 的性能分析	69
3.4.5	实验结果与分析	70
	参考文献	80
第 4 章	基于核聚类的雷达高分辨距离像识别	83
4.1	引言	83
4.2	基于核 C 均值聚类的 HRRP 识别	84
4.2.1	C 均值聚类算法	84
4.2.2	核 C 均值聚类算法	84
4.2.3	核 C 均值聚类算法的核参数优化选取	86
4.2.4	基于核 C 均值聚类的高分辨距离像识别方法	92
4.2.5	实验结果与分析	93
4.3	基于模糊核 C 均值聚类的 HRRP 识别	96
4.3.1	模糊 C 均值聚类算法	96
4.3.2	模糊核 C 均值聚类算法	97
4.3.3	自适应模糊核 C 均值聚类算法	98
4.3.4	基于模糊核 C 均值聚类的高分辨距离像识别方法	101
4.3.5	实验结果与分析	102
	参考文献	105
第 5 章	基于 SVM 的多目标分类识别	107
5.1	引言	107
5.2	基于 K 最近邻的 SVM 快速训练算法	108
5.2.1	KNN-SVM 算法的基本思路	108
5.2.2	KNN-SVM 算法的基本步骤	110
5.2.3	实验结果与分析	111

5.3	基于 K 最近邻的快速 SVM 增量学习算法	116
5.3.1	KNN-ISVM 算法的基本思路	116
5.3.2	边界向量提取方法	117
5.3.3	KNN-ISVM 算法的主要步骤	119
5.3.4	实验结果与分析	119
5.4	基于核分级聚类的 SVM 多类分类算法	125
5.4.1	分级聚类算法的改进	125
5.4.2	核分级聚类算法	129
5.4.3	基于核分级聚类的 SVM 多类分类算法	130
5.4.4	实验结果与分析	131
	参考文献	141
第 6 章	基于单空间 SVDD 的雷达高分辨距离像识别	143
6.1	引言	143
6.2	SVDD 超球空间分布特性和拒判域	144
6.2.1	HRRP 在 SVDD 超球空间的分布特性	144
6.2.2	SVDD 多目标识别中的拒判问题	147
6.3	SVDD 序贯最小相对距离多类目标识别	148
6.3.1	SVDD 多目标识别的拒判域处理	148
6.3.2	SVDD 多目标识别训练算法	149
6.3.3	最小相对距离 SVDD 多目标识别算法	150
6.3.4	序贯最小相对距离 SVDD 多类目标识别算法	150
6.3.5	实验结果与分析	151
6.4	基于 SVDD 的雷达多目标模糊识别方法	157
6.4.1	基于 SVDD 的雷达多目标模糊识别方法	157
6.4.2	实验结果与分析	159
	参考文献	162
第 7 章	基于自适应 SVDD 的雷达高分辨距离像识别	164
7.1	引言	164
7.2	SVDD 模型参数的影响	164
7.2.1	核参数的影响	164
7.2.2	惩罚因子 C 的影响	165
7.3	基于自适应 SVDD 的雷达高分辨距离像识别	167
7.3.1	二次训练	167
7.3.2	最优超球半径选择	168
7.3.3	基于常规 SVDD 的 HRRP 目标识别流程	171
7.3.4	噪声分析	172

7.3.5	自适应超球半径模型	175
7.3.6	基于自适应 SVDD 的 HRRP 目标识别流程	177
7.4	实验结果与分析	178
7.4.1	高斯白噪声情况	179
7.4.2	海杂波	182
	参考文献	184
第 8 章	基于双空间 SVDD 的雷达高分辨距离像识别	186
8.1	引言	186
8.2	超球空间 HRRP 的归属特性分析	186
8.3	延拓空间的样本分布建模	187
8.3.1	延拓样本分布的隶属度模型	188
8.3.2	延拓样本分布的云模型	189
8.3.3	延拓样本分布的高斯混合模型	191
8.3.4	参数估计	192
8.4	基于双空间 SVDD 的高分辨距离像识别方法	195
8.5	实验结果与分析	197
8.5.1	模型参数选择	198
8.5.2	识别实验结果	199
8.5.3	实验结果分析	201
	参考文献	204
第 9 章	基于 ISVDD 的雷达高分辨距离像在线识别	206
9.1	引言	206
9.2	支持向量数据描述的增量泛化性能分析	207
9.2.1	SVDD 的 KKT 条件	208
9.2.2	SVDD 的训练样本分布特性	208
9.2.3	增量样本的超球空间分布特性	209
9.2.4	实验结果与分析	212
9.3	一种适于在线学习的增量支持向量数据描述算法	215
9.3.1	增量支持向量数据描述	216
9.3.2	ISVDD 性能分析	220
9.3.3	实验结果与分析	221
9.4	基于 ISVDD 的雷达高分辨距离像在线识别方法	228
9.4.1	分方位帧建模	228
9.4.2	全方位 HRRP 的等间隔分帧	229
9.4.3	基于 ISVDD 的 HRRP 在线识别方法	231
9.4.4	实验结果与分析	232

参考文献	240
第 10 章 特征空间数据核矩阵收缩方法	243
10.1 引言	243
10.2 非线性可分与不可分问题	244
10.3 特征空间数据核矩阵收缩方法	245
10.3.1 数据在特征空间的收缩因子	245
10.3.2 数据在特征空间收缩后的核矩阵	247
10.3.3 数据在特征空间收缩方法	249
10.4 实验结果与分析	249
10.4.1 二维数据收缩实验	249
10.4.2 特征空间数据核矩阵收缩实验	250
附录 A 式(3.70)的推导	255
附录 B 式(3.75)的推导	256
附录 C 式(3.90)的推导	260
附录 D 式(3.94)的推导	261
参考文献	262

第 1 章 核方法基本理论

1.1 引言

从 20 世纪 90 年代中期开始,出现了被称为基于核学习方法的模式识别方法,简称核方法(Kernel Method)。核方法的基本思想是基于 Mercer 核展开定理^[1],通过非线性映射,把样本由原始空间映射到一个高维乃至无穷维的特征空间(这个空间称为 Hilbert 空间)。Cover 定理表明^[2],通过这种非线性变换将样本映射到足够高维的特征空间,输入空间处于非线性可分状态的样本将以极大的可能性变为线性可分。通过引入核函数这一概念避免了维数灾难的发生,使原始样本空间中的高度非线性分类、特征分析等问题,能够在高维特征空间中使用线性学习的方法解决。

核方法作为一类可有效处理非线性问题的方法,它从理论上为机器学习提供了一种系统而有原则的学习方式^[3-6],已经应用于特征提取与核分类器设计、信号处理^[7]、统计概率密度估计^[8]、图像分割^[9]等领域。

核方法按其用途主要分为两种:一种是分类方法,如支持向量机(Support Vectors Machine, SVM)^[14]、支持向量数据描述(Support Vector Data Description, SVDD)^[17];另一种是特征提取方法(或称为特征压缩方法),如核主分量分析(Kernel Principal Component Analysis, KPCA)^[18]、核 Fisher 判别分析(Kernel Fisher Discriminant Analysis, KFDA)^[19]等。

SVM 是应用核方法的典型例子,基于核方法的 SVM 分类器用于手写字符的分类,取得了好的效果。SVDD 方法最初成功应用于故障和图像检测领域^[22-25]。广泛使用的核非线性特征提取方法主要包括 KPCA、KFDA、核直接判别分析(Kernel Direct Discriminant Analysis, KDDA)^[21]、核子类判别分析(KS-DA)^[26]、完整核 Fisher 判别分析(Complete Kernel Fisher Discriminant Analysis, CKFD)^[27]、广义判别分析(Generalized Discriminant Analysis, GDA)^[20]、核独立分量分析(Kernel Independent Component Analysis, KICA)方法^[28]、核正交局部保持投影(Kernel Orthogonal Locality Preserving Projection, KOLPP)^[29]、核熵分量分析(Kernel Entropy Component Analysis, KECA)^[30]等。

1.2 核方法的基本概念

如图 1.1 所示,两类原始样本满足多模分布,处于非线性可分状态,难以直

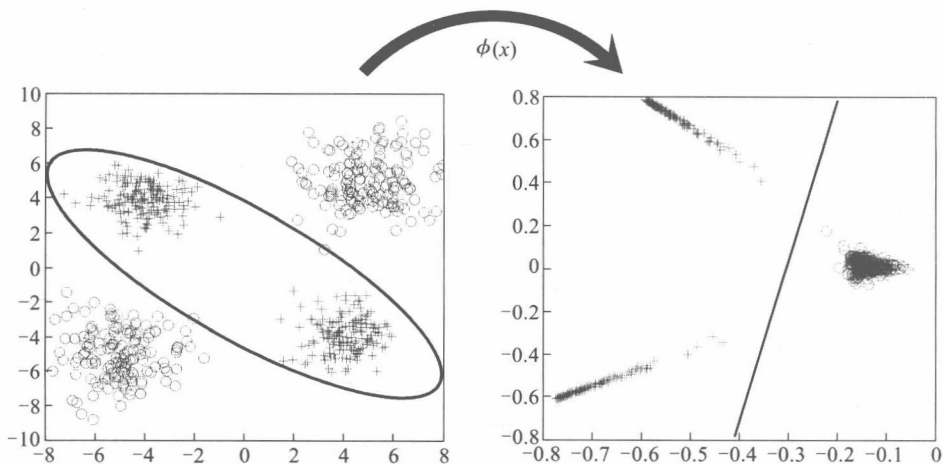


图 1.1 核映射变换前后的数据可分性示意图

接使用一个线性判决面实现两类数据的分离,而原始样本经过一个非线性映射函数变换以后,在核特征空间两类数据之间呈现出线性可分性,可以使用一个线性判决面将两类数据进行分离。

通过图 1.1 的例子可知,核映射将数据从低维空间映射到高维空间,原线性不可分的目标样本在高维空间具有良好的可分性,从而使许多在原始空间线性方法难以解决的问题,在核映射空间中可以用线性方法实现良好的分类。然而核特征空间往往是高维。甚至是无限维的,因此如果直接使用映射样本进行分类器训练,计算复杂度将会迅速增加,随着高维空间维数的不断增长,求解分类面的代价会越来越大,甚至无法求解,即产生维数灾难问题^[10],因此,需要找到一种新的方法既可有效利用核映射空间的高维优势又能避免维数灾难。此时,核函数为我们提供了一种有效的解决方法,它只需要知道样本映射空间中的样本内积,而不需要明确知道映射到高维空间的样本显式表达式(一般也无法获得映射样本的显式表达式),就可以实现高维空间的样本训练,避免了维数灾难问题的产生。

定义 1.1 (核函数)^[1] 设 $\phi: \mathbf{x} \in X \rightarrow \phi(\mathbf{x}) \in F$ 是从原始输入空间 R^D 到某个高维空间 F 的非线性映射,如果对于任意的 $\mathbf{x}_i, \mathbf{x}_j \in R^D$, 函数 k 都满足

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (1.1)$$

则函数 k 称为核函数,其中, $\langle \cdot, \cdot \rangle$ 表示内积运算。

实际上,利用核函数 k ,事先不需要知道映射 ϕ 的具体形式,就可以在原始空间中高效地计算出高维空间中样本的内积,根据 Hilbert - Schmidt 原理^[11],任意函数 k 满足 Mercer 条件都可以用作核函数。

引理 1.1 (Mercer 条件)^[12] 对于任意的对称函数 $k(\mathbf{x}, \mathbf{x}')$,它是某个空间

内积运算的充分必要条件是对于任意满足 $\int g^2(\mathbf{x}) d\mathbf{x} < \infty$ 的函数 $g(\mathbf{x})$, 有

$$\iint k(\mathbf{x}, \mathbf{x}') g(\mathbf{x}) g(\mathbf{x}') d\mathbf{x} d\mathbf{x}' \geq 0 \quad (1.2)$$

对于很多函数来说, Mercer 条件并不难满足, 满足 Mercer 条件的函数称为 Mercer 核(或者再生核), 也称为核函数, 它所定义的空间称为再生核 Hilbert 空间(Reproducing Kernel Hilbert Space, RKHS)^[13]。常用的核函数有高斯核函数 $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \sigma^2)$ 和多项式核函数 $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + b)^q$ 等。

综上所述, 核方法的基本概念可以描述为: 首先定义一个满足 Mercer 条件的核函数, 将原始样本通过非线性映射从低维原始空间变换到高维(甚至无限维)空间, 然后在高维空间中进行线性运算, 这种运算可以是分类器模型训练, 也可以是特征映射空间构造等。由于映射样本的内积运算可以通过核函数直接计算, 因此, 对于任意一种算法, 可以用核函数来代替内积, 进而将在原始空间通过线性运算实现的方法推广至高维空间实现, 从而形成核方法来描述目标样本之间的非线性关系。

1.3 支持向量机(SVM)

SVM 是 Vapnik 等人根据统计学习理论提出的一种新的学习方法^[14,15], 它建立在 VC 维理论和结构风险最小原理基础上, 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, 以期望获得最好的推广能力。

1.3.1 最优分类超平面

与线性决策规则(分类超平面)不同, SVM 采用一种特殊的超平面, 即所谓的最优分类超平面(Optimal Hyperplane)。

首先考虑训练数据是线性可分情况下的最优分类超平面, 然后推广到线性不可分的情况。

假定训练数据集为

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N), x \in \mathbb{R}^d, y \in \{+1, -1\} \quad (1.3)$$

式中: N 为训练样本的个数; y 为样本类别标号; $+1$ 表示样本属于 $+1$ 类(即正类), -1 表示样本属于 -1 类(即负类)。

d 维空间中线性判别函数的一般形式为 $g(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$, 分类面方程为

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (1.4)$$

将判别函数进行归一化, 使两类所有样本都满足 $|g(\mathbf{x})| \geq 1$, 使离分类面最近的样本的 $|g(\mathbf{x})| = 1$, 这样分类间隔就等于 $\frac{2}{\|\mathbf{w}\|}$, 因此使间隔最大等价于使

$\|w\|$ (或者 $\|w\|^2$) 最小; 而要求分类面对所有样本正确分类, 就是要求它满足

$$y_i [w \cdot x_i + b] - 1 \geq 0, i = 1, 2, \dots, N \quad (1.5)$$

因此, 满足上述条件且使 $\|w\|^2$ 最小的分类面就是最优分类面。最小化是关于向量 w 和标量 b 同时进行的。最优分类面如图 1.2 所示。

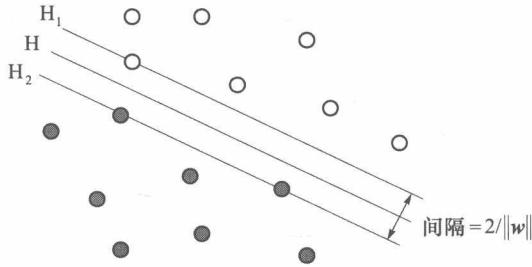


图 1.2 最优分类面示意图

1.3.2 构造最优超平面

要构造最优超平面, 必须用系数模最小的超平面把属于两个不同类别的样本向量分开。要找到这个超平面, 需要求解下面的二次规划问题, 即最小化泛函

$$\Phi(w) = \frac{1}{2} \langle w, w \rangle \quad (1.6)$$

式中: 符号 $\langle w, w \rangle$ 表示向量 w 的内积。由于 w 为实数向量, 根据实内积空间内积的定义可知: $\langle w, w \rangle$ 等于向量 w 的点积, 向量点积一般用 $(w \cdot w)$ 表示。以下直接用 $(w \cdot w)$ 表示向量 w 的内积。约束条件为不等式类型

$$y_i [w \cdot x_i + b] - 1 \geq 0, i = 1, 2, \dots, N \quad (1.7)$$

这个优化问题的解由下面的拉格朗日泛函的鞍点给出:

$$L(w, b, \alpha) = \frac{1}{2} (w \cdot w) - \sum_{i=1}^m \alpha_i [y_i \cdot (w \cdot x_i + b) - 1] \quad (1.8)$$

式中: $\alpha_i \geq 0$ 为拉格朗日乘子。

在鞍点上, 解 w, b 和 α 必须满足以下条件:

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \quad (1.9)$$

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \quad (1.10)$$

由式(1.9)和式(1.10)可以得到

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (1.11)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (1.12)$$

把式(1.11)、式(1.12)代入式(1.8),得到在约束条件 $\sum_{i=1}^N \alpha_i y_i = 0$ 和 $\alpha_i \geq 0$ 下对 α_i 求解下面泛函的最大值:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (1.13)$$

若 α_i^* 为最优解,则

$$\mathbf{w}^* = \sum_{i=1}^N \alpha_i^* y_i x_i \quad (1.14)$$

即最优分类面的权系数向量 \mathbf{w} 是训练样本向量的线性组合。

式(1.13)是一个不等式约束下的二次函数极值问题,存在唯一解。且根据 Karush - Kuhn - Tucker 条件,这个优化问题的解必须满足

$$\alpha_i [y_i (\mathbf{w} \cdot \mathbf{x}_i + b)] - 1 = 0, i = 1, 2, \dots, N \quad (1.15)$$

因此在所得的解中,只有一部分(通常是少部分) α_i 不为零,记为 α_i^* ,即 $\alpha_i^* \neq 0$,它们对应于使式(1.7)等号成立的样本,也就是支持向量(Support Vector, SV),通常只是所有样本的极少一部分,支持向量的个数用 N_s 来表示。

利用这一点可以求出分类阈值 b^* 。设任意一个支持向量 (\mathbf{x}_i, y_i) 使等式成立,则

$$b^* = \frac{1}{2} [\mathbf{w} \cdot \mathbf{x}^*(1) + \mathbf{w} \cdot \mathbf{x}^*(-1)] \quad (1.16)$$

式中: $\mathbf{x}^*(1)$ 表示属于第一类的某个(任意一个)支持向量; $\mathbf{x}^*(-1)$ 表示属于第二类的的一个支持向量。

求解上述问题后,就可以得到最优分类函数是

$$f(x) = \text{sgn} \{ (\mathbf{w}^* \cdot \mathbf{x}) + b^* \} = \text{sgn} \left\{ \sum_{i=1}^{N_s} \alpha_i^* y_i (x_i \cdot \mathbf{x}) + b^* \right\} \quad (1.17)$$

式中: $\text{sgn}(\cdot)$ 为符号函数。

1.3.3 广义最优分类面

最优分类面是在线性可分的前提下讨论的,在线性不可分的情况下,就是某些训练样本不能满足式(1.7)的条件,因此可以在条件中增加一个松弛项 $\xi_i \geq 0$,就变成

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i \geq 0 \quad (1.18)$$

对于足够小的 $\sigma > 0$,只要使下式最小就可以使错分样本数最小,对应线性

可分情况下的使分类间隔最大:

$$F_{\sigma}(\xi) = \sum_{i=1}^N \xi_i^{\sigma} \quad (1.19)$$

在线性不可分情况下引入约束条件

$$\|\mathbf{w}\|^2 \leq c_k \quad (1.20)$$

在约束条件式(1.18)、式(1.20)下对式(1.19)求最小,就得到了线性不可分情况下的最优分类面,称作广义最优分类面。为了计算方便,取 $\sigma = 1$ 。

为了使计算进一步简化,广义最优分类面问题可以进一步演化为在条件式(1.18)、式(1.20)下求下列函数的最小值:

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^N \xi_i \quad (1.21)$$

式中: C 为某个指定的正常数,它实际上控制错分样本的惩罚程度,实现在错分样本的比例与算法复杂度之间的折中。

1.3.4 高维空间中的最优分类面

由上面的讨论得知,分类器最终的分类判别函数式(1.17)中只包含待分类样本与训练样本中的支持向量的内积运算($x \cdot x_i$),同样它的求解过程也只涉及训练样本之间的内积运算($x_i \cdot x_j$),可见,要解决一个特征空间中的最优线性分类问题,只需要知道这个空间的内积运算即可。

考虑 Hilbert 空间中内积的一个一般表达式

$$(z_i \cdot z) = K(x, x_i) \quad (1.22)$$

式中: z 为输入空间中的向量在特征空间中的像。

根据 Hilbert - Schmidt 理论, $K(x, x_i)$ 可以是满足引理 1.1 一般条件的任意对称函数。

1.3.5 构造 SVM

如果用内积 $K(x, x_i)$ 代替最优分类面中的点积,就相当于把原输入空间变换到了某一新的高维特征空间,在这个空间中构造线性分类超平面,如图 1.3 所示。

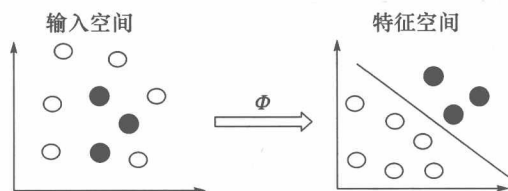


图 1.3 原始数据非线性映射到特征空间

此时的判别函数变为

$$f(x) = \text{sgn}\{(\mathbf{w}^* \cdot \mathbf{x}) + b^*\} = \text{sgn}\left\{\sum_{i=1}^{N_s} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^*\right\} \quad (1.23)$$

在可分的情况下(不可分的情况类似),要求得系数 α^* , 只要寻找泛函

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (1.24)$$

的最大值,约束条件为 $\sum_{i=1}^N \alpha_i y_i = 0$ 和 $\alpha_i \geq 0$ 。

构造式(1.23)类型决策函数的学习机器叫做支持向量机(SVM)。SVM 的复杂程度取决于支持向量的数目,而不是输入数据空间的维数。图 1.4 给出了 SVM 结构的图解。

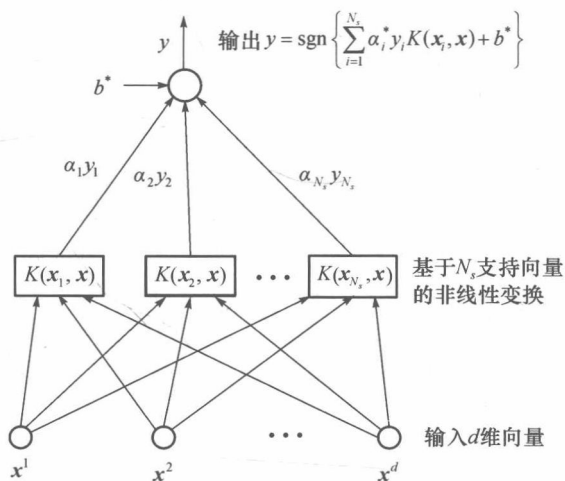


图 1.4 SVM 示意图

采用不同的函数作为内积函数 $K(\mathbf{x}, \mathbf{x}_i)$, 可以构造实现输入空间中不同类型的非线性决策面的学习机器, 目前主要研究的内积函数形式有三类:

(1) 多项式核函数, 即

$$K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^q, q > 0 \quad (1.25)$$

(2) 高斯核函数, 即

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma^2}\right\} \quad (1.26)$$

式中: $\|\cdot\|$ 表示内积空间的范数, 选择高斯核函数的 SVM 网络结构与径向基(RBF)网络结构相同。但是, 对于 RBF 网络, 高维空间的映射先完成, 然后估算 RBF 函数的中心, 而 SVM 方法中, 节点和中心的数量是优化处理的结果^[16]。