



李君宝 乔家庆 尹洪涛 刘大同 著

模式识别中的 核自适应学习及应用



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

模式识别中的核自适应 学习及应用

李君宝 乔家庆 尹洪涛 刘大同 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书以模式识别领域的重要前沿课题——核学习为研究对象,介绍了核自适应学习及在人脸识别、医学图像分类和三维碎片分类等方面的应用,主要包括基本原理、数学基础、参数选择、递归分析方法、函数构造方法、判别分析方法、主成分分析及核自适应学习机的典型应用。

本书可供计算机专业的本科生、研究生参考阅读,旨在帮助读者透彻理解和掌握模式识别中的核自适应学习基本原理和方法,并初步了解核自适应学习在人脸识别、医学图像分类及三维目标碎片分类中的应用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

模式识别中的核自适应学习及应用 / 李君宝等著. —北京: 电子工业出版社, 2013.12

ISBN 978-7-121-21331-1

I. ①模… II. ①李… III. ①核函数—应用—机器学习—研究 ②模式识别—应用—机器学习—研究 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2013) 第 200334 号

责任编辑: 董亚峰 特约编辑: 王 纲

印 刷: 北京市海淀区四季青印刷厂

装 订: 北京市海淀区四季青印刷厂

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 720×1 000 1/16 印张: 13.5 字数: 345 千字

印 次: 2013 年 12 月第 1 次印刷

定 价: 38.00 元



凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

模式识别是计算机、电子及信息领域的重要研究领域，也是人工智能的基础技术，21 世纪是智能化、信息化、计算化、网络化的世纪，作为人工智能技术基础学科的模式识别技术，必将获得巨大的发展空间。在国际上，各大权威研究机构，各大公司都纷纷开始将模式识别技术作为公司的战略研发重点。

近年提出的核学习是模式识别的重要研究分支和活跃的研究主题，出现了大量理论和实际应用研究成果，已广泛用于模式识别、数据挖掘、计算机视觉、图像与信号处理等研究领域。核方法在一定程度上解决了实际应用中的非线性问题，大大提高了实际系统的识别正确率、预测精度等性能指标。然而，核学习方法仍然面临着一个重要问题，即核函数及其参数的选择问题。研究表明，核函数及其参数直接影响非线性特征空间内数据分布结构，不恰当的核函数及其参数选择直接影响核学习器的性能。研究核函数及其参数的自适应学习对于解决目前核学习方法广泛面临的核选择问题具有重要理论研究意义，对于提高核学习器的应用系统性能具有重要意义。因此，核自适应学习是一个兼具理论价值和实际意义的研究主题。

近年来，作者在模式识别中的核自适应学习方法及应用方面开展了不懈的探索研究，本书综合展现了作者在该领域研究过程中的最新研究成果。本书以模式识别中的核自适应学习及应用为研究对象，结合作者自己的研究成果和文献调研，介绍了核自适应学习及在人脸识别、医学图像分类和三维碎片分类等方面的应用。旨在帮助读者透彻理解和掌握核自适应学习的基本原理和方法，并了解核自适应学习在图像分类及三维数据分类上的应用，为深入研究模式识别领域的核学习问题提供技术支撑。

在研究过程中先后得到了国家自然科学基金（编号 61001165、61371178）、黑龙江省自然基金（编号 QC2010066）、哈工大基础人才培养项目（编号 HIT.BRETH.201206）、教育部新世纪优秀人才支持项目以及总装预研基金等 10 项国家和省部级科研项目的资助，再次表示衷心的感谢。

由于核自适应学习领域的研究发展迅速，加之作者水平有限，错漏之处在所难免，恳请读者批评指正。

作 者

2013 年 8 月 于哈尔滨工业大学

目 录

第 1 章 绪论	1
1.1 机器学习概念	1
1.1.1 学习的定义	1
1.1.2 学习问题的一般描述	1
1.1.3 学习的实现	3
1.1.4 学习的基本形式	4
1.1.5 学习在数据降维上的应用	5
1.2 机器学习中的核学习	7
1.2.1 线性特征提取算法及存在的问题	7
1.2.2 核的引入	9
1.2.3 主要核学习算法	11
1.3 核学习的研究现状	15
1.4 核学习存在的问题	19
第 2 章 核学习的数学基础	21
2.1 核理论基础	21
2.1.1 再生核理论	23
2.1.2 Mercer 定理	24
2.2 多项式空间和多项式核函数	25
2.2.1 有序齐次多项式空间	25
2.2.2 有序多项式空间	27
2.2.3 无序多项式空间	27
2.3 Mercer 核	28
2.3.1 半正定矩阵的特征展开	28
2.3.2 半正定积分算子的特征展开	30
2.4 正定核	31
2.5 核函数的构造	34

第 3 章 自适应多核学习	37
3.1 多核学习	37
3.1.1 基于多核学习的特征提取方法的应用	38
3.1.2 存在的问题	38
3.2 基于图嵌入的特征提取原理	40
3.2.1 直接图嵌入	41
3.2.2 直接图嵌入的核扩展形式	42
3.3 多核学习原理	43
3.3.1 核函数定义及性质	43
3.3.2 多核函数构造原理	45
3.4 基于多核的图嵌入特征提取算法原理	47
3.4.1 多核矩阵的构造	47
3.4.2 图嵌入方法的多核扩展	48
3.5 基于多核映射的图像识别算法程序设计	50
3.5.1 训练样本预处理和读入定制参数	52
3.5.2 样本训练	53
3.5.3 测试分类	53
3.6 对比图像分类算法程序设计	54
3.7 实验对比与分析	55
3.7.1 ORL 数据库	56
3.7.2 Yalefaces 数据库	57
3.7.3 Iris (UCI) 数据库	58
3.7.4 Image Segmentation (UCI) 数据库	60
3.8 算法效率比较和分析	61
第 4 章 核自适应递归分析	63
4.1 核函数对 Online SVR 算法性能的影响分析	64
4.1.1 SVR 算法基本原理	64
4.1.2 基于增量学习的 Online SVR 算法	68
4.1.3 基于 Online SVR 的在线时间序列预测	71
4.1.4 核函数类型及其参数影响分析	72
4.2 基于核函数组合的 Online SVR 算法	75
4.2.1 基于核函数组合的 Online SVR 在线时间序列预测算法	75
4.2.2 仿真实验和算法评估	77

4.3	基于残差修正的局部 Online SVR 算法	80
4.3.1	离线与在线算法分析	80
4.3.2	基于残差修正的局部 Online SVR 在线时间序列预测算法	81
4.3.3	仿真实验和算法评估	83
4.3.4	两种核函数组合 Online SVR 算法对比分析	86
第 5 章	核函数优化及构造	88
5.1	高斯核函数及核函数优化的意义	88
5.2	数据相关核及其扩展	90
5.3	核函数优化算法	92
5.3.1	基于 Fisher 准则的核函数优化算法	93
5.3.2	基于最大间隔准则的核函数优化算法	95
5.3.3	算法比较与分析	98
5.3.4	仿真实验	99
5.4	基于图像矩阵的高斯核函数及改进	103
5.4.1	基于图像矩阵的高斯核函数	104
5.4.2	基于图像矩阵的数据相关高斯核函数	107
5.4.3	仿真实验	108
第 6 章	核自适应判别分析	110
6.1	核自适应判别分析算法	110
6.1.1	核判别分析算法	111
6.1.2	改进算法	112
6.1.3	仿真实验	114
6.2	无参数核判别分析算法	116
6.2.1	算法框架	116
6.2.2	仿真结果与分析	119
6.3	自适应多核图嵌入判别分析	123
6.3.1	多核图嵌入目标方程求解	123
6.3.2	核函数选择	125
6.3.3	基本核函数参数优化	127
6.3.4	仿真实验与分析	128
第 7 章	核自适应流形学习算法	137
7.1	流形学习	137

7.2 基于核自适应学习的局部判别分析	137
7.2.1 局部保持映射算法	138
7.2.2 监督局部保持映射算法	139
7.2.3 核监督局部保持映射算法	141
7.2.4 核自适应局部保持判别分析	144
7.2.5 实验仿真与结果分析	146
第 8 章 核自适应主成分分析	150
8.1 主成分分析算法	150
8.2 稀疏核主成分分析算法	152
8.3 核自适应稀疏主成分分析算法	155
8.4 仿真实验	160
第 9 章 核自适应学习机应用	165
9.1 三维碎片分类	165
9.1.1 算法	168
9.1.2 仿真实验	170
9.2 乳腺 X 射线图像分类	173
9.2.1 算法步骤	174
9.2.2 仿真实验	175
9.3 人脸识别	178
9.3.1 算法描述	178
9.3.2 仿真实验	184
9.4 基于 Gabor 小波和 CKFD 结合的人脸图像特征提取算法	188
9.4.1 算法描述	188
9.4.2 仿真实验	191
9.5 KPCA 和 PCA 自融合的人脸图像特征提取算法	195
9.5.1 算法描述	196
9.5.2 仿真实验	200
参考文献	204

1.1 机器学习概念

1.1.1 学习的定义

“学习”涉及多个学科和领域，很难对其做一个精确而统一的定义，因此学者们对学习的定义也不同。下面是不同学者从不同的角度给出的定义。从系统的角度，H. A. Simon 认为“如果一个系统能够通过执行某种过程而改进它的性能，这就是学习，学习就是系统的适应性，这意味着这些改进使得系统能够更有效地完成同样的或者类似的工作”；从神经网络的角度，S. Haykin 把学习定义为“学习是一个网络的自由参数通过网络所处环境的刺激而自适应地调整的过程，学习的形式由参数改变的方式来确定”；从模式识别的角度，Duda 等人认为“最广义地讲，任何设计分类器时所用的方法，只要它利用了训练样本的信息，都可以认为运用了学习”。此外，最新的定义是由 V. Vapnik 提出的，他把学习问题看做“利用有限数量的观测来寻找待求的依赖关系的问题”。学习包括许多特殊的问题，但模式识别、回归函数估计和概率密度估计为机器学习的三个基本问题。

1.1.2 学习问题的一般描述

为了更好地理解学习的定义，利用系统模型对其进行描述。根据学习的定义，

学习问题可以看做利用观测数据来寻找待求的依赖关系的问题，一般的学习系统模型如图 1.1 所示。

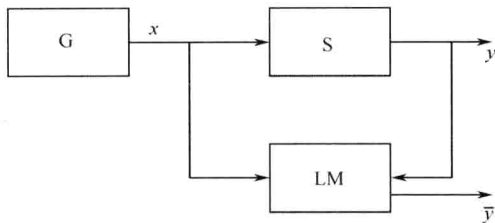


图 1.1 一般的学习系统模型

其中， G 为数据产生器，产生随机样本向量 $x \in R$ ，它们是从固定但未知的概率分布函数独立抽取的。 S 代表训练器，对每个输入向量 x 返回一个输出值 y ，产生输出的根据是条件分布函数 $F(y|x)$ 。 LM 代表学习机器，它能够实现一定的函数集 $f(x, \alpha)$ ， $\alpha \in S$ ，其中 S 是参数集合。学习的目的是从给定的函数集 $f(x, \alpha)$ ， $\alpha \in S$ 中选择出能够最好地逼近训练响应的函数。训练集有根据联合分布函数抽取出的 l 个独立分布观测数据 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ 。因此，机器学习问题可以表示为：根据 l 个独立同分布的观测样本，在一组函数 $\{f(x, \cdot)\}$ 中求一个最优函数 $f(x, \alpha)$ 对训练器的响应进行估计，使期望风险 $R(\alpha) = \int L[y, f(x, \alpha)] dF(x, y)$ 最小，其中， $F(x, y)$ 是未知的， $L[y, f(x, \alpha)]$ 是损失函数，不同类型的机器学习问题有不同形式的损失函数。对学习中的模式识别、回归函数估计和概率密度估计三个基本问题来说，损失函数有不同的定义。

模式识别问题的损失函数为

$$L[y, f(x, \alpha)] = \begin{cases} 0, & y = f(x, \alpha) \\ 1, & y \neq f(x, \alpha) \end{cases} \quad (1.1)$$

其中，训练器的输出 y 是类标识 $y = \{0, 1\}$ 或 $y = \{1, -1\}$ 。

回归函数估计问题的损失函数为

$$L[y, f(x, \alpha)] = [y - f(x, \alpha)]^2 \quad (1.2)$$

其中， $f(x, \alpha) = \int y dF(y|x)$ ，训练器的输出 y 为实数值，并令 $f(x, \alpha)$ ， $\alpha \in S$ 为实函数集和。

概率密度估计问题的损失函数为

$$L[p(x, \alpha)] = -\log p(x, \alpha) \quad (1.3)$$

其中， $p(x, \alpha)$ 为要估计的概率密度函数。

根据损失函数的定义可以看出：

(1) 模式识别学习的目的是得到指示函数(判别函数), 对于这个指示函数, 期望风险就是分类误差的概率。指示函数给出的答案与训练器输出不同, 则称为分类错误。这样学习问题就成为在概率分布函数 $F(x, y)$ 未知、但数据已知的条件下, 寻找一个最小化分类误差的函数。

(2) 回归函数估计学习的目的是在概率分布函数 $F(x, y)$ 未知时, 对给定的采集样本, 最小化风险函数 $R(\alpha) = \int L[y, f(x, \alpha)] dF(x, y)$ 。

(3) 概率密度估计学习的目的是根据独立分布的训练样本确定 x 的概率分布, 在概率分布 $F(x)$ 未知、但给出了独立同分布数据的情况下, 使风险函数最小化。在概率分布 $F(x, y)$ 未知的情况下, 只有样本的信息可以利用, 这导致定义的期望风险是无法直接计算和最小化的。传统的学习方法采用了所谓经验风险最小化(ERM)准则。神经网络和其他经典的统计推理学习算法(例如, 函数回归的最小二乘法、概率密度估计的最大似然法等)都是建立在经验风险最小化归纳原理基础上的。对于分类问题的损失函数, 经验风险就是训练样本错误率; 对于回归估计问题的损失函数, 经验风险最小化准则是通过最小二乘法来实现的; 对于密度估计问题中采用的损失函数, 经验风险最小化准则就等价于最大似然法。

1.1.3 学习的实现

目前学习的实现方法还没有统一的理论框架, 根据目前的研究现状, 大体可分为以下三种。

1. 经典的参数统计估计方法

模式识别、神经网络等机器学习方法的重要理论基础之一是统计学。参数方法正是基于传统统计学的, 参数的相关形式是已知的, 训练样本用来估计参数的值。该方法有很大的局限性, 它需要已知样本分布形式, 这需要花费很大代价。此外, 传统统计学研究是基于样本数目趋于无穷大时的渐近理论进行的, 相关的学习方法也大多基于此假设。但在实际问题中, 样本数是有限的, 一些理论上很优秀的学习方法在实际的应用中却存在很大的局限性。

2. 经验非线性方法

以人工神经网络为例, 该学习方法利用已知样本建立非线性模型, 克服了传统参数估计方法的困难, 然而, 这种方法缺乏一种统一的数学理论。

3. 统计学习理论

与传统统计学相比,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。该理论针对小样本统计问题建立了一套新的理论体系,该体系下的统计推理规则不仅考虑渐近性能要求,而且在现有有限信息条件下寻求得到最优结果。该理论的研究始于 20 世纪 60~70 年代,到 20 世纪 90 年代中期,随着其理论的不断发展和成熟,统计学习理论开始受到越来越广泛的重视。统计学习理论的一个核心概念就是 VC 维概念,它是描述函数集或学习机器的复杂性或者说学习能力的一个重要指标,在此概念基础上发展出了关于统计学习的一致性、收敛速度、推广能力等一系列重要理论。在有限样本情况下,学习精度和推广性似乎是一对不可调和的矛盾,采用复杂的学习机器虽然容易使学习误差更小,却往往丧失推广能力。以分类问题为例,所谓学习推广能力是指训练好的学习机不仅对训练样本有良好的分类能力,对新的数据分类正确率也很高,也就是学习机的实际风险和经验风险都小。如果在训练时学习机表现出良好分类能力,在识别新的数据时正确率却很低,也就是学习机的实际风险很高和经验风险很小,则认为这个学习机的学习推广能力很差。统计学习理论为提高学习的推广能力,在提出 VC 维的基础上,提出了结构风险最小化 (Structure Risk Minimization, SRM) 准则。在模式识别方法中 VC 维的直观定义是: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。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。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,只对一些特殊函数集的 VC 维可以准确知道,如 n 维实数空间中线性分类器和线性实函数的 VC 维是 $1+n$ 。而对于一些比较复杂的学习机器(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的影响,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。

1.1.4 学习的基本形式

根据学习过程中有没有考虑样本类标识信息,学习可以分为监督学习、增强学习和无监督学习三种形式。

(1) 监督学习 (Supervised Learning): 在监督学习中,外界存在一个“教师”,

它可对一组给定输入提供应有的输出结果。

(2) 增强学习 (Reinforcement Learning): 在增强学习中, 一个输入-输出映射的学习是通过与环境连续交互被执行的, 从而使得某个性能的标度值达到最小。在增强学习过程中没有“教师”提供正确答案, 而是一个具有“评论者”的学习。

(3) 无监督学习 (Unsupervised Learning): 在无监督或者自组织学习中没有外部“教师”或者“评论者”去监视学习过程。

1.1.5 学习在数据降维上的应用

基于学习的方法主要是用维数约简方法对原始样本用映射矩阵进行降维, 将降维所得向量作为该样本的特征向量。经典的学习方法是线性映射的学习方法, 线性学习方法的基本思路是找到一个最优映射矩阵, 使得样本在特征空间内可用一个最优分类面将不同类样本线性分开。以两类样本分类为例, 如图 1.2 所示, 对原始空间内混叠的样本, 在线性映射后在特征空间内可以找到一条直线将两类样本分开。该方法对大多数特征提取的应用问题是有效的, 在图像的特征提取中也得到了很好的应用。然而, 对于复杂的图像特征提取问题, 如人脸识别问题, 不同人的人脸图像的灰度分布、轮廓信息相近, 由于姿态、光照和表情变化的影响, 同一个人的几张图像样本也呈现复杂的变化, 这种变化不能简单地描述为线性变化。此时, 以前面所述的两类问题为例, 则无法找到最优的分类面将两类样本分开, 如图 1.3 所示。为解决该问题, 人们引入了非线性映射方法, 基本思路是首先将样本进行非线性映射, 而后利用线性映射方法得到特征向量, 期望此时可以找到最优的分类面将两类样本分开, 如图 1.4 所示。现在的问题是很难确定非线性映射的形式和数学解析式。研究表明, 用核函数代替上述非线性映射内积的方法是一种可行的方法, 该方法在计算过程中无须直接计算样本的非线性映射便可达到非线性特征提取的目的。因此, 核方法被认为解决非线性特征提取问题的有效手段, 该方法具有计算复杂度低、效果好、实用性强等特点, 已广泛应用于非线性特征提取中。

基于学习的方法在图像特征提取上应用广泛, 图像处理的方法大致可以分为基于信号处理的方法和基于学习的方法两类。前者发展较早, 典型的算法有傅里叶变换、Gabor 滤波、小波变换等。后者是近十几年兴起的方法, 该方法是用维数约简方法将数据从原始空间经过线性映射到某一低维空间, 低维空间的数据能够极大地反映原始数据的本质特征, 典型的算法有主成分分析 (PCA)、线性判别分析 (LDA)、局部保持映射 (LPP)、二维主成分分析 (2DPCA)、核主成分分析 (KPCA)

等。基于信号处理的特征提取方法是在变换域内提取特征，强调图像样本的个体信息，主要用于图像的边缘检测、图像分割、图像的降噪等；而对于图像识别来说，主要目的是提取具有最大类区分能力的信息，信号处理的方法不考虑整体图像样本之间在数据空间内存在的统计特性，因此基于信号处理方法的特征提取方法不适合图像的分类。基于学习的方法主要考虑全部样本内在的统计特性，旨在提取具有最大类区分能力的特征，该方法适合图像识别与分类等应用的图像特征提取任务。因此，基于学习的图像特征提取逐渐受到模式识别和图像处理领域的广泛关注。

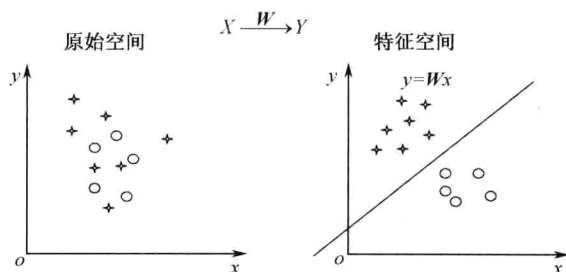


图 1.2 线性可分问题

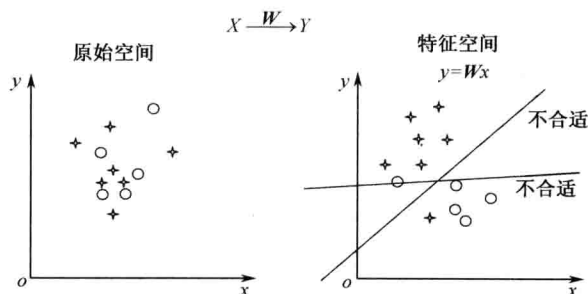


图 1.3 线性不可分问题

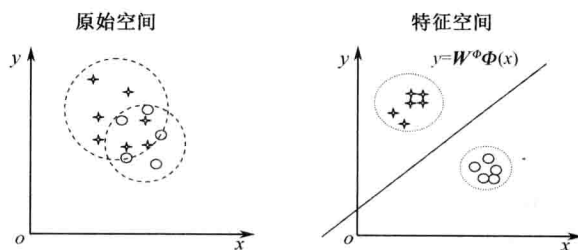


图 1.4 非线性特征提取示意图

1.2 机器学习中的核学习

1.2.1 线性特征提取算法及存在的问题

一般把原始数据组成的空间叫做输入空间，把分类识别赖以进行的空间称为特征空间，把数据从输入空间到特征空间的转换过程称为特征提取。因此，特征提取的完整定义就是使处于原始特征空间的样本数据通过映射（或变换）的方法可以用低维空间来表示的过程。映射后的特征是原始特征空间（也叫输入空间）的某种组合。从广义上讲，所谓的特征提取就是一种变换。若 Y 是输入空间， X 是特征空间，则变换 $A: Y \rightarrow X$ 就称特征提取器。常见的线性特征提取算法的主要思路是利用训练数据在目标函数的约束下进行训练映射矩阵，不同的目标函数也定义了不同的学习方法。不同的特征提取算法采用不同的准则设计目标函数。例如，主成分分析算法（PCA）设计目标函数的思想是使原始样本在特征空间内的特征达到极大不相关；线性判别分析算法（LDA）的目标是在特征空间的同类样本更加靠近，不同类样本远离；而局部保持映射算法（LPP）的目标是在原始空间相邻近的样本在特征空间也互为邻近。用于分类的线性学习器的基本框架如图 1.5 所示。训练样本 $x_1, x_2, \dots, x_N, i = 1, 2, \dots, N$ ，使最优映射矩阵 W 满足于目标函数，用映射矩阵 W 对样本进行映射得到特征向量 $y_1, y_2, \dots, y_N$ ，即 $y_i = Wx_i$ 。对于分类问题，期望在特征空间内 $y_1, y_2, \dots, y_N$ 有很好的类区分能力，可以找到一个最优的分类平面（超平面）把 $y_1, y_2, \dots, y_N$ 区分开，如图 1.5 所示。

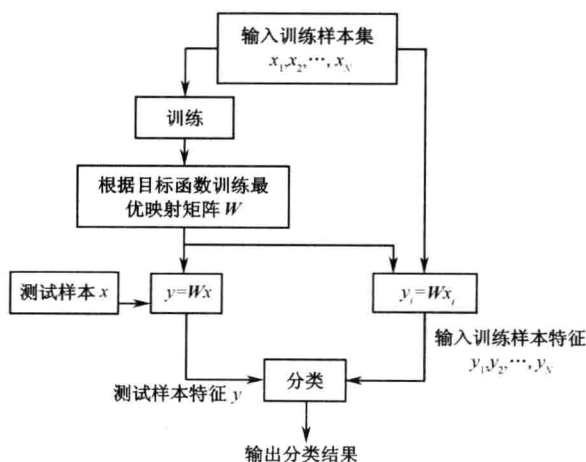


图 1.5 线性特征提取的原理框图

然而不是所有的实际应用都能达到图 1.2 所示的情况。一般说来, 真实世界中一些复杂的应用需要比线性函数更富表达能力的假设空间。换句话说, 目标概念通常不能由给定属性的简单线性函数组合产生, 而是需要利用数据的更抽象的特征。对于一些特定的应用场合, 如人脸识别问题, 姿态、光照和表情变化的影响使人脸图像的特征呈现非线性特性。此时, 用线性特征提取的方法提取特征效果不理想, 在特征空间内, 有少量样本不能由分类面正确分类, 不能达到图 1.2 所示的理想状况, 如图 1.3 所示, 此问题称为线性不可分问题, 也是传统的线性特征提取算法所面临的困难。

为了解决该问题, 采取非线性映射的方法, 将样本从输入空间映射到非线性映射空间, 然后利用线性映射方法将非线性映射空间的数据映射到特征空间, 特征空间内的样本线性可分。令样本为 $x_1, x_2, \dots, x_N$, 非线性映射函数为 $\Phi(x)$, 那么, 样本从输入空间到非线性映射空间 $F = \Phi(x)$ 的映射可以表示为

$$\begin{aligned} \Phi: \mathbb{R} &\rightarrow F \\ x_i &\rightarrow \Phi(x_i), \quad i=1, 2, \dots, N \end{aligned} \quad (1.4)$$

然后对非线性映射空间 F 内的数据进行线性映射, 原始数据 x_i 在映射矩阵 W 下的特征为

$$y_i = W^{\Phi} \Phi(x_i) \quad (1.5)$$

由特征向量 $y_i, i=1, 2, \dots, N$ 表征的样本 $x_i \in R, i=1, 2, \dots, N$ 在特征空间内线性可分, 如图 1.4 所示。

该问题解决思路最大的优势是可以利用线性提取算法的研究成果, 在非线性映射空间 F 中设计线性学习算法。然而, 这种思路也遇到了两个难题: ① 对于某一具体应用问题, 如人脸识别问题, 需要确定到底什么样的非线性映射适合人脸图像的非线性映射, 而对于其他实际应用问题, 又需要确定另外的适合应用对象的非线性映射; ② 即使清楚适合某一具体应用问题的非线性映射, 如何确定非线性映射 $\Phi(x_i)$ 的数学解析式仍然是一个难题。然而, 利用该思路解决线性不可分问题是合理的, 只要找到一种函数来代替非线性映射 $\Phi(x_i)$, 该方法就是可行的。替代函数必须有三个特点: 第一, 具有通用性, 对于大多数的实际应用问题都适用; 第二, 利用该函数可以代替数据在非线性映射空间内的操作; 第三, 该函数形式简单, 在计算上是可实现的。研究表明, 核函数具有以上特点, 是解决上述问题的最佳途径。

1.2.2 核的引入

核方法的首次提出是在 1992 年召开的计算学习理论国际会议上,会议上提出的支持向量机 (SVM) 理论引起了机器学习领域的重要革新,其关键技术是用核函数替换非线性映射的内积来解决非线性问题。利用核函数,通过核代入方法将数据隐含地映射到高维特征空间,这样线性算法即可转换成非线性算法。此后,众多研究者纷纷开始“核化”传统线性算法,通过核代入把传统的线性算法推广为相应的非线性算法,开辟了一个称为“核学习”的研究领域。核学习理论已受到研究者的广泛关注,核方法在模式识别、回归估计等领域得到成功的应用。

1. 核函数的定义

所谓核函数,就是把向量 x 和 y 经过非线性映射在非线形空间上的内积 $\langle \Phi(x), \Phi(y) \rangle$ 用原空间的两向量的一个函数表示,以实现非线性映射。数学表达式为

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (1.6)$$

以上是从函数表达式上进行定义的,从矩阵的角度,核还有另外一个定义。该定义是基于 Gram 矩阵和正定矩阵提出来的, Gram 矩阵和正定矩阵的定义如下。

(1) Gram 矩阵。

给定一个函数 $k: \mathbb{R}^2 \rightarrow K$ 和样本 $x_1, \dots, x_n \in \mathbb{R}$, 其中 $K = C$ 或 $K = R$, C 表示复数集, R 表示实数集,那么,大小为 $n \times n$ 的矩阵 $\mathbf{K}_{ij} := k(x_i, x_j)$ 称为关于 $x_1, \dots, x_n$ 的 Gram 矩阵。

(2) 正定矩阵。

对于 $n \times n$ 矩阵 $\mathbf{K}$, 如果对所有 $c_i \in C$ 都满足 $\sum_{ij} c_i c_j \mathbf{K}_{ij} \geq 0$, 则称 $\mathbf{K}$ 为正定矩阵。

核的定义为: 令 $\mathbb{R}$ 为一非空集, 一个定义在 $\mathbb{R} \times \mathbb{R}$ 上的函数 k , 如果满足对所有 $n \in N$ (N 为自然数集合) 和 $x_1, \dots, x_n \in \mathbb{R}$ 都产生一个正定 Gram 矩阵, 则称其为正定核, 简称为核。

2. 构造核函数

核函数是核学习中的关键因素,不同的核函数会导致 SVM 的泛化能力不同。核函数一直是研究的重点,如何根据所给数据选择合适的核函数成为人们所关注的核心问题。人们根据不同的实际应用构造不同的核函数,在目前大多数核学习