高等学校教材

模式识别

蔡元龙 编

西北电讯工程学院出版社

高等学校教材

模式识别

蔡元龙 编

西北电讯工程学院出版社

1986

内 容 简 介

本书分统计模式识别和句法(结构)模式识别两大部分。前者包括聚类分析；判别函数和可训练的确定性分类器；统计判别及其可训练的模式分类器；特征选择和提取四章。后者为句法模式识别一章。此外，在导引一章有必需的概率统计知识。全书有较多的数值计算例题，各章都有习题，并有上机作业。附录中的模式样本数据系供上机练习之用。

本书系模式识别的基础，介绍了目前采用较多的分类方法。书中注意物理概念的说明和具体分类算法的实现，数学推导则力求简洁。适用于工科院校高年级学生、研究生和有关专业的科技工程人员使用。

高等学校教材

模式识别

蔡元龙 编

责任编辑 夏大平

西北电讯工程学院出版社出版

西北电讯工程学院印刷厂印刷

陕西省新华书店发行 各地新华书店经售

开本 787×1092 1/16 印张 12 8/16 字数 303 千字

1986 年 6 月第一版 1986 年 6 月第一次印刷 印数 1-6,000

统一书号：15322·47 定价：2.10 元

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校工科电子类专业课教材的编审、出版的组织工作。从一九七七年底到一九八二年初，由于各有关院校，特别是参与编审工作的广大教师的努力和有关出版社的紧密配合，共编审出版了教材159种。

为了使工科电子类专业教材能更好地适应社会主义现代化建设培养人才的需要，反映国内外电子科学技术水平，达到"打好基础、精选内容、逐步更新、利于教学"的要求，在总结第一轮教材编审出版工作经验的基础上，电子工业部于一九八二年先后成立了高等学校《无线电技术与信息系统》、《电磁场与微波技术》、《电子材料与固体器件》、《电子物理与器件》、《电子机械》、《计算机与自动控制》，中等专业学校《电子类专业》、《电子机械类专业》共八个教材编审委员会，作为教材工作方面的一个经常性的业务指导机构，并制定了一九八二～一九八五年教材编审出版规划，列入规划的教材、教学参考书、实验指导书等共217种选题。在努力提高教材质量，适当增加教材品种的思想指导下，这一批教材的编审工作由编审委员会直接组织进行。

这一批教材的书稿，主要是从通过教学实践、师生反映较好的讲义中评选择优和从第一轮较好的教材中修编产生出来的。广大编审者，各编审委员会和有关出版社都为保证和提高教材质量作出了努力。

这一批教材，分别由电子工业出版社、国防工业出版社、上海科学技术出版社、西北电讯工程学院出版社、湖南科学技术出版社、江苏科学技术出版社、黑龙江科学技术出版社和天津科学技术出版社承担出版工作。

限于水平和经验，这一批教材的编审出版工作肯定还会有许多缺点和不足之处，希望使用教材的单位，广大教师和同学积极提出批评建议，共同为提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前　言

本教材由无线电技术与信息系统教材编审委员会电路与系统编审小组评选审定并推荐出版。

该教材由西安交通大学蔡元龙教授编写，上海交通大学李介谷教授主审。编审者均依据电路与系统编审小组审定的编写大纲进行了编写和审阅。

本课程的参考教学时数课内(外)为40(80)学时，其内容包括统计模式识别和句法(结构)模式识别两大部分。前者有非监督分类的聚类分析方法，以及监督分类中的感知器算法、贝叶斯分类器、参数估计和势函数法等，这些都是常用的基本分类方法。为简化分类器，专用一章讨论了特征的选择和提取，分析了采用K-L变换进行特征提取的方法。在导引一章还归纳了本书中必需的概率统计知识。句法模式识别部分，在简单介绍集合论中关系运算和形式语言理论的基础上，阐述了句法识别和自动机识别，然后针对图形结构识别，逐一讨论了图形的基元提取、采用形式语言的图形识别方法、句法分析和文法推断。全书例题较多，各章都附有参考书目和习题，并有上机作业，附录中的模式样本数据，就是供统计模式识别上机练习之用的。使用本教材应先修概率统计和线性代数；为能进行上机作业，亦必须掌握FORTRAN或BASIC语言。考虑到许多学生常常也选修《数字图象处理》课程，最好能结合该课进行一个大型课程作业。例如利用遥感图象进行特征提取和统计分类，利用医学或工业上的图形来提取基元并用句法进行结构识别分类，这对提高学生的独立工作能力大有好处。

由于编者水平有限，加以该学科发展迅速，书中难免还存在一些缺点和错误，殷切希望广大读者批评指正。

编　者

1985年2月

此为试读，需要完整PDF请访问：www.ertongbook.com

目　录

导　引

模式识别(Pattern Recognition)就是利用计算机对某些物理对象进行分类，在错误概率最小的条件下，使识别的结果尽量与客观事物相符。这里的"模式"有广泛的含义，它或者是图形、波形，或者是不同的疾病，或者是各种动植物的类别，或者是不同成份的矿石，总之它包括自然界中各种各样需要识别的对象。

§0.1　模式识别系统的组成

将模式识别用于医学的一个例子，就是病症的计算机数值诊断。医生看病，目的是要看出病人患什么病。他首先从病人获取临床表现，比如要测量体温、血压，询问病人有什么不舒服，例如是否有胃痛、腹泻等，然后根据这些临床表现逐步缩小诊断范围，再通过有目的地化验某几项指标，经过头脑的综合分析和根据自己的学识经验，进行判断，最后给出诊断的病名。显然，诊断的准确性与医生的学识、专业、经验有着直接的关系。

计算机数值诊断与医生的临床诊断过程相仿，它也是在接受病人临床表现的基础上，将临床的检验结果数值化，并按给定的临床表现与身体器官状态变量之间的数学模型，对这些数据进行加工，再经过反复试验，不断接受新的病例资料，通过学习，不断修改诊断模型。所以计算机有能力接受大量的过去病例资料，综合各家医生的经验，定量分析各种临床表现，并有可能从提供的临床表现中获取比较多的信息。尽管医生要赋予计算机这些能力也得进行大量工作，但医生可利用计算机提高诊断的效果。

因此，模式识别的过程大致如图 0.1-1 的框图所示，虚线下部是学习训练过程，上部是识别过程。在学习过程中，将已知的模式样本(例如已经确诊的病例资料)进行"数值化"后，

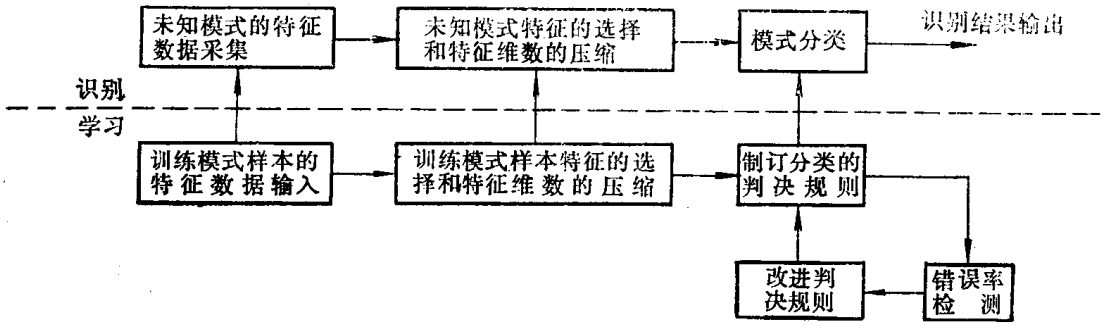

图 0.1-1　模式识别的过程

送入计算机，经过分析，去掉对分类无效或反而易造成混淆的那些特征，尽量保留对分类判别特别有效的数值特征(医学上常称之为特异性的指标)，这就是所谓特征选择。有时，还采用某种变换技术，得出数目上比原来少的综合性特征作为分类用(这里称之为特征维数压缩，习惯上亦称特征提取)，然后按设想的分类判决的数学模型进行分类，并将分类结果与已知类别的输入模式作对比，不断修改，制定出错误率最小的判别规则。学习过程中所制定的数学模型是用来对未知模式分类之用的。

图象(或图形)的识别与上述过程相同，只是输入的是两维图象。例如显微镜下的染色体涂片的放大图象，它们先经电视摄象机转换成图象的电信号，再经过数字化成为具有 $n \times m$ 个象元的空间排列矩阵，每一象元的数值代表该点上的亮度值(或称灰阶值)。数字化图象的原始数据通常先经过预处理，如改善图象质量，去除噪声，增强边缘轮廓等等，然后将图象中感兴趣的对象分割出来(这里就是将染色体图象从背景中分割出来)，并对该对象进行各项测量，例如几何形状、尺寸大小、灰阶数值变化或纹理参数等。凡是能表征被测对象的主要性质、对分类识别有用的参数都可测量，它们就是模式的数值特征，用作识别的输入。上述这个处理过程要用数字图象处理技术来完成。

通常这些测量得的特征参数包含有多余的信息，同样需要经过特征选择或特征维数的压缩，才进入分类器进行分类判别，它与图 0.1-1 中所示的过程相同。

本书主要讨论特征选择、维数压缩和各种分类方法，这也是通常的"模式识别"的内容。至于图象的预处理，则在图象处理中讨论。关于各种特征数值的测量，更是种类繁多、方法不一。如医学上有医生的各种化验检查的手段，地质上有其化学和物理的分析方法，等等，它们不是一门课程所能解决的。本书只是讨论在已经获得模式的数值特征之后怎样进行识别的问题。

§0.2 模式识别的方法

(1) 统计模式识别和句法模式识别

模式识别所采用的处理方法分统计方法和句法方法。我们先看两个简单例子：

例 1 19 名男女进行体检，测量了身高和体重，但事后发现其中有 4 人忘了填写性别，试问(在最小错误的条件下)这 4 人是男是女？体检的数值如表 0-1 所示。

表 0-1

编 号	身高(cm)	体重(kg)	性 别	编 号	身高(cm)	体重(kg)	性 别
1	170	68	男	11	140	62	男
2	130	66	女	12	150	64	女
3	180	71	男	13	120	66	女
4	190	73	男	14	150	66	男
5	160	70	女	15	130	65	男
6	150	66	男	α	140	70	?
7	190	68	男	β	150	60	?
8	210	76	男	γ	145	65	?
9	100	58	女	δ	160	75	?
10	170	75	男				

这里待识别的模式是男或女性别，测量的特征是身高和体重，15 名已知性别的样本特征作为"训练样本"，希望藉助他们的特征建立判别函数(即数学模型)，对未知的模式进行分类。为直观起见，绘成图 0.2-1 所示的特征空间分布图。从图上训练样本的分布情况，发现男、女的两类特征各有其聚类的特点，我们有可能找出一个判别函数(直线或曲线)，这样，只要给出待分类的模式特征的数值，看它在特征平面上落在判别函数的哪一侧，就可判别它是男还是女了。

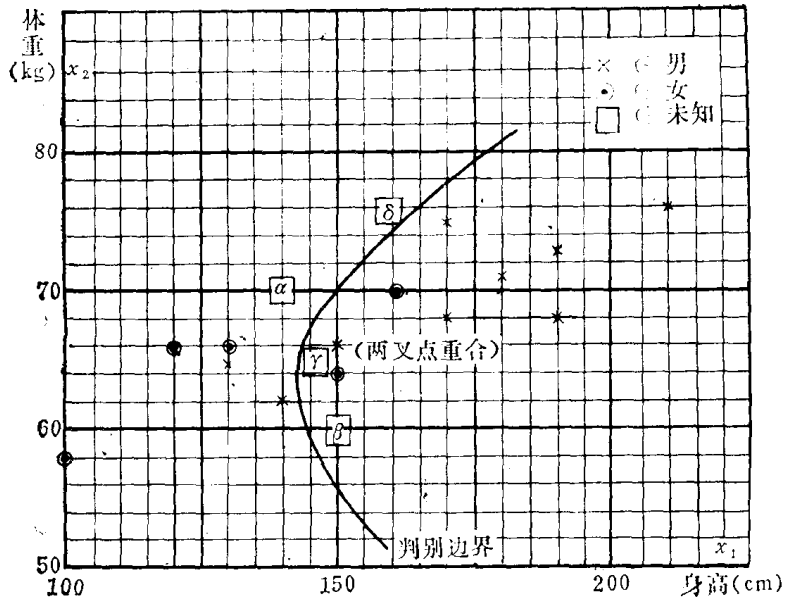

图 0.2-1　模式样本在特征空间中的分布图

这个问题可采用概率统计的方法来识别，求出判别函数式

$$d(x_1, x_2) = 0$$

若上表中 α 的特征值使 $d(x_1, x_2)<0$，则判决 α 为女；若 β 特征值使 $d(x_1, x_2)>0$，则判决 β 为男。为了判别的计算简单，常采用线性或二次型的判别函数式(图 0.2-1 所示为二次曲线)；为使误判概率最小，常采用统计方法选择最优化的判别函数。

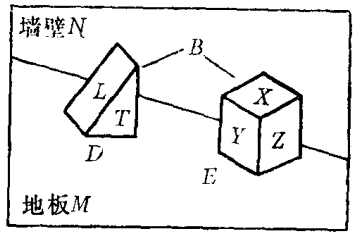

图 0.2-2　具有三角体和长方体的图片

例 2　识别图 0.2-2 所示的图片，需要对图象的结构信息作描述。因图象结构比较复杂，就设想用一些较简单的子图象组合来表示，而子图象又用一些更为简单的基本图形单元来表示，且所有这些基元又是按一定的结构关系来组织，这就可采用多级树结构来进行描述，如图 0.2-3 所示。这种描述是利用了形式语言理论的技巧，例如英文句子“This

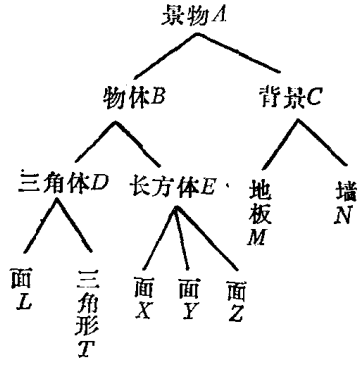

图 0.2-3　图0.2-2 中图片的结构描述

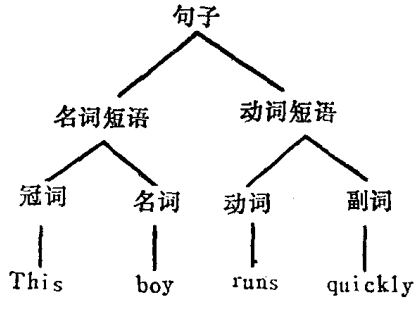

图 0.2-4　简单英文句子的句法描述

boy runs quickly”的句法亦可描述成多级树状结构，如图0.2-4。该句子可用比较简单的名词短语和动词短语来描述，而它们可用一些更简单的词(这里有冠词、名词、动词和副词)作

为基元，按一定的句法规则组合起来。这些词好比树的叶子，按树的结构汇集起来，就形成这个句子。

对照上述的场景，如果用已知结构信息的图象作为训练样本，先识别出基元(就是图0.2-2中的X、Y、Z等简单的面)和它们之间的连接关系(例如长方体E是由X、Y和Z三个面拼接而成)，并用字母符号代表之，然后用构造句子的文法来描述生成这幅场景的过程，由此推断出生成该场景的一种文法规则，这就是训练过程。在识别过程中，同样要对未知结构信息的图象进行基元识别及其相互结构关系的分析，然后用训练过程获得的文法作句法分析。如果它能被已知结构信息的文法分析出来，则该幅未知图象具有相同的图形结构，否则就不是这种结构。这就是所谓模式识别的句法(结构)方法。因为诸如景物分析这类问题所研究的模式通常十分复杂，需要特征数目也很多，所以用一些比较简单的子模式组成多级结构来描述一个复杂的模式就成为一种可取的想法，但只能在获得预期的描述文法结构之后，才能将它指定到一种特定的类别。

句法方法和统计方法在模式识别的许多方面是互相补充的，模式识别开始是从统计方法发展起来的，而句法方法更扩大了模式识别的能力，特别是用于景物分析和结构识别。

(2) 监督分类与非监督分类

监督分类与非监督分类亦称为有人管理与无人管理的分类。

从图0.2-1的例子中可看出，如果模式样本的特征有n个，一个样本构成一个n维的特征向量，它在n维特征空间中是对应于一个点。分类问题就是把特征空间分割成对应于不同类别的互不相容区域，每一区域对应于一个特定的模式类，不同类别之间的界面，在数学上表示成"判别函数"。

监督分类，要依靠已知所属类别的训练样本集，按它们特征向量的分布来确定判别函数，只有在判别函数确定之后才能用它来对未知的模式进行分类判别。这就要求我们对于要分类的图象，有足够的先验知识，而要做到这一点，往往要付出相当大的代价。例如医学的计算机数值诊断，要采集足够数量的病人临床诊断资料，并且应是完全确诊的，还要有典型性。再如利用遥感图象作农业估产，则要指定若干个训练区，请专业人员先到现场进行估产，有了这个标准才能对图象的其它区域进行计算机估产。

在没有先验知识的情况下可用非监督分类，通常采用聚类分析方法，它基于"物以类聚"，的观点，用数学方法分析各特征向量之间的距离及分散情况。这些特征向量集可能聚集成若干个群，按群间距离远近划分成类。这种按各类之间的亲疏程度的划分，如事先能知道应划分成几类，则更为合适。

(3) 统计识别中的参数和非参数方法

统计模式识别是以模式集在特征空间中分布的类概率密度函数为基础。当模式样本的类概率密度函数的形式是已知的，或者从提供的作为设计分类器用的训练样本能估出类概率密度函数的近似式，则在具体计算中，密度函数的未知参数就可用合适的估计量来取代。例如在多数场合，常用多维的正态分布来近似类密度分布，这不单是符合自然界多数分布的情况，也是为使分析计算比较简单，因正态分布用均值向量和协方差矩阵就可完全代表。计算判别函数只需用估计参数来运算，即将估计类概率密度函数的问题简化成估计密度函数的参数问题，这样运算就大为简化。

参数估计是数理统计中的经典问题，最常用的有最大贝叶斯估计和最大似然估计。

此为试读，需要完整PDF请访问：www.ertongbook.com

假如类概率密度函数不知，或者所提供的训练样本的数目不足以估计出其概率密度函数，这时就不可能通过密度函数的参数估计方法来计算判别函数，而只能借助于非参数方法。

非参数方法有多种形式，且各不相关。常用的有 k-最近邻判决规则，它直接按 k 个最近邻样本的不同类别分布，将未知类别的特征向量分类。也可采用直接确定判别函数的形式，例如确定判别函数是线性方程或二次多项式等，再利用训练样本的特征向量值直接计算判别函数的系数。还有其它许多方法，这里不一一列举。总之，凡是不采用某种概率密度函数的估计参数来确定判别函数的，都可归之为非参数方法。

用参数或非参数方法来确定判别函数，都可采用逐次估计的方式进行，常称为训练或学习。这时，每提供一个训练样本，便教分类器学习一次，使估计的参数或判别函数的系数更为准确一点。显然，开始训练时，可先用一个很粗的初始猜测来近似，然后依次用每一个样本观察值来修改前一次的估计。这种情况，往往需要首先分析一下所要采用的估计量，看它能否在某种意义上收敛于真实参数且其收敛是否足够快。

模式识别这门学科需要用到较多的概率统计及线性代数方面的知识，下面将扼要地复习一下有关的内容，如读者已较熟悉，则可跳过。

§0.3 随机向量及其分布

(1) 随机向量

如果一个对象的特征观察值为$\{x_1, x_2, \cdots, x_n\}$，它可构成一个 n 维的特征向量值 $\boldsymbol{x}$，即

$$\boldsymbol{x}=(x_1 \quad x_2 \quad \cdots \quad x_n)'$$

式中，$x_1, x_2, \cdots, x_n$ 为特征向量 $\boldsymbol{x}$ 的各个分量。因此，一个特征可以看作 n 维空间中的向量或点，此空间就称为模式的特征空间 R_n。

在模式识别过程中，要对许多具体对象进行测量，这样便可获得许多次观察值。每次观察值不一定相同，所以对许多对象而言，它们各个特征分量都是随机变量，即许多对象的特征向量在 n 维空间中亦呈随机性分布，故名为随机向量。本章以 X_i 表示随机变量，$\boldsymbol{X}$ 表示随机向量。

(2) 分布函数

随机向量 $\boldsymbol{X}$ 的联合分布函数 F 可用概率 P 来定义，即

$$F(x_1, x_2, \cdots, x_n)=P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}$$

或写成向量形式 即

$$F(\boldsymbol{x})=P(\boldsymbol{X} \leqslant \boldsymbol{x})$$

随机向量的联合概率密度函数定义为

$$p(\boldsymbol{x})=\lim_{\substack{\Delta x_1 \to 0 \\ \vdots \\ \Delta x_n \to 0}} \frac{P\{x_1 < X_1 \leqslant x_1+\Delta x_1, \cdots, x_n < X_n \leqslant x_n+\Delta x_n\}}{\Delta x_1 \cdots \Delta x_n}$$

$$=\frac{\partial^n F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n}$$

(3) 给定类别的条件密度函数

模式识别中，设某个模式集合 Ω 由 m 类模式所组成，将其中第 i 类模式记为 ω_i。对属于给定类别 ω_i 的模式特征有其自己的概率密度函数，称为类密度函数。对给定类别的条件分布函数与密度函数分别定义为

$$F(\boldsymbol{x}|\omega_i)=P\{\boldsymbol{X}\leqslant\boldsymbol{x}|\omega_i\}$$

$$p(\boldsymbol{x}|\omega_i)=\frac{\partial^n F(\boldsymbol{x}|\omega_i)}{\partial x_1\cdots\partial x_n}$$

(4) 随机向量的参数

n 维随机向量 $\boldsymbol{X}$ 的数学期望(或称均值向量) $\boldsymbol{m}$ 定义为

$$\boldsymbol{m}=E(\boldsymbol{X})=\int_{R_n}\boldsymbol{x}p(\boldsymbol{x})d\boldsymbol{x}$$

式中 R_n 包括整个 $\boldsymbol{X}$ 的样本空间。$\boldsymbol{m}$ 的第 i 个分量 m_i 代表第 i 维的随机变量 X_i 的平均值，它为

$$\begin{aligned}m_i=E(X_i)&=\int_{R_n}x_i p(\boldsymbol{X})d(\boldsymbol{X})\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_i p(x_1,\ \cdots,\ x_n)dx_1\cdots dx_n\\&=\int_{-\infty}^{\infty}x_i p(x_i)dx_i\end{aligned}$$

式中 $p(x_i)$ 是 $\boldsymbol{X}$ 的第 i 个分量的边缘密度，即

$$p(x_i)=\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{(n-1)\text{重积分}}p(x_1,\ \cdots,\ x_{i-1},\ x_{i+1},\ \cdots,\ x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n$$

因此，$\boldsymbol{m}$ 的每个分量可按对应的一维边缘密度的各个变量的期望值来计算。

随机向量 $\boldsymbol{X}$ 的条件期望向量定义为

$$\boldsymbol{m}=E(\boldsymbol{X}|\omega_i)=\int_{\boldsymbol{x}\in\omega_i}p(\boldsymbol{x}|\omega_i)d\boldsymbol{x}$$

协方差矩阵与自相关矩阵：

协方差阵说明随机向量 $\boldsymbol{X}$ 的各分量的分散情况，它的定义为

$$\begin{aligned}C&=E\{(\boldsymbol{X}-\boldsymbol{m})(\boldsymbol{X}-\boldsymbol{m})^t\}\\&=E\begin{bmatrix}(X_1-m_1)\\\vdots\\(X_n-m_n)\end{bmatrix}[(X_1-m_1)\cdots(X_n-m_n)]\\&=\begin{bmatrix}E(X_1-m_1)(X_1-m_1)&\cdots&E(X_1-m_1)(X_n-m_n)\\\vdots&&\vdots\\E(X_n-m_n)(X_1-m_1)&\cdots&E(X_n-m_n)(X_n-m_n)\end{bmatrix}\\&=\begin{bmatrix}\lambda_{11}&\cdots&\lambda_{1n}\\\vdots&&\vdots\\\lambda_{n1}&\cdots&\lambda_{nn}\end{bmatrix}\end{aligned}$$

式中，协方差阵诸分量为

$$\lambda_{ij}=E[(X_i-m_i)(X_j-m_j)]$$
$$=\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n\text{重积分}}(x_i-m_i)(x_j-m_j)p(x_1,\ \cdots,\ x_n)dx_1\cdots dx_n$$

上式中，如 $i\neq j$，λ_{ij} 便是 $\boldsymbol{X}$ 的第 i 分量与第 j 分量的协方差；如 $i=j$，λ_{ii} 便是随机变量 X_i 的方差，是协方差阵的对角分量。可以看出，协方差矩阵都是对称阵。

表示随机向量$\boldsymbol{X}$的分散度还有自相关矩阵，用 S 表示，它定义为

$$S=E(\boldsymbol{X}\boldsymbol{X}^t)$$

协方差阵与自相关阵之间的关系为

$$C=E[(\boldsymbol{X}-\boldsymbol{m})(\boldsymbol{X}-\boldsymbol{m})^t]$$
$$=E[\boldsymbol{X}\boldsymbol{X}^t-\boldsymbol{m}\boldsymbol{X}^t-\boldsymbol{X}\boldsymbol{m}^t+\boldsymbol{m}\boldsymbol{m}^t]$$
$$=S-\boldsymbol{m}\boldsymbol{m}^t$$

将协方差阵各协方差分量变为相关系数，可将它归一化。令$\lambda_{ij}=\sigma_{ij}^2$，$(i=1,\ \cdots,n)$，则第 i 与第 j 分量间的相关系数为

$$r_{ij}=\frac{\lambda_{ij}}{\sigma_{ii}\sigma_{jj}},\quad i,\ j=1,\ \cdots,\ n$$

相关(系数)矩阵为

$$R=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{nn} \end{pmatrix}$$

(5) 协方差阵的正定性

协方差阵与自相关阵都是对称阵。对于任意随机向量 $\boldsymbol{X}$，$\boldsymbol{X}^tC\boldsymbol{X}$是 C 的二次型。如果

$$\boldsymbol{X}^tC\boldsymbol{X}\geqslant 0$$

并对一切$\boldsymbol{X}$都成立，则称 C 为非负定阵。如果 C 是非负定的，而且$\boldsymbol{X}^tC\boldsymbol{X}=0$的必要条件是 $\boldsymbol{X}=0$，则 C 为正定阵，即 C 不仅是非负定，而且 C 的行列式$|C|\neq 0$。

(6) 随机向量的不相关、正交、独立的条件

如果随机向量的第 i 分量 X_i 与第 j 分量 X_j 不相关，则

$$\lambda_{ij}=0,\quad i,j=1,\ \cdots,\ n$$

但因 $\lambda_{ij}=E(X_iX_j)-m_im_j=0$

所以 $E(X_iX_j)=m_im_j$

即 $E(X_iX_j)=E(X_i)E(X_j)$

随机向量$\boldsymbol{X}$和$\boldsymbol{Y}$不相关的条件也可以是

$$E(\boldsymbol{X}^t\boldsymbol{Y})=E(\boldsymbol{X}^t)E(\boldsymbol{Y})$$

及

$$E(\boldsymbol{X}\boldsymbol{Y}^t)=E(\boldsymbol{X})E(\boldsymbol{Y}^t)$$

这时，$\boldsymbol{X}$和$\boldsymbol{Y}$的协方差阵为零矩阵。

随机向量$\boldsymbol{X}$与$\boldsymbol{Y}$的内积符合 $E(\boldsymbol{X}^t\boldsymbol{Y})=0$ 的条件时，称$\boldsymbol{X}$与$\boldsymbol{Y}$正交。

随机向量$\boldsymbol{X}$与$\boldsymbol{Y}$的联合概率密度函数等于各自密度函数的相乘，即

$$p(\boldsymbol{x},\ \boldsymbol{y})=p(\boldsymbol{x})p(\boldsymbol{y})$$

时，则称随机向量$\boldsymbol{X}$与$\boldsymbol{Y}$彼此独立。

不相关、正交、独立这三个条件之间的几个重要关系为：

(i) 若$\boldsymbol{X}$与$\boldsymbol{Y}$是彼此独立的，则一定是不相关，但反过来不成立。因此独立性比不相关有更强的条件，前者是对每一随机向量$p(\boldsymbol{x},\boldsymbol{y})=p(\boldsymbol{x})p(\boldsymbol{y})$的条件都成立，而后者只是$p(\boldsymbol{x},\boldsymbol{y})$的一个总体性质。

(ii) 若$\boldsymbol{X}$或$\boldsymbol{Y}$的期望向量是零向量，则不相关等价于正交。

§0.4 多维正态分布

对许多实际的数据集，正态分布常是比较合理的近似，特别对于那种是由n个独立变量之和组成的随机变量，当n很大时，它将是正态分布。

(1) 一维正态密度函数

一维随机变量X的正态密度函数表示为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$$

式中 均值$m=E(X)=\int_{-\infty}^{\infty}xp(x)dx$；

方差$\sigma^2=E\{(X-m)^2\}$

$$=\int_{-\infty}^{\infty}(x-m)^2p(x)dx$$

σ称为标准差。一维正态分布的密度函数如图0.4-1所示。

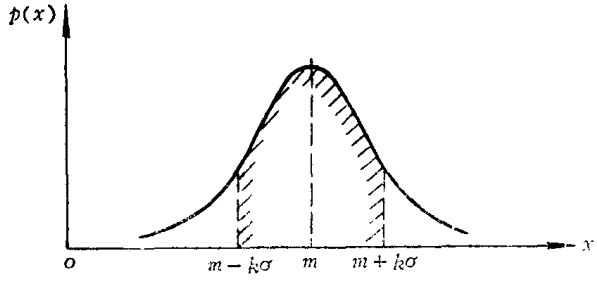

图 0.4-1 一维正态的密度函数曲线

在m的左、右各为$k\sigma$的范围内，概率为

$$P\{m-k\sigma\leqslant x\leqslant m+k\sigma\}=\int_{m-k\sigma}^{m+k\sigma}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]dx$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-k}^{+k}\exp\left[-\frac{y^2}{2}\right]dy$$

上式中令$y=\frac{x-m}{\sigma}$，则P值与k的关系为

$$P\{m-k\sigma\leqslant x\leqslant m+k\sigma\}=\begin{cases}0.683, & \text{当 } k=1\\ 0.954, & \text{当 } k=2\\ 0.997, & \text{当 } k=3\end{cases}$$

因此在区间$|x-m|\leqslant 3\sigma$内，差不多包含了全部由正态母体抽取的子样。

正态密度曲线可完全由均值m和方差σ^2所确定，因此可用下式表示：

$$p(x)\sim N(m,\ \sigma^2)$$

(2) 多维正态密度函数

n维正态随机向量的密度函数用下式：

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|C|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})'C^{-1}(\boldsymbol{x}-\boldsymbol{m})\right\}$$

或 $$p(\boldsymbol{x})\sim N(\boldsymbol{m},\ C)$$

表示。式中

$$\boldsymbol{x}=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix},\quad \boldsymbol{m}=\begin{bmatrix}m_1\\ \vdots\\ m_n\end{bmatrix},\quad C=\begin{bmatrix}\sigma_{11}^2 & \cdots & \sigma_{1n}^2\\ \vdots & & \vdots\\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2\end{bmatrix}$$

$|C|$为协方差阵C的行列式。多维正态密度函数完全由它的均值$\boldsymbol{m}$和协方差阵C所确定。

二维随机向量$\boldsymbol{X}$的正态密度分布如图0.4-2(a)所示。将其等高线投影到x_1ox_2平面上，得图(b)，图中，原点到M点的向量为$\boldsymbol{m}$。等密度点的轨迹是椭圆，椭圆中心为M，其形状由

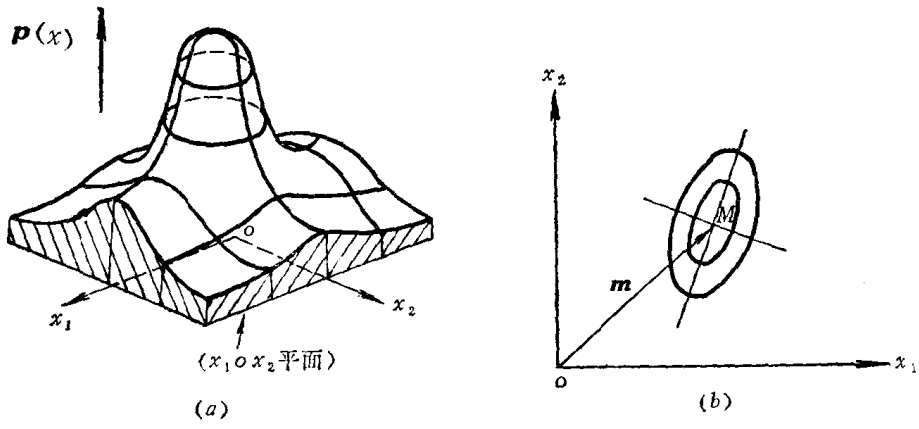

图0.4-2 二维正态密度函数图示

协方差阵C决定。椭圆的主轴方向，由C的本征向量决定，主轴的长度则由C的本征值决定。式中的C，只在$|C|\neq 0$情况(即正定阵情况)有意义，但当$\boldsymbol{X}$中的一个分量为零，或者两个分量完全相同时，$|C|=0$，这时C为非负定阵，它的逆不存在。

对于n维正态随机向量，其指数项有

$$d^2=(\boldsymbol{x}-\boldsymbol{m})'C^{-1}(\boldsymbol{x}-\boldsymbol{m})$$

称之为$\boldsymbol{X}$到$\boldsymbol{m}$的马氏(Maharanobis)距离平方，也称距离函数，它不同于欧几里德距离。

(3) 不相关-独立

正态分布的诸分量$X_i(i=1,\ 2,\ \cdots,\ n)$若互不相关，则它们也是独立的。这时

$$\sigma_{ij}^2=E\{(X_i-m_i)(X_j-m_j)\}=0,\quad i\neq j$$

因此C变为对角矩阵

$$C=\begin{bmatrix}\sigma_{11}^2 & & 0\\ & \ddots & \\ 0 & & \sigma_{nn}^2\end{bmatrix}\qquad 及\ C^{-1}=\begin{bmatrix}\dfrac{1}{\sigma_{11}^2} & & 0\\ & \ddots & \\ 0 & & \dfrac{1}{\sigma_{nn}^2}\end{bmatrix}$$

所以可得

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}\prod_{i=1}^{n}\sigma_{ii}^2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-m_i)^2}{\sigma_{ii}^2}\right\}$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}\sigma_{ii}} \exp\left\{-\frac{1}{2}\frac{(x_i - m_i)^2}{\sigma_{ii}^2}\right\} = \prod_{i=1}^{n} p(x_i)$$

即满足了独立的条件。

(4) 正态边际密度和条件密度

正态分布的边际密度和条件密度都是正态分布，可用二维正态密度说明此结论。当

$$p(x_1, x_2) = \frac{1}{2\pi|C|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t C^{-1}(\boldsymbol{x}-\boldsymbol{m})\right\}$$

则边际密度为

$$p(x_1) \sim N(m_1, \sigma_{11}^2)$$
$$p(x_2) \sim N(m_2, \sigma_{22}^2)$$

式中 $\boldsymbol{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$，$C = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}$

条件密度为

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} \sim N\left\{m_1 + \frac{\sigma_{12}^2}{\sigma_{22}^2}(x_2 - m_2), \frac{|C|}{\sigma_{22}^2}\right\}$$

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} \sim N\left\{m_2 + \frac{\sigma_{12}^2}{\sigma_{11}^2}(x_1 - m_1), \frac{|C|}{\sigma_{11}^2}\right\}$$

§0.5 随机向量的变换

(1) 密度函数的变换和雅可比行列式

设随机向量$\boldsymbol{Y}$是另一随机向量$\boldsymbol{X}$的函数，即

$$\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$$

或写成

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} g_1(X_1, \cdots, X_n) \\ \vdots \\ g_n(X_1, \cdots, X_n) \end{pmatrix}$$

并且设$\boldsymbol{X}$与$\boldsymbol{Y}$之间有一一对应关系，则这两个随机向量的密度函数之间的关系为

$$p(\boldsymbol{y}) = \frac{p(\boldsymbol{x})}{|J|}$$

式中，J是雅可比行列式，即

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

而$|J|$是行列式J的绝对值，这是由于变换函数可能是递增也可能是递减，但密度函数总是正值的缘故。在变换中存在雅可比行列式是因为$\boldsymbol{Y}$坐标系中的面积$dy_1 \cdots dy_n$相应于$\boldsymbol{X}$坐标系中面积$|J|dx_1 \cdots dx_n$，雅可比行列式是衡量由变换引起的尺度变化的一个系数。

此为试读，需要完整PDF请访问：www.ertongbook.com