



普通高等教育“十一五”国家级规划教材

高等学校规划教材

# 模式识别

李晶皎 赵丽红 王爱侠 编著

计算机学科教学计划



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

普通高等教育“十一五”国家级规划教材  
高等学校规划教材

# 模式识别

李晶皎 赵丽红 王爱侠 编著

电子工业出版社  
Publishing House of Electronics Industry  
北京·BEIJING

## 内 容 简 介

本书系统阐述了模式识别的原理与方法，并在此基础上介绍了模式识别的应用。全书分为：基础部分和应用部分：基础部分主要包括统计模式识别、模糊模式识别、神经网络模式识别等内容；应用部分有车牌识别和语音识别。本书将理论与实践相结合，有利于读者加深对理论方法的理解，可使读者比较系统地掌握模式识别的理论和相关技术。书中给出的两个应用实例，为读者应用模式识别方法来解决实际问题提供了具体思路和方法。附录给出的习题解答，有利于学生学习理解原理与方法。

本书可以作为高等院校自动化、计算机、生物医学工程等学科本科生、研究生的教材或教学参考书，亦可供有关工程技术人员参考。

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 / 李晶皎, 赵丽红, 王爱侠编著. —北京: 电子工业出版社, 2010.11  
高等学校规划教材  
ISBN 978-7-121-04401-4

策划编辑: 许菊芳

责任编辑: 许菊芳 特约编辑: 王 崧

印 刷: 北京市李史山胶印厂

装 订:

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: 787×1092 1/16 印张: 20 字数: 512 千字

印 次: 2010 年 11 月第 1 次印刷

定 价: 35.00 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](mailto:zltz@phei.com.cn)，盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

服务热线：(010)88258888。

# 前 言

模式识别诞生于 20 世纪 20 年代, 在 20 世纪 60 年代初发展成为一门学科。随着计算机性能的不断提高, 模式识别技术迅速发展, 被广泛应用于人工智能、光学字符识别、生物身份认证、DNA 序列分析、人脸识别、手势识别、语音识别、信息检索、数据挖掘和信号处理、图像识别和理解、视频跟踪识别等。

全书共分 10 章。第 1 章为绪论, 概述了模式识别的概念、方法和应用; 第 2 章为贝叶斯决策理论, 主要介绍了贝叶斯理论、正态分布模式、概率密度函数的估计等; 第 3 章为线性判别函数, 主要介绍了线性判别函数、感知器算法、最小平方误差准则函数、Fisher 线性判别函数等; 第 4 章为模式特征提取与选择, 主要介绍了离散 K-L 变换、离散傅里叶变换、离散余弦和正弦变换、小波变换等; 第 5 章为聚类分析, 主要介绍了相似性测度和聚类准则、层次聚类、 $K$  均值聚类、ISODATA 聚类算法; 第 6 章为人工神经网络, 主要介绍了人工神经网络的构成、多层前馈网络学习算法、联想记忆网络学习算法、Hamming 网络分类学习算法、径向基函数网络等; 第 7 章为支持向量机, 主要介绍了支持向量机的理论基础、常用的几种支持向量机、支持向量回归机等; 第 8 章为核函数方法及应用, 主要介绍了核函数的可分性条件、核函数的参数确定、核函数的构造方法、KPCA 等几种核方法; 第 9 章为模糊模式识别, 主要介绍了模糊数学的基本概念、最大隶属原则和择近原则的模糊模式识别基本方法、模糊  $C$  均值聚类算法、基于模糊等价矩阵的聚类分析等; 第 10 章为模式识别应用, 详细给出了车牌识别和语音识别的系统组成, 以及实现方法和结果; 附录 A 给出了模式识别文献中最著名的数据集之一——鸢尾属植物样本 Iris 数据; 附录 B 给出了各章习题的详细解答。

本书是作者在结合多年教学实践和相关科研成果的基础上编写的。第 1、5、9、10 章由李晶皎、吴秀丽、宋光杰、吴鹏编写, 第 6、7、8 章由赵丽红、王骄、杜玉远、李景宏编写, 第 2、3、4 章由王爱侠、闫爱云、李贞妮、马学文编写。全书由李晶皎负责整理与统稿。

由于作者学识有限, 书中难免有错误和不准确之处, 恳请广大读者批评指正。

作 者

2010 年 8 月于东北大学

# 目 录

<b>第 1 章 绪论</b> .....	(1)
1.1 模式和模式识别的概念 .....	(1)
1.2 模式识别的研究方法 .....	(1)
1.2.1 识别方法 .....	(1)
1.2.2 模式识别系统的组成 .....	(2)
1.3 模式识别的应用 .....	(3)
参考文献 .....	(5)
<b>第 2 章 贝叶斯决策理论</b> .....	(6)
2.1 基于最小错误率的贝叶斯判别法 .....	(6)
2.2 基于贝叶斯公式的几种判别规则 .....	(10)
2.2.1 基于最小风险的贝叶斯决策 .....	(10)
2.2.2 最小最大决策 .....	(13)
2.3 正态分布模式的统计决策 .....	(15)
2.3.1 正态分布概率密度函数的定义及性质 .....	(15)
2.3.2 多元正态概率模型的贝叶斯判别函数 .....	(20)
2.4 概率密度函数的估计 .....	(24)
2.4.1 最大似然估计 .....	(25)
2.4.2 贝叶斯估计 .....	(28)
2.5 离散情况的贝叶斯决策 .....	(31)
2.6 贝叶斯分类器的错误率 .....	(33)
习题 2 .....	(37)
参考文献 .....	(37)
<b>第 3 章 线性判别函数</b> .....	(39)
3.1 线性判别函数 .....	(39)
3.2 广义线性判别函数 .....	(42)
3.3 感知器算法 .....	(44)
3.3.1 基于赏罚概念的感知器训练算法 .....	(44)
3.3.2 梯度下降法 .....	(46)
3.4 最小平方误差准则函数 .....	(47)
3.5 多类问题 .....	(50)
3.5.1 多类问题的基本概念 .....	(50)
3.5.2 决策树简介 .....	(51)



3.6 Fisher 线性判别函数 .....	(54)
习题 3 .....	(56)
参考文献 .....	(57)
<b>第 4 章 模式特征提取与选择 .....</b>	<b>(58)</b>
4.1 离散 K-L 变换 .....	(58)
4.1.1 离散 K-L 展开式 .....	(59)
4.1.2 基于 K-L 变换的数据压缩 .....	(60)
4.1.3 基于 K-L 变换的特征提取 .....	(62)
4.2 离散傅里叶变换 .....	(64)
4.2.1 一维离散傅里叶变换 .....	(64)
4.2.2 二维离散傅里叶变换 .....	(65)
4.3 离散余弦和正弦变换 .....	(67)
4.3.1 余弦变换 .....	(67)
4.3.2 正弦变换 .....	(69)
4.4 Hadamard 变换 .....	(70)
4.5 Haar 变换 .....	(72)
4.6 小波变换 .....	(73)
4.6.1 连续小波变换 .....	(73)
4.6.2 离散小波变换 .....	(75)
4.6.3 多分辨率分析 .....	(75)
4.6.4 正交小波包 .....	(78)
习题 4 .....	(79)
参考文献 .....	(80)
<b>第 5 章 聚类分析 .....</b>	<b>(81)</b>
5.1 相似性测度和聚类准则 .....	(82)
5.1.1 相似性测度 .....	(82)
5.1.2 聚类准则 .....	(83)
5.2 聚类算法 .....	(86)
5.2.1 聚类算法的分类 .....	(86)
5.2.2 层次聚类算法 .....	(87)
5.2.3 K 均值算法 .....	(90)
5.2.4 核聚类 .....	(93)
5.2.5 ISODATA 算法 .....	(95)
5.3 聚类有效性 .....	(99)
习题 5 .....	(101)
参考文献 .....	(102)
<b>第 6 章 人工神经网络 .....</b>	<b>(103)</b>
6.1 人工神经网络的构成 .....	(103)

6.1.1	神经元的结构模型	(103)
6.1.2	人工神经网络的连接方式	(105)
6.1.3	神经网络模型分类	(107)
6.1.4	神经网络学习规则	(108)
6.2	多层前馈网络学习算法	(109)
6.2.1	前馈网络模型	(109)
6.2.2	感知器分类学习算法	(113)
6.2.3	BP 网络分类学习算法	(115)
6.3	联想记忆网络学习算法	(118)
6.3.1	反馈网络模型	(119)
6.3.2	联想记忆分类学习算法	(124)
6.4	海明网络分类学习算法	(127)
6.4.1	海明神经网络结构	(127)
6.4.2	海明网络分类学习算法	(128)
6.5	特征映射网络分类学习算法	(130)
6.5.1	特征映射网络结构	(130)
6.5.2	特征映射分类学习算法	(131)
6.6	前馈网络分类机理	(133)
6.6.1	前馈网络分类的几何机理	(133)
6.6.2	前馈网络分类的代数机理	(137)
6.7	径向基函数网络	(139)
6.7.1	径向基函数	(139)
6.7.2	径向基函数网络的特点	(140)
6.7.3	径向基函数网络的正则化	(142)
	习题 6	(145)
	参考文献	(146)
<b>第 7 章</b>	<b>支持向量机</b>	<b>(149)</b>
7.1	最优分类超平面	(149)
7.2	支持向量机的理论基础	(153)
7.2.1	支持向量机的三种分类形式	(153)
7.2.2	统计学习理论	(160)
7.2.3	优化理论	(166)
7.3	常用的几种支持向量机	(168)
7.3.1	C-支持向量分类机	(168)
7.3.2	C-支持向量机的变形	(174)
7.3.3	广义支持向量机	(175)
7.3.4	$\nu$ -支持向量机	(176)
7.4	支持向量回归机	(178)

7.4.1	回归问题	(178)
7.4.2	线性回归	(179)
7.4.3	非线性回归	(182)
7.4.4	$\epsilon$ -支持向量回归机	(184)
7.4.5	$\nu$ -支持向量回归机	(185)
习题 7		(187)
参考文献		(187)
<b>第 8 章</b>	<b>核函数方法及应用</b>	(189)
8.1	核函数的可分性条件	(190)
8.1.1	输入空间中样本点线性可分的判别条件	(190)
8.1.2	特征空间中样本点线性可分的判别条件	(191)
8.2	核函数的参数确定	(195)
8.3	核函数的构造方法	(196)
8.3.1	基于特征变换的核函数构造	(196)
8.3.2	利用 Mercer 核函数的性质组合核函数	(197)
8.3.3	借助其他领域知识构造核函数	(198)
8.4	几种核方法	(198)
8.4.1	KPCA 的基本思想	(198)
8.4.2	基于类内散布的最优 kernel PCA 展开方法	(201)
8.4.3	融合先验类别信息的非线性主元分析算法	(202)
8.4.4	PKPCA 与 KPCA 和 KFD 的关系	(205)
习题 8		(205)
参考文献		(206)
<b>第 9 章</b>	<b>模糊模式识别</b>	(207)
9.1	模糊数学的基本理论	(207)
9.1.1	模糊集合	(207)
9.1.2	模糊关系	(210)
9.1.3	模糊集合的度量	(213)
9.2	模糊模式识别的基本方法	(217)
9.2.1	最大隶属原则	(217)
9.2.2	择近原则	(218)
9.3	模糊聚类分析方法	(220)
9.3.1	基于模糊等价矩阵的聚类分析	(220)
9.3.2	模糊 $C$ 均值聚类算法	(224)
9.3.3	模糊聚类的有效性	(228)
习题 9		(232)
参考文献		(233)



<b>第 10 章 模式识别应用</b> .....	(235)
<b>10.1 车牌识别</b> .....	(235)
10.1.1 车牌图像预处理 .....	(235)
10.1.2 车牌定位 .....	(239)
10.1.3 字符分割 .....	(246)
10.1.4 字符识别 .....	(247)
<b>10.2 语音识别</b> .....	(252)
10.2.1 语音识别研究的发展与现状 .....	(252)
10.2.2 语音识别方法简介 .....	(254)
10.2.3 DHMM 语音识别系统 .....	(256)
<b>参考文献</b> .....	(280)
<b>附录 A 鸢尾属植物样本数据(Iris Data)</b> .....	(283)
<b>附录 B 习题解答</b> .....	(285)
习题 2 .....	(285)
习题 3 .....	(288)
习题 4 .....	(289)
习题 5 .....	(292)
习题 6 .....	(298)
习题 7 .....	(299)
习题 8 .....	(302)
习题 9 .....	(303)

# 第1章 绪 论

## 1.1 模式和模式识别的概念

模式识别诞生于20世纪20年代。随着20世纪40年代计算机的出现,20世纪50年代人工智能的兴起,模式识别在20世纪60年代迅速发展成为一门学科。在20世纪60年代以前,模式识别主要限于统计学领域的理论研究,计算机的出现增加了对模式识别实际应用的需求,也推动了模式识别理论的发展。

在日常生活中,我们经常进行模式识别活动,例如,收听广播就是在做语音识别,阅读报纸就是在做文字识别,看照片就是在做图像识别。人通过自己的感觉器官从外界获取信息,经过思维、分析、判断,建立对客观世界各种事物的认识;人通过视觉获得形状、大小、色彩等信息,通过听觉获得各种声音的信息,通过触觉获得温度、湿度、材质等信息。人从各个方面获取信息,进行综合思维,认识各种客观事物。随着计算机的发展,人们一直希望计算机能够具有人的能力。

模式识别(Pattern Recognition)就是研究用计算机实现人类的模式识别能力的一门学科,目的是利用计算机将对象进行分类。这些对象与应用领域有关,它们可以是图像、信号,或者是任何可测量且需要分类的对象,对象的专业术语就是模式(Pattern)。按照广义的定义,存在于时间和空间中可观察的事物,如果可以区别它们是否相同或相似,都可以称为模式。

模式识别是一个多领域的交叉学科,它涉及人工智能、统计学、计算机科学、工程学、医学等众多的研究问题。例如,语音识别、字符识别、医学图像识别、医疗诊断、商品销售分析等,吸引了众多的研究人员,且人们提出了许多新方法。在20世纪80年代,基于知识的系统和神经网络发展迅速。近年来,在概率和统计交叉的领域取得重大进展,例如,核函数方法的核贝叶斯计算方法。到目前为止,模式识别的理论和技术的还远未完善,尚有很多课题有待人们去研究和探索。

## 1.2 模式识别的研究方法

### 1.2.1 识别方法

根据所采用的数学模型,模式识别的分类主要如下。

(1) 统计模式识别。主要是指依据模式特征数据的统计分析而建立的数学模型的方法。

(2) 结构模式识别。主要依据模式内部结构关系数据和模式之间的结构关系数据,采用语言结构分析方法进行识别。

模式识别仍然是一门发展中的学科,新的理论和方法不断出现。在模式识别中应用模糊数学和人工神经网络的方法,取得了良好效果。

模糊模式识别就是利用模糊数学的理论和方法来分析解决模式识别问题。这种方法既有模糊数学基础，又接近于人的思维方法，因此适合于分类识别对象本身或要求识别结果具有模糊性的场合。典型的模糊模式识别方法有模糊  $K$  均值和模糊 ISODATA 算法。

人工神经网络是对人脑结构和功能的简单模拟和近似，它是由大量神经元相互连接构成的非线性动力学系统。人工神经网络在自学习、自组织、联想记忆及容错方面具有较大的能力，可以处理一些环境信息十分复杂、背景知识不清楚、推理规则不明确的识别问题。

## 1. 统计模式识别

在统计模式识别中，被研究的模式用特征向量来描述，特征向量中的每一个元素代表模式的一个特征，特征向量构成的空间称为特征空间。一般情况下，合理的假设是：同类模式在特征空间中相距较近，而不同类模式在特征空间中相距较远。这是因为相距较近的模式各个特征相差不多，属于同一类的可能性较大。如果用某种方法分割特征空间，使得同一类模式大体上都在特征空间的同一个区域中，对于待分类的模式，就可以根据它的特征向量位于特征空间中的哪一个区域来判定它属于哪一类模式。统计模式识别的任务就是用不同的方法划分特征空间，从而达到识别的目的。

## 2. 结构模式识别

结构模式识别方法主要分析模式的结构信息。由于模式是由一些模式基元按一定的结构规则组合而成的，因此结构分析的内容就是分析模式如何由基元构成的规则。目前比较成功的是句法结构模式识别方法，它通过检查代表这个模式的句子是否符合事先给定的某一类文法规则来达到识别的目的，即如果符合，那么该模式就属于这个语法所代表的那个模式类。

介绍模式识别最新方法和理论的中英文期刊很多，主要有《模式识别与人工智能》、《中国图像图形学报》、*Pattern Recognition Letters*、*Pattern Recognition*、*IEEE Transactions on Pattern Analysis and Machine Intelligence* 等，以及有关的图像处理 and 模式识别应用类杂志。

### 1.2.2 模式识别系统的组成

本节以统计模式识别为例，介绍模式识别系统的组成。图 1.1 给出了一个统计模式识别系统的简单框图，它主要由信息获取、预处理、特征提取与选择、分类器设计、分类决策这五个模块组成。

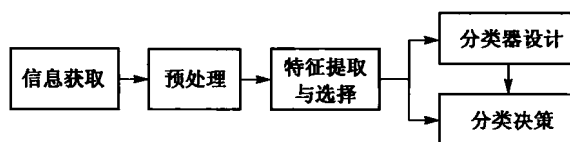


图 1.1 模式识别系统的基本构成

#### 1. 信息获取

为了使计算机能够分类识别对象(模式)，必须首先将对象用计算机所能接受的形式表示。目前待识别模式大多是非电量输入模式，如灰度、色彩、声音、压力、温度等，需要将

这些以各种不同形式表现的信息通过相应的传感器转换成电信号，然后经过 A/D 变换，最终转换为计算机能接受的数字量。通常输入对象的信息有下列三种类型：

- (1) 二维图像。如指纹、照片、文字等。
- (2) 一维图像。如语音信号、心电图、机械振动波等。
- (3) 物理参数和逻辑值。如体温、各种实验数据等。

通过测量、采样和量化，可以用矩阵或向量表示二维图像或一维波形。这就是信息的获取过程。

## 2. 预处理

预处理的目的是去除信息获取过程中掺入的干扰和噪声，人为地加强有用信息，并对各种因素造成的退化现象进行复原。

## 3. 特征提取与选择

由信息获取部分得到的原始数据量一般都很大。例如，一幅文字图像可以有几千个数据，100 ms 的语音信号也有几千个数据。为了有效地实现分类识别，要对原始数据进行选择和变换，得到最能反映分类本质的特征，而由这些特征组成的向量则称为特征向量。上述过程就是特征提取与选择过程。一般我们把原始数据组成的空间称为测量空间，把经过特征提取与选择后得到的特征向量空间称为特征空间。特征空间中的一个点就是一个特征向量，它代表一个模式或样本，特征向量的每一个分量就是模式的一个特征。在模式识别中，特征提取与选择占有重要的地位，但尚无通用的理论指导，只能通过分析具体识别对象来确定选取何种特征。

## 4. 分类器设计

为了把待识别模式分配到各自的模式类中，必须设计出一套分类判别规则。基本做法是：用一定数量的样本(或训练样本集)确定出一套分类判别规则，使得按这套分类规则对待识别模式进行分类所造成的错误识别率最小或引起的损失最小。这就是分类器设计过程。

## 5. 分类决策

分类器按照已经确定的分类规则将待识别模式进行分类判别，输出分类结果。这就是分类器的使用过程，也称为分类决策。

在模式识别系统中，信息获取和预处理部分通常是数字信号处理和数字图像处理等课程的研究课题，本书只讨论特征提取与选择、分类器设计以及分类决策的理论和方法。

# 1.3 模式识别的应用

随着模式识别的迅速发展，模式识别技术已在越来越多的领域中得到应用。

## 1. 文字识别

迄今为止，在模式识别领域中发展最成熟、应用最广泛的一个方面就是文字识别。各种成熟的光学字符识别(Optical Character Recognition, OCR)系统已经在使用，如可识别手写

体阿拉伯数字的邮政信函自动分拣系统，可以识别数字及数字个数的银行支票机器识别系统，等等。

按照识别对象划分，文字识别分为英文字符识别、阿拉伯数字识别和汉字识别。按照书写方式划分，文字识别可分为印刷体识别和手写体识别。由于汉字结构复杂、种类多，因而汉字识别难度最大。联机手写汉字识别，由于利用书写板输入了汉字的笔顺信息，因此降低了识别难度，故已有实际应用；而脱机手写汉字识别还处于实验室阶段。

## 2. 语音识别

由于语音是人类最自然的沟通和交换信息的方式，因而成为计算机人机接口的关键技术。从原理上看，语音识别虽然实现起来并不困难，但在实际实现时会遇到很多困难，主要表现为：①发音的多变性。不同人发同一个音、同一个人不同的条件下发同一个音等，都会有不同的发音特征参数；②发音的模糊性。在实际的连续语音流中，语音声学变量与音素变量之间不存在一一对应的关系，语音流中存在变化多端的音变现象，这些音变对人类的听觉系统来说很容易辨认，但机器识别却很不容易。

语音识别的应用很广，如声控打字、用声音控制计算机等。如将语音识别与语音合成结合起来，则可以实现甚低比特率的语音通信。目前从事语音识别研究的组织很多，如 IBM、Microsoft、Motorola、Siemens、Nokia、Toshiba 等公司，以及美国的卡内基·梅隆大学(CMU)、麻省理工学院(MIT)和我国的清华大学等。一些中、小词汇量的孤立词或连续语音识别系统已经进入市场。据预测，带有语音功能的计算机将很快成为大众化产品，语音识别将可能取代键盘和鼠标成为计算机的主要输入手段。

## 3. 指纹识别

手掌及其手指、脚、脚趾内侧表面的皮肤凹凸不平所产生的纹路会形成各种各样的图案，而这些皮肤的纹路在图案、断点和交叉点上各不相同，具有唯一性。依靠这种唯一性，就可以将一个人与其指纹对应起来，通过将其指纹和预先保存的指纹进行比较，便可以验证他的真实身份。从 20 世纪 60 年代开始，随着计算机技术的发展，人们开始研究利用计算机处理指纹，自动指纹识别系统在法律实施方面的研究和应用已在世界许多国家展开。20 世纪 80 年代，随着计算机和光学扫描技术的迅速发展，使得它们作为指纹取像工具成为现实，从而指纹识别得到了广泛应用。20 世纪 90 年代后期，随着低价位的取像设备的飞速发展，以及可靠的比对识别算法的实现，指纹识别在个人身份识别中得到了广泛应用。

## 4. 生物医学应用

模式识别已经广泛应用于生物医学，例如，心电图和心电向量图的分析，脑电图的分析，染色体的自动分类，癌细胞分类，血相分析，X 光片、CT 片、磁共振片等医学图片的分析。

## 5. 其他方面的应用

(1) 对地球资源和环境的调查研究。在这类应用中，人们利用遥感卫星或飞机获取的大量信息，经过图像处理、模式识别来调查资源情况，例如调查林业资源、水资源及矿藏资源

的分布, 监测大气变化, 调查环境污染, 调查土地利用和规划状况, 进行农作物长势监测, 虫灾监测, 等等。

(2) 生产自动化过程中的应用。在这类应用中, 人们使用模式识别来对自动化生产线上的产品进行质量检验。例如, 在大规模集成电路的生产中, 芯片内的缺损检验以及芯片上引出端的自动识别和引出线焊接等。

(3) 军事应用。在这类应用中, 人们使用模式识别来对可见光、雷达、红外线图像进行分析与识别, 检出和鉴别目标的出现, 判断目标的类别, 并对运动中的目标进行监视和跟踪; 采用地形匹配方法校正飞行器轨道, 提高导弹命中率; 等等。

## 参考文献

- [1] 边肇祺编著. 模式识别(第二版). 北京: 清华大学出版社, 2000.
- [2] 杨光正编著. 模式识别. 合肥: 中国科学技术大学出版社, 2007.
- [3] 钟珞编著. 模式识别. 武汉: 武汉大学出版社, 2006.
- [4] Andrew R. Webb 著, 王萍译. 统计模式识别(第二版). 北京: 电子工业出版社, 2004.
- [5] Richard O. Duda 著, 李宏东译. 模式分类. 北京: 机械工业出版社, 2003.
- [6] 蔡元龙编. 模式识别. 西安: 西北电讯工程学院出版社, 1986.
- [7] J. P. Marques de Sa 著, 吴逸飞译. 模式识别——原理、方法及应用. 北京: 清华大学出版社, 2002.
- [8] 苏宁编著. 模式识别的理论与方法. 武汉: 武汉大学出版社, 2004.
- [9] Sergios Theodoridis 著, 李晶皎译. 模式识别(第三版). 北京: 电子工业出版社, 2006.
- [10] 张宏林编著. Visual C++数字图像模式识别技术及工程实践. 北京: 人民邮电出版社, 2003.
- [11] Lawrence Rabiner. *Fundamentals of Speech Recognition*. 北京: 清华大学出版社, 1999.
- [12] Alan Bovik. *Handbook of Image and Video Processing, Second Edition*. 北京: 电子工业出版社, 2006.

## 第2章 贝叶斯决策理论

在模式识别方法中，通过模式分类将特征空间分割成若干区域，使每个区域对应一个模式类别。在理想情况下，基于这些区域分割而进行的决策不应该产生错误，如果在实际中做不到这一点，则要求分类的错误代价尽量小；如果一些错误比另一些错误要付出更多的代价，则要求分类错误的平均代价最小<sup>[1]</sup>。

模式识别的一种主要处理方法是贝叶斯(Bayes)决策理论。这种方法的基本思路是，在假设决策问题可以用概率的形式描述，并且所有有关的概率结构均已知的的前提下，决策者根据已经获得的历史资料数据以及主观知识(包括经验、直觉、判断等)，对未来事件发生的概率做出主观估计(即先验概率)，最后根据期望值的计算结果做出决策选择。由于先验状态分布与实际情况存在一定的误差，所以它很难准确地反映客观真实情况，而且有时候决策结果对先验概率又非常敏感，所以必须通过市场调查等方法收集有关自然状态的补充信息，以修正对事件的先验概率估计(得到后验概率)，最后用后验状态分布进行决策。贝叶斯决策理论提供了一种修正先验概率的科学方法。

用贝叶斯理论进行分类时要求满足两点：第一，要决策的类别数是一定的。例如两类样本(正常状态 $\omega_1$ 和异常状态 $\omega_2$ )，或 $L$ 类样本 $\omega_1, \omega_2, \dots, \omega_L$ ；第二，各类别总体的概率分布是已知的，即每一类样本出现的先验概率 $p(\omega_i)$ 以及各类概率密度函数 $p(x|\omega_i)$ 是已知的。

显然， $0 \leq p(\omega_i) \leq 1, i = 1, 2, \dots, L$ ，且 $\sum_{i=1}^L p(\omega_i) = 1$ <sup>[2]</sup>。对于两类故障诊断问题，相当于在识别前已知正常状态 $\omega_1$ 的概率 $p(\omega_1)$ 和异常状态 $\omega_2$ 的概率 $p(\omega_2)$ ，它们是由先验知识确定的状态先验概率。如果不做进一步的仔细观测，仅依靠先验概率做决策，那么就会给出这样的决策规则：若 $p(\omega_1) > p(\omega_2)$ ，则做出状态属于 $\omega_1$ 类的决策；反之，则做出状态属于 $\omega_2$ 类的决策。例如，某设备在365天运行中，发生故障是少见的，无故障是经常的，有故障的概率远小于无故障的概率。因此，若无特别明显的异常状况，就应判断为无故障。由于只利用先验概率提供的分类信息太少了，所以这样判断对于某一个实际的待检状态根本达不到诊断的目的。为此，我们还要对系统状态进行状态检测，分析所观测到的信息。根据观测信息，结合先验概率再对状态进行归类。

如今贝叶斯理论广泛应用于各个领域，如工程技术、管理科学、系统运筹、医疗诊断等。本章介绍这种方法的基本内容。

### 2.1 基于最小错误率的贝叶斯判别法

在模式分类中，要尽量减少分类错误的概率。从这样的要求出发，利用概率论中的贝叶斯公式，能得到使错误率最小的分类规则，这称为基于最小错误率的贝叶斯判别法<sup>[3]</sup>。

假定一个两类问题，先验概率分别为 $p(\omega_1)$ 和 $p(\omega_2)$ 。令 $x$ 为 $N$ 维向量， $x$ 的类条件概率密度为 $p(x|\omega_i), i = 1, 2$ 。根据全概率公式，模式样本 $x$ 出现的全概率密度为



$$p(\mathbf{x}) = \sum_{i=1}^2 p(\mathbf{x}|\omega_i)p(\omega_i) \quad (2.1)$$

根据贝叶斯公式, 在模式  $\mathbf{x}$  出现的条件下, 两个类的后验概率为

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{\sum_{i=1}^2 p(\mathbf{x}|\omega_i)p(\omega_i)} \quad (2.2)$$

因此, 贝叶斯公式实际上是通过观察  $\mathbf{x}$ , 把状态的先验概率  $p(\omega_i)$  转化为状态的后验概率  $p(\omega_i|\mathbf{x})$ 。要判断  $\mathbf{x}$  是属于  $\omega_1$  类还是属于  $\omega_2$  类, 从概率统计的观点来看,  $\mathbf{x}$  来自于哪类的概率大, 就属于哪类, 这样能够使错误概率最小。因此基于最小错误率的贝叶斯判别规则为

$$\text{如果 } p(\omega_1|\mathbf{x}) > p(\omega_2|\mathbf{x}), \text{ 那么 } \mathbf{x} \in \omega_1; \text{ 否则 } \mathbf{x} \in \omega_2 \quad (2.3)$$

由于  $p(\omega_1|\mathbf{x}) + p(\omega_2|\mathbf{x}) = 1$ , 因此上述规则可写成

$$\begin{cases} \text{如果 } p(\omega_1|\mathbf{x}) > \frac{1}{2}, \text{ 那么 } \mathbf{x} \in \omega_1 \\ \text{如果 } p(\omega_2|\mathbf{x}) > \frac{1}{2}, \text{ 那么 } \mathbf{x} \in \omega_2 \end{cases} \quad (2.4)$$

由贝叶斯定理, 上述规则可进一步表示成

$$\begin{cases} \text{如果 } p(\omega_1)p(\mathbf{x}|\omega_1) > p(\omega_2)p(\mathbf{x}|\omega_2), \text{ 那么 } \mathbf{x} \in \omega_1 \\ \text{如果 } p(\omega_1)p(\mathbf{x}|\omega_1) < p(\omega_2)p(\mathbf{x}|\omega_2), \text{ 那么 } \mathbf{x} \in \omega_2 \end{cases} \quad (2.5)$$

或者可以写成

$$\begin{cases} \text{如果 } l_{12}(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{p(\omega_2)}{p(\omega_1)} = \theta_{21}, \text{ 那么 } \mathbf{x} \in \omega_1 \\ \text{如果 } l_{12}(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} < \frac{p(\omega_2)}{p(\omega_1)} = \theta_{21}, \text{ 那么 } \mathbf{x} \in \omega_2 \end{cases} \quad (2.6)$$

其中,  $l_{12}(\mathbf{x})$  称为似然比函数,  $\theta_{21}$  称为似然比的判决阈值<sup>[4]</sup>。

**【例 2.1】** 对一大批人进行癌症普查, 设  $\omega_1$  类代表正常人,  $\omega_2$  类代表癌症患者。已知先验概率如下: 正常状态  $p(\omega_1) = 0.9$ , 异常状态  $p(\omega_2) = 0.1$ 。以一个化验结果作为特征  $\mathbf{x}$ : {阳性, 阴性}, 正常人和癌症患者的化验结果为阳性的概率分别为

$$p(\mathbf{x} = \text{阳性}|\omega_1) = 0.2, \quad p(\mathbf{x} = \text{阳性}|\omega_2) = 0.4$$

现有一人化验结果为阳性, 问此人是否为癌症患者<sup>[2]</sup>?

解: 利用贝叶斯公式, 分别计算出  $\omega_1$  及  $\omega_2$  的后验概率如下:

$$p(\omega_1|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)p(\omega_1)}{\sum_{j=1}^2 p(\mathbf{x}|\omega_j)p(\omega_j)} = \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.818$$

$$p(\omega_2|\mathbf{x}) = 1 - p(\omega_1|\mathbf{x}) = 0.182$$

根据贝叶斯决策规则公式, 有

$$p(\omega_1 | x) > p(\omega_2 | x)$$

所以合理的判别是把  $x$  归类于  $\omega_1$ , 属于正常状态。

从这个例子可以看出, 决策结果取决于实际观察到的先验概率  $p(\omega_i)$  和类条件概率密度  $p(x | \omega_i)$ , 根据先验概率  $p(\omega_i)$  和类条件概率密度  $p(x | \omega_i)$  计算后验概率  $p(\omega_1 | x)$  和  $p(\omega_2 | x)$ , 基于最小错误率的贝叶斯决策理论就是根据后验概率的大小进行分类决策的。因为  $p(\omega_1 | x) > p(\omega_2 | x)$ , 所以做出  $x \in \omega_1$  的决策; 而如果  $p(\omega_1 | x) < p(\omega_2 | x)$ , 则  $x \in \omega_2$ 。在这个例子中, 由于状态  $\omega_1$  的先验概率比  $\omega_2$  的先验概率大几倍, 故先验概率在最后决策中起了主导作用。

在前面我们只是给出了最小错误率贝叶斯决策规则, 而没有证明按照这种规则进行分类能够使错误率最小。现在仅以一维情况为例来完成证明, 其结果不难推广到多维情形。

所谓错误率, 是指平均错误率, 它以  $P(e)$  来表示, 其定义为

$$P(e) = \int_{-\infty}^{+\infty} p(e, x) dx = \int_{-\infty}^{+\infty} p(e | x) p(x) dx \quad (2.7)$$

其中  $\int_{-\infty}^{+\infty} (\ ) dx$  表示在整个特征空间上的积分。

对于两类别问题, 由式(2.3)可知, 如果  $p(\omega_1 | x) < p(\omega_2 | x)$ , 那么判断结果应为  $x \in \omega_2$ 。显然在做出决策  $x \in \omega_2$  时,  $x$  的条件错误概率为  $p(\omega_1 | x)$ ; 反之,  $x$  的条件错误概率应为  $p(\omega_2 | x)$ 。从而  $x$  的条件错误概率可表示为

$$p(e | x) = \begin{cases} p(\omega_1 | x), & \text{当 } p(\omega_1 | x) < p(\omega_2 | x) \\ p(\omega_2 | x), & \text{当 } p(\omega_1 | x) > p(\omega_2 | x) \end{cases} \quad (2.8)$$

如果令  $t$  为  $\omega_1, \omega_2$  两类的分界面, 则当特征向量  $x$  是一维向量时,  $t$  为  $X$  轴上的一点, 而且  $t$  点将  $X$  轴分为两个区域  $R_1$  和  $R_2$ 。  $R_1$  的范围为  $(-\infty, t)$ ,  $R_2$  的范围为  $(t, +\infty)$ , 这样就有

$$\begin{aligned} P(e) &= \int_{-\infty}^t p(\omega_2 | x) p(x) dx + \int_t^{+\infty} p(\omega_1 | x) p(x) dx \\ &= \int_{-\infty}^t p(x | \omega_2) p(\omega_2) dx + \int_t^{+\infty} p(x | \omega_1) p(\omega_1) dx \end{aligned} \quad (2.9)$$

上式可以写为

$$\begin{aligned} P(e) &= p(x \in R_1, \omega_2) + p(x \in R_2, \omega_1) \\ &= p(x \in R_1 | \omega_2) p(\omega_2) + p(x \in R_2 | \omega_1) p(\omega_1) \\ &= p(\omega_2) \int_{R_1} p(x | \omega_2) dx + p(\omega_1) \int_{R_2} p(x | \omega_1) dx \\ &= p(\omega_2) P_2(e) + p(\omega_1) P_1(e) \end{aligned} \quad (2.10)$$

所以总的错误率是两种分类错误率的加权和。

图2.1<sup>[1]</sup>说明了一维模式的情况。  $t$  为  $R_1$  和  $R_2$  区域的分界。显然  $t$  的位置不同, 则错误率也不同。图中两个画线(斜线和纹线)部分之和, 包括黑色区域  $A$ , 为总的错误率。当把决策面  $t$  左移时, 可以减小代表误分类的三角区域  $A$  的面积, 从而减小分类错误概率。若选取决策面  $t$  使得