

77
GEO-SPATIAL INFORMATION SCIENCE

● 高等学校摄影测量与遥感专业教材

模式识别的理论与方法

The Theories and Methods
of Pattern Recognition

舒宁 马洪超 孙和利 编著



全国优秀出版社
武汉大学出版社

高等学校摄影测量与遥感专业教材

0235
16

模式识别的理论与方法

The Theories and Methods
of Pattern Recognition

舒宁 马洪超 孙和利 编著

北方工业大学图书馆



00589975



全国优秀出版社
武汉大学出版社

12/163 01

模式识别的理论与方法/舒宁,马洪超,孙和利编著. —武汉:武汉大学出版社,2004.12

高等学校摄影测量与遥感专业教材

ISBN 7-307-04363-7

I. 模… II. ①舒… ②马… ③孙… III. 模式识别—高等学校—教材 IV. TP391.4

中国版本图书馆 CIP 数据核字(2004)第 101096 号

责任编辑:王金龙 责任校对:黄添生 版式设计:支 笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

印刷:武汉市楚风印刷有限公司

开本:787×1092 1/16 印张:12 字数:277 千字

版次:2004 年 12 月第 1 版 2004 年 12 月第 1 次印刷

ISBN 7-307-04363-7/TP·157 定价:20.00 元

版权所有,不得翻印;所购教材,如有缺页、倒页、脱页等质量问题,请与当地教材供应部门联系调换。

前 言

模式识别的理论与方法教材是在参考有关教材的基础上，为适应遥感科学与技术、摄影测量与遥感、遥感信息工程等专业方向本科生教学的需要，并结合近年来有关科研成果和目前国内外遥感影像计算机处理的研究进展而编写出来的，包括了模式识别的四大方面，即统计模式识别、模糊模式识别、句法模式识别和智能模式识别的基本理论和方法，目的在于使学生对模式识别有个总的认识。论述的重点在统计模式识别，因为这是理论上十分成熟、应用最广的方法。其他方面也作了基本内容的介绍。由于内容很多，教学中将有选择地讲授。这本教材以尽可能做到适于自学为原则，在文字叙述、理论与公式推导上均有所斟酌。笔者编写本教材历时两年，但仍感仓促，书中缺点、错误在所难免，希望采用者及相关研究人员提出宝贵意见。

本教材前七章由舒宁教授编写，第八章由孙和利副教授编写，马洪超教授编写了所有的思考与练习题，并编写了有关概率论和矩阵论的一些基础知识作为附录一的内容，编写了一些很好的应用实例作为附录五的内容，为帮助学生更好地理解 and 掌握知识提供了资料。愿读者能从中受益。

编 者

2004 年 7 月

目 录

前 言	1
第一章 概论	1
第二章 贝叶斯决策理论	4
第一节 基于最小错误率的贝叶斯判别法	4
第二节 基于贝叶斯公式的几种判别规则	8
一、基于最小风险的贝叶斯决策	8
二、聂曼-皮尔逊决策法	10
三、最小最大决策	13
四、序贯分类决策	15
第三节 正态分布模式的统计决策	16
一、正态分布概率密度函数的定义及性质	16
二、多元正态概率模型的贝叶斯判别函数	20
第四节 概率密度函数的估计	23
一、均值向量和协方差矩阵的参数估计	23
二、概率密度函数的函数近似	24
三、后验概率的函数估计	27
四、均值向量和协方差矩阵的贝叶斯估计	29
第五节 贝叶斯分类器的错误概率	32
第三章 判别函数与确定性分类器	39
第一节 线性判别函数与广义线性判别函数	39
一、两类别问题	39
二、多类别问题	40
第二节 广义线性判别函数的二分能力	44
一、二分法	44
二、广义线性判别函数的二分能力	46
第三节 感知器算法	47
一、基于赏-罚概念的感知器训练算法	47
二、梯度法	49
第四节 最小平方误差算法(LMSE)	50

第五节 多类模式的分类器迭代算法	52
第六节 势函数法	55
一、判别函数的产生	55
二、势函数的选取	56
第四章 聚类分析	60
第一节 按距离聚类的概念	60
第二节 模式相似性测度与聚类准则	61
一、模式相似性测度	61
二、聚类准则	63
第三节 聚类算法	65
一、简单搜索算法	65
二、最大的最小距离算法	65
三、K 均值算法	67
四、ISODATA 算法	69
五、系统聚类法(Hierarchical Clusterring Method)	73
六、模式样本的选取和初始聚类中心的确定	74
第四节 对聚类的评价	75
一、聚类几何分布的显示	75
二、图表分析	75
三、屏幕显示与实地(或实物)抽样检核	76
四、借助有关文献、图件和已有成果评价聚类结果	77
第五章 模式特征分析与选取	78
第一节 描述模式分布状态的测度	79
一、欧氏距离	79
二、离散度	80
三、特征变换与排序	81
四、熵	82
第二节 特征提取的方法	83
一、独立特征的选取准则	83
二、一般情况下的特征选取准则	83
三、穷举法	85
四、最小熵方法	85
五、OIF(Optimum Index Factor)指标	86
第三节 离散 K-L 变换	86
一、离散 K-L 展开	87
二、根据 K-L 展开式选择特征	88
第四节 利用 K-L 变换的分类特征提取	91

一、按总体散布矩阵作 K-L 变换	91
二、按类内散布矩阵作 K-L 变换	91
三、按类间散布矩阵作 K-L 变换	91
第五节 增维问题	95
一、辅助数据的应用	96
二、数据复合技术的利用	96
三、纹理分析	97
四、智能化——专家意识的注入	98
第六章 模糊集合理论在模式识别中的应用	100
第一节 引言	100
第二节 模糊子集	100
第三节 模糊集的简单运算、模糊度与模糊关系	103
一、模糊集的简单运算	103
二、模糊度	104
三、模糊关系和模糊关系矩阵	105
第四节 隶属函数的确定	107
一、模糊统计法	107
二、二元对比排序法	110
三、逻辑推理法	113
四、专家评判法	113
五、采用常见的隶属函数的函数逼近法	113
第五节 模糊集理论在模式识别中的应用	114
一、最大隶属原则识别法	114
二、择近原则识别法	116
三、模糊聚类分析法	116
第七章 句法模式识别	123
第一节 形式语言基础和文法	123
一、字母表和符号串	123
二、文法	124
三、句子与语言	125
四、文法的四种基本类型	125
第二节 一维及高维文法	127
一、串文法	127
二、图像描述语言 PDL(Picture Description Language)	128
三、树文法	129
第三节 基元提取和文法推断	130
一、基元提取	131

二、文法推断	132
第四节 句法分析和自动机识别	134
一、句法分析	134
二、句法结构的自动机识别	135
第八章 人工神经网络在模式识别中的应用	137
第一节 人工神经网络概述	137
一、概述	137
二、人工神经网络基础	138
第二节 前馈神经网络及其主要算法	142
一、前馈神经网络	142
二、感知器	142
三、三层前馈网络	143
四、反向传播算法(BP法)	144
五、径向基函数网络	146
第三节 Hopfield 网络	148
一、Hopfield 网络概述	148
二、离散 Hopfield 网络(DHNN)	148
三、联想存储器	149
四、优化计算	150
五、连续 Hopfield 网络(CHNN)	150
第四节 自适应共振理论神经网络	151
一、概述	151
二、ART 网络的结构及原理	151
三、ART1 网络算法步骤	153
第五节 自组织特征映射神经网络	154
一、概述	154
二、SOFM 网络模型及功能	155
三、SOFM 网络原理	156
附录	
一、概率论和矩阵论的有关内容	160
二、 n 维模式的线性二分法公式推导	166
三、感知器算法的收敛证明	167
四、多元正态随机向量边缘分布、条件分布和线性变换的正态性证明	170
五、应用实例	172
六、几种典型的隶属函数分布	176
参考文献	182

第一章 概 论

“模式”是一个抽象的概念。客观世界和主观世界即物质和意识的所有方面、所有个体、所有单元、所有事物都可以称之为模式。客观世界的事物，形形色色，种类万千，每一种事物都叫一个模式。主观意识，形态各异，方式有别，每一种思维也叫一个模式。整个宇宙，整个世界也可以称为模式。人类为了改造世界，首先必须认识它，还要认识人类自己，包括自己的意识、思维，这就是对模式的识别。

人通过自己的感觉器官从外界获取信息，然后经过思维、分析、判断，建立对客观世界各种事物的认识，比如通过视觉获得形状、大小、色彩等特征信息，映入脑海中构成一幅幅图像；通过听觉取得各种音响的信息；通过触觉得知温度、湿度、材料强度等；还有从嗅觉闻到种种气味，等等。人们从各方面获得信息，然后进行综合思维，认识各种客体。

所谓模式识别技术，是随着现代科学技术的发展，特别是计算机技术的发展而形成的一种模拟人的各种识别能力（当前主要是对视觉能力和听觉能力的模拟）和方法的技术。它基本上属于一种自动判别和分类的理论。这一理论孕育于 20 世纪 60 年代，随着科学技术的发展，特别是 20 世纪 70 年代遥感技术的发展和地球资源卫星的发射，人们通过遥感从卫星取得的巨量信息，需要进行空前规模的处理、识别和应用，在此推动之下，模式识别技术便得以迅速发展。

模式识别技术的目标是模拟人的识别方法，但这一点是极不容易完全做到的。图像识别、机器人发展起来了，而人工智能、专家系统还多停留在理论研究阶段，实际的成功事例，寥寥无几，所以目前模式识别技术仍处于初级阶段，它的理论还要完善和发展，它的应用领域还有待于拓宽和深入，它的尖端堡垒还有待于攻克。

尽管如此，模式识别技术仍然在不少领域得到了应用，有的已经取得了很大的成效，有的正在其发展过程中。这些领域大致有以下几个方面：

1. 对地球资源、环境的调查研究。这主要是利用遥感卫星或飞机所获取的大量信息，经过图像处理、模式识别来调查资源情况，如林业资源、水资源以及矿藏资源的分布等；监测大气变化，以作预报和全局分析；调查人类生存环境的污染状况，分析根源，以求对策；调查土地利用状况，以利合理规划与开发；进行农作物长势监测，虫灾监测，产量预估；作城市多维信息分析，以利于规划与管理……这些应用对模式识别理论与技术发展的推动非常大，所取得的成效也是显著的。

2. 在生物医学工程中的应用。在这个领域里的应用也是十分广泛的。各种生理信号的分析、识别，病症判断，健康检查，都需要分析各种各样的数据。这些数据或为曲线，或为图像，如脑电图、心电图、血压波等，是一维波形数据，而放射性图片，如 X 光片，显微图片如染色体、白血球、癌细胞组织的图片等，是二维信号，这种二维信号若形成时

间的序列或轴向（垂直于图片方向）序列，就是三维的信息了。对于诸如此类的波形、图片的分析都需要利用模式识别技术。

3. 生产自动化过程中的应用。这主要是在自动化生产线上的质量检验和成品安装中的应用，例如大规模集成电路生产中，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。发达工业国家广泛利用智能机器人，它具有“视觉”和“听觉”，产品品种改变时，机器人可以随之改变操作，很容易形成新的流水线，因而提高了流水线的灵活度。

4. 文件处理和管理自动化。在现代信息社会，各类大量数据都利用计算机进行存储和管理，诸如图书馆的书籍登录、索引，商店的进出账目，货物库存结算，银行的各种票据核对等，都采用资料、数据自动识别、登记、分类。

5. 军事应用。这主要表现在瞄准和搜索对方目标方面，如利用带有地面特征的图像制导等。

6. 公安侦破上的应用。这主要是对指纹、人面、足迹的识别。

7. 商业自动化中的应用，如自动售货机、硬币兑换机、自动检票机等。

目前，模式识别技术的应用领域还在不断地拓展，视觉识别方面是比较成熟了，听觉识别的模拟正在发展之中，如口语翻译机，电话自动订票等，正有待于进一步完善。

模式识别技术是在计算机上实现的，而计算机只识别数字和字符，故所有模式都必须首先数值化或符号化，才能进行自动识别。也就是说前提是对模式的特征测量数值化。具体对象不同，采用的数值化方法也不同，通常有以下几种方式：

1) 连续量的量化。例如长度、重量、面积、体积、时间、电流、电磁波长等，都是连续量，其度量结果即特征量化值。

2) 量级的数值化。某些事物的特征有时不需要详尽的数值，而只需要划分成有次序的量化等级。如癌症的检查，按患病程度的不同，有：

0——代表无症状；

1——代表早期症状；

2——代表中期症状；

3——代表晚期。

3) 名义尺度。特征度量时没有数量关系，也没有明显的次序关系，只是需要一些指定的指标。如黑和白，用“b”和“w”表示，或用“0”和“1”表示；又如男性和女性用“G”和“L”表示等。

在模式的特征量化之后，有时需要去掉量化时的“噪声”或在几何位置上进行纠正，或增强所需的信息，这些称为预处理。有时还需要经过分析，去掉对分类识别无效或易造成混淆的那些特征，保留对分类判别有效的特征数据，即特征选择，例如，采取某种变换算法，对原特征数据进行变换，得出有利于分类的综合性特征，或减少特征维数，以提高效率，这些都属于特征选择的工作。在这以后，按所取的分判类的数学模型对模式样本集进行分类，并将结果进行分析（例如与已知类别的模式作对比），以图改进，制定出错误率最小的判别规则，这一过程称之为学习。经过学习所制定的数学模型最后用来对所有未知类别的模式进行分类。上述的模式识别过程大致如图 1-0-1 所示。

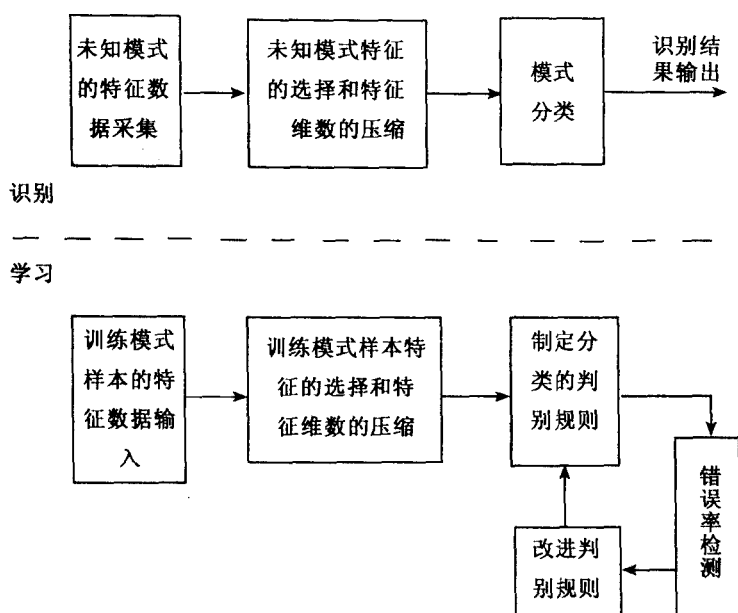


图 1-0-1

在模式识别中所采用的数学模型各种各样，主要依据模式特征数据的统计分析而建立数学模型的方法称为统计模式识别。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的方法称为句法模式识别。在模式识别过程中，引入事物归属的模糊概念和模糊度量，利用模糊数学方法进行识别的方法称为应用模糊集合论的模式识别法。还有近年来发展起来的交互式模式识别法，它综合人的智能与计算机的智能，是一种有实用价值的方法。

有关模式识别的中英文期刊很多，主要的有：《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》，等等，以及有关图像处理和模式识别应用类的杂志。

思考与练习

1. 查阅有关文献，把目前常见的期刊和有关模式识别的中英文专著或期刊列一个表格。查阅有关网站、文献，罗列目前模式识别中的常用算法，并对照自己的数学基础知识，看看哪些是熟悉的，哪些是不熟悉的。对于不熟悉的内容，尽量找相关的数学参考书学习。

第二章 贝叶斯决策理论

模式识别是通过对待识别模式的多种观察或测量,将观测数据构成其特征向量,作为输入,然后按照某一种判决法则来进行分类的。在观察与测量中,当模式表现为具有确定的特征,具有明显的线性可分性(或广义线性可分性),即类别之间具有明显的界限,或大多数类别具有这样的特性,或各类模式的大多数具有这种特征时(如图 2-0-1(a)所示),采用确定性分类器或线性判别分类器,可以收到较好的效果。然而,客观实际是十分复杂的,许多现象在观察与测量时都具有某种不确定性,例如卫星遥感影像。由于那种影像每一个像素相应于地面几十米见方甚至 $1\text{km} \times 1\text{km}$,因而具有很大的综合性,往往是地面上多种地物反射光谱的综合效应,只是在某种地物面积很大的情况下,一个像素才代表某一特定地物。而随着时间的、环境的不同,即使同一地物的光谱也会出现变化。由于这样一些原因,像素的灰度值才具有某种不确定性。模式特征量测与观察值一般都具有随机性,只有在基本条件保持不变,进行大量观测的情况下,才呈现出某种规律性。但是,实际工作中往往不允许或不可能作大量观测,使观测结果带有局限性。另外,观测时的主观臆断或片面性也会使观测结果出现误差。我们称这些误差和观测的局限带来的偏差为观测噪声。这些观测噪声也就是产生上述不确定性的原因。由于这种不确定性,各个类别之间呈现混淆、浑沌的表象,给分类带来了困难。这时,需要采取统计方法,对模式的统计特性进行观测,并采用统计判别的分类器,分析归属概率的大小,按照某种方法如在分类错误发生的概率最小的前提下进行分类。贝叶斯决策理论和方法正是就上述问题展开讨论的,它是统计模式识别的基本方法。

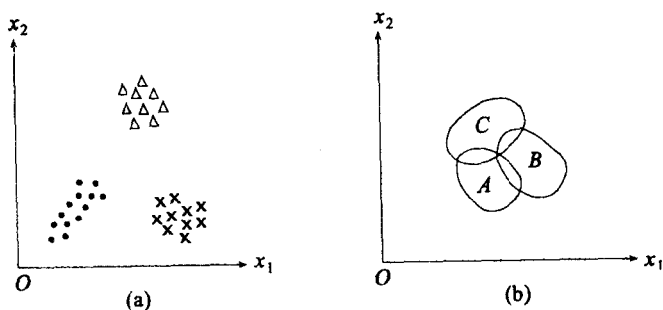


图 2-0-1

第一节 基于最小错误率的贝叶斯判别法

在两类问题中,对于某一模式样本 x ,要判断它是属于 ω_1 类,还是属于 ω_2 类,从概率

的角度讲,就是看是 $x \in \omega_1$ 的可能性大,还是 $x \in \omega_2$ 的可能性大,要比较条件概率 $p(\omega_1|x)$ 和 $p(\omega_2|x)$ 的大小。

$$\text{如果 } p(\omega_1|x) > p(\omega_2|x), \text{ 则有 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (2-1-1)$$

或者比较 ω_i 与 x 同时发生的概率的大小,即比较

$$p(\omega_i, x) = p(\omega_i|x)p(x) \quad (2-1-2)$$

的大小。

由于

$$\begin{aligned} p(\omega_i, x) &= p(x, \omega_i) = p(x|\omega_i)p(\omega_i) \\ p(\omega_i|x) &= \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \end{aligned} \quad (2-1-3)$$

这就是所谓的贝叶斯公式,将它代入式(2-1-1),若:

$$p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2), \text{ 则有 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (2-1-4)$$

或写成:

$$\text{若 } l(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{p(\omega_2)}{p(\omega_1)}, \text{ 则有 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (2-1-5)$$

这里 $l(x)$ 称为似然比, $Q = p(\omega_2)/p(\omega_1)$ 称为似然比的判决阈值。若对(2-1-5)式取自然对数的负值,则有:

$$h(x) = -\ln[l(x)] = -\ln[p(x|\omega_1)] + \ln[p(x|\omega_2)]$$

$$\text{若 } h(x) \leq \ln\left[\frac{p(\omega_1)}{p(\omega_2)}\right], \text{ 则有 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (2-1-6)$$

由式(2-1-4)可见,在 m 类问题中,对于 ω_i 类的判别函数为

$$d_i(x) = p(x|\omega_i)p(\omega_i), \quad i=1,2,\dots,m \quad (2-1-7)$$

当 $d_i(x) > d_j(x)$, $j \neq i$ 时, $x \in \omega_i$ 。

按贝叶斯公式,还可以有:

$$d_i(x) = p(\omega_i|x) \quad (2-1-8)$$

当 $d_i(x) = \max[p(\omega_j|x)], j=1,2,\dots,m$ 时, $x \in \omega_i$ 。

所以,由贝叶斯公式可以有几种形式的判别法则,针对具体问题可以选取合适的形式。不管选取何种形式,其基本思想均是要求判别归属时依概率最大作出决策,这样的结果就是分类的错误率最小。为了证明这一点,我们以两类问题中的一维情况为例,其结果不难推广到多维。

首先,所谓错误率是指平均错误率,以 $p(e)$ 表示,其定义为:

$$p(e) = \int_{-\infty}^{\infty} p(e, x) dx = \int_{-\infty}^{\infty} p(e|x)p(x) dx \quad (2-1-9)$$

表示在整个多维特征空间上的积分。

对两类问题,由式(2-1-1)可知,如果 $p(\omega_1|x) > p(\omega_2|x)$, 则有 $x \in \omega_1$, 显然,如果此时作出 $x \in \omega_2$ 的决策就是错误的,其条件错误概率为 $p(\omega_2|x)$, 反之,应为 $p(\omega_1|x)$ 。所以,

$$p(e|x) = \begin{cases} p(\omega_1|x), p(\omega_2|x) > p(\omega_1|x) \\ p(\omega_2|x), p(\omega_1|x) > p(\omega_2|x) \end{cases} \quad (2-1-10)$$

如果令 t 为两类的分界面, 且在特征向量 x 是一维时, t 为 x 轴上的一个点, 它将 x 轴分为两个区域 R_1 和 R_2 。 R_1 为 $(-\infty, t)$, R_2 为 (t, ∞) , 这样就有:

$$\begin{aligned} p(e) &= \int_{-\infty}^t p(\omega_2|x)p(x)dx + \int_t^{\infty} p(\omega_1|x)p(x)dx \\ &= \int_{-\infty}^t p(x|\omega_2)p(\omega_2)dx + \int_t^{\infty} p(x|\omega_1)p(\omega_1)dx \end{aligned} \quad (2-1-11)$$

此式还可以写为:

$$\begin{aligned} p(e) &= p(x \in R_1, \omega_2) + p(x \in R_2, \omega_1) \\ &= p(x \in R_1 | \omega_2)p(\omega_2) + p(x \in R_2 | \omega_1)p(\omega_1) \\ &= p(\omega_2) \int_{R_1} p(x|\omega_2)dx + p(\omega_1) \int_{R_2} p(x|\omega_1)dx \\ &= p(\omega_2)p_2(e) + p(\omega_1)p_1(e) \end{aligned} \quad (2-1-12)$$

图 2-1-1 所示斜线面积即 $p(\omega_2)p_2(e)$, 纹线面积即 $p(\omega_1)p_1(e)$, 这两部分之和为 $p(e)$ 。以上讨论可以推广到多维特征空间。

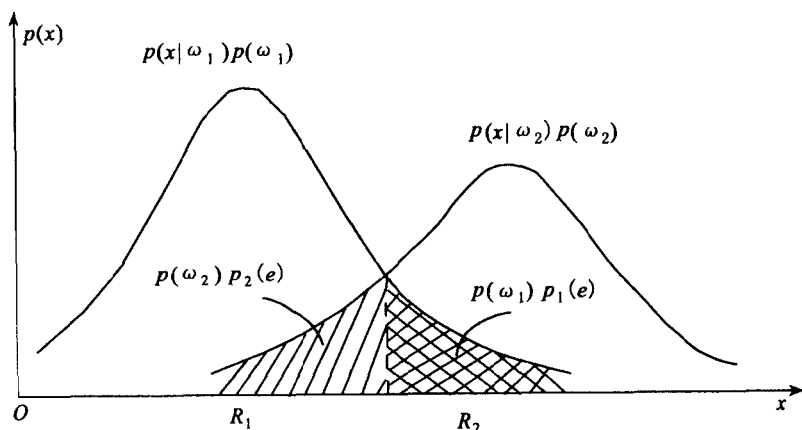


图 2-1-1

按式(2-1-1)或式(2-1-8)作决策, 实际上是对每个 x 都使 $p(e|x)$ 取值最小, 这样式(2-1-9)的积分必然达到最小, 即平均错误率 $p(e)$ 达到最小。这就证明了贝叶斯决策确实使错误率最小。

对于多类决策问题, 也可以写出相应的最小错误率贝叶斯决策规则, 即式(2-1-7)与式(2-1-8)。多类别决策要把特征空间分割成 $R_1, R_2, \dots, R_m$ 个区域, 可能错分的情况很多, 平均错误概率 $p(e)$ 由 $m(m-1)$ 项组成, 即

$$\begin{aligned} p(e) &= [p(x \in R_2 | \omega_1) + p(x \in R_3 | \omega_1) + \dots + p(x \in R_m | \omega_1)]p(\omega_1) \\ &\quad + [p(x \in R_1 | \omega_2) + p(x \in R_3 | \omega_2) + \dots + p(x \in R_m | \omega_2)]p(\omega_2) \end{aligned}$$

$$\begin{aligned}
 & + \cdots + [p(x \in R_1 | \omega_m) + p(x \in R_2 | \omega_m) + \cdots + p(x \in R_{m-1} | \omega_m)]p(\omega_m) \\
 & = \sum_{i=1}^m \sum_{j=1, j \neq i}^m [p(x \in R_j | \omega_i)]p(\omega_i)
 \end{aligned} \tag{2-1-13}$$

显然,直接求 $p(e)$ 的计算量较大。如果先计算平均正确分类概率 $p(c)$,则有:

$$p(c) = \sum_{j=1}^m p(x \in R_j | \omega_j) p(\omega_j) = \sum_{j=1}^m \int_{R_j} p(x | \omega_j) p(\omega_j) dx \tag{2-1-14}$$

这里仅对 m 项求和,而 $p(e) = 1 - p(c)$,显然比直接计算 $p(e)$ 简单得多。

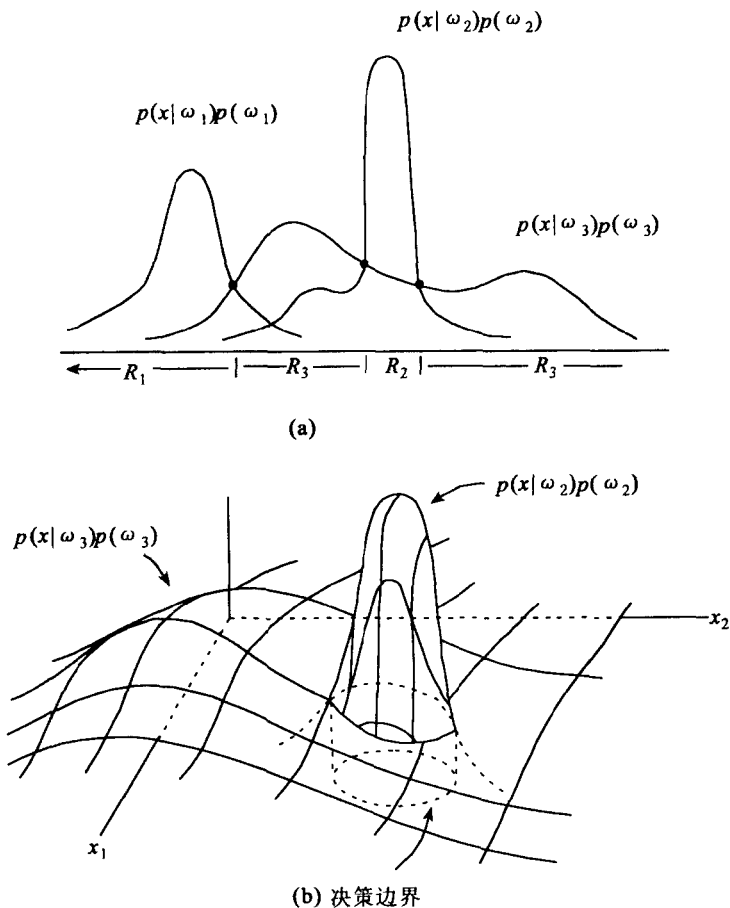


图 2-1-2

上面介绍了最小错误贝叶斯决策规则的几种等价形式,下面以这几种形式说明判别函数与决策面或界面的概念。

对于 m 类问题,按照决策规则可以把多维特征空间分成 m 个类别区域,划分这些区域的边界面称为决策面,在数学上用解析形式可以表示为决策面方程。用以描述决策规则的某种函数则称为判别函数。我们通常定义一组判别函数 $d_i(x)$, $i = 1, 2, \cdots, m$, 用于表示多

类判别规则。如果有 $d_i(x) > d_j(x)$, $j \neq i$, 则将 x 归于 ω_i 类, 如式(2-1-7), 即为贝叶斯决策的一种判别函数。对于其他几种形式, 也不难写出其判别函数。

显然, 各类别在多维特征空间中为决策面或界面所分割。这些决策面是特征空间中的超曲面。相邻的两个类别在决策面上的判别函数值是相等的, 如图 2-1-2 所示。如果 ω_i 和 ω_j 是相邻的, 则分割它们的决策面就应为:

$$d_i(x) = d_j(x) \text{ 或 } d_i(x) - d_j(x) = 0 \tag{2-1-15}$$

对于两类问题, 由式(2-1-7), 决策面方程即

$$p(x|\omega_1)p(\omega_1) - p(x|\omega_2)p(\omega_2) = 0 \tag{2-1-16}$$

第二节 基于贝叶斯公式的几种判别规则

前节所述的判别规则是直接以贝叶斯公式的各种形式为依据的, 下面介绍在此基础上
的几种判别规则或决策规则。

一、基于最小风险的贝叶斯决策

前节介绍了基于最小错误率的决策方法。在某些情况下, 引入风险的概念, 以求风险最小的决策则更为合理。例如对癌细胞的识别, 风险的概念比错误率似乎更恰当。因为识别的正确与否, 直接关系到病人的身体甚至生命。风险的概念常与损失相联系, 所以, 在下面的讨论中, 我们引入损失函数 $\lambda(\alpha_i, \omega_j)$, $i = 1, 2, \dots, a, j = 1, 2, \dots, m$ 。这个函数表示当处于状态 ω_j 时采取决策为 α_i 所带来的损失。在决策论中, 常以决策表一目了然地表示各种情况下的决策损失, 如表 2-2-1 所示。下面在已知先验概率 $p(\omega_j)$ 及类条件概率密度 $p(x|\omega_j)$, $j = 1, 2, \dots, m$ 的条件下进行讨论。

表 2-2-1

损 失 决 策	状 态	ω_1	ω_2	...	ω_j	...	ω_m
α_1		$\lambda(\alpha_1, \omega_1)$	$\lambda(\alpha_1, \omega_2)$	...	$\lambda(\alpha_1, \omega_j)$	...	$\lambda(\alpha_1, \omega_m)$
α_2		$\lambda(\alpha_2, \omega_1)$	$\lambda(\alpha_2, \omega_2)$	...	$\lambda(\alpha_2, \omega_j)$	...	$\lambda(\alpha_2, \omega_m)$
$\vdots$		$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
α_i		$\lambda(\alpha_i, \omega_1)$	$\lambda(\alpha_i, \omega_2)$	...	$\lambda(\alpha_i, \omega_j)$	...	$\lambda(\alpha_i, \omega_m)$
$\vdots$		$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
α_a		$\lambda(\alpha_a, \omega_1)$	$\lambda(\alpha_a, \omega_2)$	...	$\lambda(\alpha_a, \omega_j)$	...	$\lambda(\alpha_a, \omega_m)$

根据贝叶斯公式,后验概率为:

$$p(\omega_j | x) = \frac{p(x | \omega_j) p(\omega_j)}{p(x)} \quad (2-2-1)$$

式中,

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

当引入“损失”的概念后,考虑错判所造成的损失时,就不能只根据后验概率的大小来作决策,而必须考虑所采取的决策是否使损失最小。对于给定的 x ,如果采取决策 α_i ,从决策表中可见,对应于决策 α_i , λ 可以在 m 个 $\lambda(\alpha_i, \omega_j)$, $j=1,2,\dots,m$ 当中任取一个,其相应概率为 $p(\omega_j | x)$ 。因此在采取决策 α_i 情况下的条件期望损失 $R(\alpha_i | x)$ 为:

$$R(\alpha_i | x) = E[\lambda(\alpha_i, \omega_j)] = \sum_{j=1}^m \lambda(\alpha_i, \omega_j) p(\omega_j | x), i = 1, 2, \dots, a \quad (2-2-3)$$

在决策论中又把采取决策 α_i 的条件期望损失 $R(\alpha_i | x)$ 称为条件风险。由于 x 是随机向量的观察值,对于 x 的不同观察值,采取决策 α_i 时,其条件风险的大小是不同的。所以,究竟采取哪一种决策将随 x 的取值而定。决策 α 可以看成随机向量 x 的函数,记为 $\alpha(x)$,于是我们可以定义期望风险 R 为:

$$R = \int R(\alpha(x) | x) p(x) dx \quad (2-2-4)$$

式中, dx 是特征空间的体积元,积分在整个特征空间进行。

期望风险 R 反映对整个特征空间所有 x 的取值都采取相应的决策 $\alpha(x)$ 所带来的平均风险;而条件风险 $R(\alpha_i | x)$ 只是反映了对某一 x 的取值采取决策 α_i 所带来的风险。显然,需采取一系列决策 $\alpha(x)$ 使期望风险 R 最小。

在考虑错判带来的损失时,我们希望损失最小。如果在采取每一个决策或行动时,都使其条件风险最小,则对所有的 x 作出决策时,其期望风险也必然最小,这样的决策就是最小风险贝叶斯决策。

最小风险贝叶斯决策规则为:

如果

$$R(\alpha_k | x) = \min_{i=1,2,\dots,a} R(\alpha_i | x), \text{ 则有 } \alpha = \alpha_k \quad (2-2-5)$$

对于实际问题,最小风险贝叶斯决策可按下列步骤进行:

(1) 在已知 $p(\omega_j)$, $p(x | \omega_j)$, $j=1,2,\dots,m$, 并给出待识别的 x 的情况下,根据贝叶斯公式可以计算出后验概率:

$$p(\omega_j | x) = \frac{p(x | \omega_j) p(\omega_j)}{\sum_{i=1}^m p(x | \omega_i) p(\omega_i)}, \quad j = 1, 2, \dots, m \quad (2-2-6)$$

(2) 利用计算出的后验概率及决策表,按式(2-2-3)计算出采取 α_i , $i=1,2,\dots,a$ 的条件风险 $R(\alpha_i | x)$ 。

(3) 对步骤(2)中得到的 a 个条件风险值 $R(\alpha_i | x)$, $i=1,2,\dots,a$, 进行比较,找出使条