模式识别及MATLAB实现

学习与实验指导

◎郭志强　主　编
◎杨　杰　副主编

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

模式识别及 MATLAB 实现：学习与实验指导

郭志强 主 编

杨 杰 副主编

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书是与电子工业出版社出版的《模式识别及 MATLAB 实现》配套的学习指导书，在章节安排上与主教材一致，各章节内容包括本章知识结构、知识要点和实验指导，实验指导部分给出了实验步骤、MATLAB 代码和实验结果。

实验的内容和训练对模式识别学习者有很大帮助，也为从事模式识别的工程技术人员提供了一定的指导。

图书在版编目（CIP）数据

模式识别及 MATLAB 实现：学习与实验指导 / 郭志强主编. —北京：电子工业出版社，2017.9
ISBN 978-7-121-32373-7

Ⅰ. ①模… Ⅱ. ①郭… Ⅲ. ①模式识别－计算机辅助计算－Matlab 软件②人工智能－计算机辅助计算－Matlab 软件 Ⅳ. ①O235-39②TP183

中国版本图书馆 CIP 数据核字(2017)第 183906 号

策划编辑：董亚峰
责任编辑：赵　娜
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
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：787×1 092　1/16　印张：13　字数：333 千字
版　　次：2017 年 9 月第 1 版
印　　次：2017 年 9 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）88254888，88258888。

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

本书咨询联系方式：（010）88254694。

前　言

本书是武汉理工大学杨杰和郭志强编写的教材《模式识别及 MATLAB 实现》的学习和实验指导用书，可与教材配套使用，也可单独作为高等学校模式识别课程的教学与学习参考书，还可作为模式识别领域专业技术人员的参考资料。

模式识别是一门理论和工程应用都发展十分迅速的学科，尤其随着大数据的出现和互联网+的兴起，模式识别已伴随着人工智能技术渗透到人们生活的方方面面。“模式识别”作为信息类专业硕士研究生的学位课，主要介绍模式识别的基础知识和基本理论，为进一步研究模式识别理论和技术打下良好的基础。同时，模式识别也是一门实践性很强的学科，通过一定量的实验训练，有助于学习者加深理解和巩固所学的基本理论知识，也有助于提高其解决实际工程问题的能力。

全书分为 7 章，每章都按本章知识结构、知识要点和实验指导三部分编写。具体内容包括贝叶斯决策、参数估计、非参数判别分类法、聚类分析法、特征选择与提取、模糊模式识别、神经网络在模式识别中的应用等，每章实验均给出了实验步骤、MATLAB 代码和实验结果。实验的内容和训练对模式识别学习者有很大帮助，也为从事模式识别的工程技术人员提供了一定的指导。

本书第 1～4 章由郭志强编写，第 5～7 章由杨杰编写，编者指导的研究生王贺、吴紫薇、林仲康和李博闻等参加了程序调试、插图和校对工作。在编写本书过程中，参阅了大量模式识别参考书，这里谨向有关作者表示衷心感谢。

由于作者水平有限，书中难免存在疏漏和不当之处，恳请读者批评指正。

目 录

第1章 贝叶斯决策

贝叶斯决策是统计决策的核心，其基本思想是根据各类特征的后验概率进行决策。通过贝叶斯公式，可以将后验概率的比较转为成类条件概率密度的比较，也可能定义两类之间的似然比或对数似然比进行决策。当样本类别的概率密度服从正态分类时，贝叶斯决策在工程实践中有较广泛的应用。

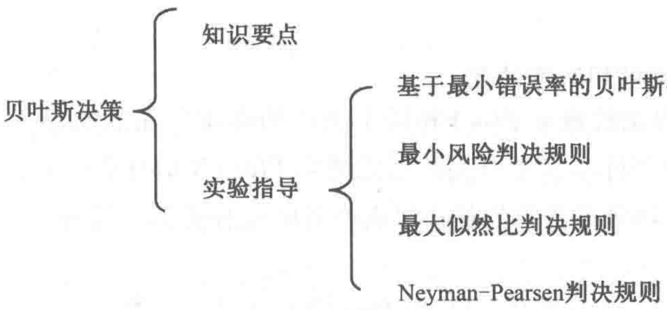

图 1-1　本章知识结构

1.1　知识要点

1. 先验概率

先验概率在分类方法中有着重要的作用，它的函数形式及主要参数或者是已知的，或者是可通过大量抽样实验估计出来的。

若用 ω_1 和 ω_2 分别表示为两个类别，$P(\omega_1)$ 和 $P(\omega_2)$ 表示各自的先验概率，此时满足

$$P(\omega_1)+P(\omega_2)=1$$

推广到 c 类问题中，$\omega_1,\omega_2,\cdots,\omega_c$ 表示 c 个类别，各自的先验概率用 $P(\omega_1),P(\omega_2),\cdots,P(\omega_c)$ 表示，则满足

$$P(\omega_1)+P(\omega_2)+\cdots+P(\omega_c)=1$$

2. 类（条件）概率密度

类（条件）概率密度是指在某种确定类别条件下，模式样本 $\boldsymbol{X}$ 出现的概率密度分布函数，常用 $p(\boldsymbol{X}\mid\omega_i)(i\in 1,2,\cdots,c)$ 来表示。在本书中，我们采用 $p(\boldsymbol{X}\mid\omega_i)$ 表示条件概率密度函数，

$P(\boldsymbol{X} \mid \omega_i)$ 表示其对应的条件概率。$p(*|\#)$ 是条件概率的通用符号，在“|”后边出现的#为条件，之前的*为某个事件，即在某条件#下出现某个事件*的概率，如 $P(\omega_k \mid \boldsymbol{X})$ 表示在 $\boldsymbol{X}$ 出现条件下，样本为 ω_k 类的概率。

3．后验概率

后验概率是在某个具体的模式样本 $\boldsymbol{X}$ 条件下，某种类别出现的概率，常以 $P(\omega_i \mid \boldsymbol{X})(i=1,2,\cdots,c)$ 表示。后验概率可以根据贝叶斯公式或式（1.1）计算出来并直接用作分类判决的依据。

$$P(\omega_i|\boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \omega_i)P(\omega_i)}{p(\boldsymbol{X})} \tag{1.1}$$

其中：

$$p(\boldsymbol{X}) = \sum_{i=1}^{c} p(\boldsymbol{X}|\omega_i)P(\omega_i) \tag{1.2}$$

先验概率是指 $\omega_i(i=1,2,\cdots,c)$ 出现的可能性，不考虑其他任何条件。类条件概率密度函数 $p(\boldsymbol{X} \mid \omega_i)$ 是指 ω_i 条件下在一个连续的函数空间出现 $\boldsymbol{X}$ 的概率密度，即第 ω_i 类样本的特征 $\boldsymbol{X}$ 是如何分布的。

4．基于最小错误率的贝叶斯决策

当已知类别出现的先验概率 $P(\omega_i)$ 和每个类中的样本分布的类条件概率密度 $p(\boldsymbol{X} \mid \omega_i)$ 时，可以求得一个待分类样本属于每类的后验概率 $P(\omega_i \mid \boldsymbol{X}), i=1,2,\cdots,c$。将其划归到后验概率最大的那一类中，这种分类器称为最小错误率贝叶斯分类器，其分类决策准则可表示为：

（1）两类情况

$$\begin{cases} 若 P(\omega_1|\boldsymbol{X}) > P(\omega_2|\boldsymbol{X})，则 \boldsymbol{X} \in \omega_1 类 \\ 若 P(\omega_2|\boldsymbol{X}) > P(\omega_1|\boldsymbol{X})，则 \boldsymbol{X} \in \omega_2 类 \end{cases} \tag{1.3}$$

（2）多类情况

$$若 P(\omega_i|\boldsymbol{X}) = \max\left\{P(\omega_j|\boldsymbol{X})\right\}, j=1,2\cdots,c \quad 则 \boldsymbol{X} \in \omega_i 类 \tag{1.4}$$

由式(1.1)，已知待识别样本 $\boldsymbol{X}$ 后，可以通过先验概率 $P(\omega_i)$ 和条件概率密度函数 $p(\boldsymbol{X} \mid \omega_i)$ 得到样本 $\boldsymbol{X}$ 分属各类别的后验概率，显然这个概率值可以作为 $\boldsymbol{X}$ 类别归属的依据。该判别依据可以有以下几种等价形式。

从贝叶斯公式式（1.1）可以看出，分母与 i 无关，即与分类无关，故分类规则又可表示为：

$$若\ p(\boldsymbol{X} \mid \omega_i)P(\omega_i) = \max\left\{p(\boldsymbol{X} \mid \omega_j)P(\omega_j)\right\}\ j=1,2,\cdots,c,\ 则\ \boldsymbol{X} \in \omega_i 类 \tag{1.5}$$

对两类问题，式（1.5）相当于

$$\begin{cases} p(\boldsymbol{X} \mid \omega_1)P(\omega_1) > p(\boldsymbol{X} \mid \omega_2)P(\omega_2), & \boldsymbol{X} \in \omega_1 \\ p(\boldsymbol{X} \mid \omega_2)P(\omega_2) > p(\boldsymbol{X} \mid \omega_1)P(\omega_1), & \boldsymbol{X} \in \omega_2 \end{cases} \tag{1.6}$$

式（1.6）可改写为

$$l_{12}(\boldsymbol{X})=\frac{P(\boldsymbol{X}|\omega_1)>P(\omega_2)}{P(\boldsymbol{X}|\omega_2)<P(\omega_1)},\quad \boldsymbol{X}\in\begin{cases}\omega_1\\ \omega_2\end{cases}\tag{1.7}$$

统计学中称 $l_{12}(\boldsymbol{X})$为似然比，$P(\omega_2)/P(\omega_1)$为似然比阈值。

对式（1.7）取自然对数，有

$$\ln l_{12}(\boldsymbol{X})=\ln p(\boldsymbol{X}\mid\omega_1)-\ln p(\boldsymbol{X}\mid\omega_2)\begin{matrix}>\ln P(\omega_2)\\ <\ln P(\omega_1)\end{matrix},\quad \boldsymbol{X}\in\begin{cases}\omega_1\\ \omega_2\end{cases}\tag{1.8}$$

式（1.5）、式（1.7）和式（1.8）都是贝叶斯决策规则的等价形式。可以发现，上述分类决策规则实为“最大后验概率分类器”，易知其分类错误的概率为

$$p(e)=\int_{-\infty}^{\infty}p(e,\boldsymbol{X})\mathrm{d}\boldsymbol{X}=\int_{-\infty}^{\infty}p(e|\boldsymbol{X})p(\boldsymbol{X})\mathrm{d}\boldsymbol{X}$$

而

$$p(e\mid\boldsymbol{X})=\sum_{i=1}^{c}p(\omega_i|\boldsymbol{X})-\max_{1\leqslant i\leqslant c}p(\omega_i|\boldsymbol{X})$$

显然，当 $p(e\mid\boldsymbol{X})$ 取最小值时，$p(e)$ 为最小值，“最大后验概率分类器”与“最小错误率分类器”是等价的。

5．最小风险判决规则

最小错误率判决规则没有考虑错误判决带来的“风险”，或者说没有考虑某种判决带来的损失。同一问题中，不同的判决有不同的风险，例如判断细胞是否为癌细胞，可能有两种错误判决：①正常细胞错判为癌细胞；②癌细胞错判为正常细胞。但两种错误带来的风险并不相同。在①中，会给健康人带来不必要的精神负担；在②中，会使患者失去进一步检查和治疗的机会，造成严重后果。显然，第②种错误判决的风险大于第①种。

正是由于有判决风险的存在，仅考虑最小错误进行判决是不充分的，还必须考虑判决带来的风险，因此引入最小风险判决规则。事实上，最小风险判决规则也是一种 Bayes 分类方法。判决风险也可以理解为由判决而付出的代价，即使在做出正确判决的情况下，也会付出一定的代价，也会有损失。

假定有 c 类问题，用 $\omega_j(j=1,2,\cdots,c)$ 表示类别，用 $a_i(i=1,2,\cdots,a)$ 表示可以做出的判决。实际应用中，判决数 a 和类别数 c 可能相等；也可能不等，即允许除 c 类的 c 个决策之外，可以采用其他决策，如“拒绝”决策，此时 $\alpha=c+1$。

对于给定的模式 $\boldsymbol{X}$，令 $L(\alpha_i\mid\omega_j)$ 表示 $\boldsymbol{X}\in\omega_j$ 而判决为 α_i 的风险。若已做出判决 α_i，对 c 个不同类别 ω_j，有 c 个不同的 $L(\alpha_i\mid\omega_j)$。

假定某样本 $\boldsymbol{X}$ 的后验概率 $P(\omega_j\mid\boldsymbol{X})$ 已经确定，则有：

$$P(\omega_1\mid\boldsymbol{X})+P(\omega_2\mid\boldsymbol{X})+\cdots+P(\omega_c\mid\boldsymbol{X})=1, j=1,2,\cdots,c\text{，且 }P(\omega_j\mid\boldsymbol{X})\geqslant 0\text{，}$$

对于每一种判决 α_i，可求出随机变量 $L(\alpha_i\mid\omega_i)$ 的条件平均风险，也叫“条件平均损失”：

$$R(\alpha_i\mid\boldsymbol{X})=E[L(\alpha_i\mid\omega_j)]=\sum_{j=1}^{c}L(\alpha_i\mid\omega_j)\cdot P(\omega_j\mid\boldsymbol{X})\qquad i=1,2,\cdots,a\tag{1.9}$$

最小风险判决规则就是把样本 $\boldsymbol{X}$ 归属于“条件平均风险最小”的那一种判决。也即

$$若\ R(\alpha_i \mid \boldsymbol{X}) = \min_{k=1,2,\cdots,a} \{R(\alpha_k \mid \boldsymbol{X})\}，\ 则\ \boldsymbol{X} \in \omega_i \tag{1.10}$$

实施最小风险判决规则的步骤如下：

（1）给定样本$\boldsymbol{X}$，计算各类后验概率$P(\omega_j \mid \boldsymbol{X})$，$j=1,2,\cdots,c$。

（2）在已知风险矩阵的条件下，按照式（1.9）求各种判决的条件平均风险$R(\alpha_i \mid \boldsymbol{X})$，$i=1,2,\cdots,a$。

（3）按照式（1.10），比较各种判决的条件平均风险，把样本$\boldsymbol{X}$归属于条件平均风险最小的那一种判决。

6．最大似然比判决规则

类条件概率密度函数$p(\boldsymbol{X} \mid \omega_i)$又称“似然函数”，两个类条件概率密度之比称为“似然比函数”。可定义为

$$l_{ij}(x) = \frac{p(\boldsymbol{X} \mid \omega_i)}{p(\boldsymbol{X} \mid \omega_j)} \qquad i,j=1,2,\cdots,c，\ 且\ i \neq j \tag{1.11}$$

最大似然比判决规则可描述为：类型ω_i分别与其他类型$\omega_j(j=1,2,\cdots,c, j\neq i)$的似然比均大于相应的门限值$\theta_{ij}$时，则样本$\boldsymbol{X}\in\omega_i$。事实上，最大似然比判决规则也是一种Bayes分类方法。

（1）由最小错误率判决规则引出最大似然比判决规则

下面以二分类问题为例，借助最小错误率判决规则引出最大似然比判决规则，若$X\in\omega_1$，由式（1.6）知最小错误率判决规则为：

$$p(\boldsymbol{X} \mid \omega_1)\cdot P(\omega_1) > p(\boldsymbol{X} \mid \omega_2)\cdot P(\omega_2)$$

两边同时除以$p(\boldsymbol{X} \mid \omega_2)\cdot P(\omega_1)$有

$$\frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$

类别ω_1与ω_2的似然比为：$l_{12}(\boldsymbol{X}) = \dfrac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)}$

则判决门限为

$$\theta_{12} = \frac{P(\omega_2)}{P(\omega_1)} \tag{1.12}$$

当先验概率已知时，可求得θ_{12}。所以“最小错误率判决规则”就变为

$$\begin{cases} l_{12}(\boldsymbol{X}) > \theta_{12}, & \boldsymbol{X}\in\omega_1 \\ l_{12}(\boldsymbol{X}) < \theta_{12}, & \boldsymbol{X}\in\omega_2 \\ l_{12}(\boldsymbol{X}) = \theta_{12}, \boldsymbol{X}\in\omega_1或\boldsymbol{X}\in\omega_1 \end{cases} \tag{1.13}$$

（2）由最小风险判决规则引出最大似然比判决规则

也可由最小风险判决规则引出最大似然比判决规则，同样以二分类问题为例，若模式$\boldsymbol{X}\in\omega_1$，根据最小风险判决规则，则有

$$R(\alpha_1=\omega_1 \mid \boldsymbol{X}) < R(\alpha_2=\omega_2 \mid \boldsymbol{X})$$

考虑到 $R(\alpha_i=\omega_i\mid\boldsymbol{X})=\sum_{j=1}^{2}L(\alpha_i\mid\omega_j)\cdot p(\omega_j\mid\boldsymbol{X})$，有

$$[L(\alpha_2\mid\omega_1)-L(\alpha_1\mid\omega_1)]P(\omega_1\mid\boldsymbol{X})>[L(\alpha_1\mid\omega_2)-L(\alpha_2\mid\omega_2)]P(\omega_2\mid\boldsymbol{X})$$

即

$$\frac{P(\omega_1\mid\boldsymbol{X})}{P(\omega_2\mid\boldsymbol{X})}>\frac{L(\alpha_1\mid\omega_2)-L(\alpha_2\mid\omega_2)}{L(\alpha_2\mid\omega_1)-L(\alpha_1\mid\omega_1)}$$

又由 Bayes 公式

$$\frac{P(\omega_1\mid\boldsymbol{X})}{P(\omega_2\mid\boldsymbol{X})}=\frac{p(\boldsymbol{X}\mid\omega_1)\cdot P(\omega_1)}{p(\boldsymbol{X}\mid\omega_2)\cdot P(\omega_2)}$$

得

$$\frac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)}>\frac{L(\alpha_1\mid\omega_2)-L(\alpha_2\mid\omega_2)}{L(\alpha_2\mid\omega_1)-L(\alpha_1\mid\omega_1)}\cdot\frac{P(\omega_2)}{P(\omega_1)} \tag{1.14}$$

即

$$l_{12}(\boldsymbol{X})>\theta_{12}$$

式中

$$\theta_{12}=\frac{L(\alpha_1\mid\omega_2)-L(\alpha_2\mid\omega_2)}{L(\alpha_2\mid\omega_1)-L(\alpha_1\mid\omega_1)}\cdot\frac{P(\omega_2)}{P(\omega_1)} \tag{1.15}$$

为判决门限。

$$\text{若 } l_{ij}(\boldsymbol{X})>\theta_{ij}\text{，则 }\boldsymbol{X}\in\omega_i \tag{1.16}$$

$$\text{其中，}\theta_{ij}=\frac{P(\omega_j)}{P(\omega_i)} \tag{1.17}$$

由最小风险判决规则导出

$$\theta_{ij}=\frac{[L(\alpha_i\mid\omega_j)-L(\alpha_i\mid\omega_j)]\cdot P(\omega_j)}{[L(\alpha_j\mid\omega_i)-L(\alpha_i\mid\omega_i)]\cdot P(\omega_i)} \tag{1.18}$$

同样在 0～1 损失函数的情况下，式（1.18）退化为式（1.17）。

由于似然函数满足 $l_{ij}(\boldsymbol{X})=\dfrac{1}{l_{ij}(\boldsymbol{X})}$，所以在 c 类问题中，若有一个 ω_i 满足式（1.16），则不可能再有另外的类别 $\omega_j(i\neq j)$ 满足式（1.16）。

7．Neyman-Pearsen 判决规则

在二分类问题中，贝叶斯判决规则的基本思想是根据类别的先验概率和类条件概率将样本的特征空间 R 划分成两个子区域 R_1 和 R_2。这时存在两种错误，一种是当样本 $\boldsymbol{X}$ 应属 ω_2 时，判决为 ω_1；另一种是当样本 $\boldsymbol{X}$ 应属 ω_1 时，判决为 ω_2。两种错误的概率分别为：$P_1(e)=\int_{R_2}p(\boldsymbol{X}\mid\omega_1)\mathrm{d}\boldsymbol{X}$，$P_2(e)=\int_{R_1}p(\boldsymbol{X}\mid\omega_2)\mathrm{d}\boldsymbol{X}$，总的错误之和 $P(e)$ 为

$$P(e)=P(\omega_2)\cdot P_2(e)+P(\omega_1)\cdot P_1(e) \tag{1.19}$$

最小错误率 Bayes 决策是使 $P(e)$ 为最小。

$$\varepsilon_{12}=P_1(e)=\int_{R_2}p(\boldsymbol{X}\mid\omega_1)\mathrm{d}\boldsymbol{X} \tag{1.20}$$

$$\varepsilon_{21} = P_2(e) = \int_{R_1} p(\boldsymbol{X} \mid \omega_2)\mathrm{d}\boldsymbol{X} \tag{1.21}$$

假定ε_{21}保持不变，为某个给定的正数，令：

$$\varepsilon = \varepsilon_{12} + \mu\varepsilon_{21} \tag{1.22}$$

为使ε_{12}最小化，就要通过适当地选择某个正数μ使ε最小。

$$\varepsilon_{12} = 1 - \int_{R_1} p(\boldsymbol{X} \mid \omega_1)\mathrm{d}\boldsymbol{X} \tag{1.23}$$

$$\varepsilon_{21} = 1 - \int_{R_2} p(\boldsymbol{X} \mid \omega_2)\mathrm{d}\boldsymbol{X} \tag{1.24}$$

把式（1.23）和式（1.21）代入式（1.22），得到

$$\varepsilon = 1 + \int_{R_1} [\mu p(\boldsymbol{X} \mid \omega_2) - p(\boldsymbol{X} \mid \omega_1)]\mathrm{d}\boldsymbol{X} \tag{1.25}$$

把式（1.24）和式（1.20）代入式（1.22），得到

$$\varepsilon = \mu + \int_{R_2} [p(\boldsymbol{X} \mid \omega_1) - \mu p(\boldsymbol{X} \mid \omega_2)]\mathrm{d}\boldsymbol{X} \tag{1.26}$$

为了使ε最小化，上两式中的被积函数最好为负数，从而得到聂曼-皮尔逊判决规则为

$$\left.\begin{aligned} &若 \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} > \mu, \quad 则\boldsymbol{X} \in \omega_1 \\ &若 \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} < \mu \quad 则\boldsymbol{X} \in \omega_2 \end{aligned}\right\} \tag{1.27}$$

从式（1.27）可以看出，聂曼-皮尔逊判决规则归结为寻找判决阈值μ，显然μ是$\boldsymbol{X}$的函数，根据上式，要求$\mu(\boldsymbol{X})$为

$$\mu(\boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} \tag{1.28}$$

为了最后确定判决阈值，利用给定的正数ε_{21}，由式（1.21），并参考图1-2，得到

$$\varepsilon_{21} = \int_{-\infty}^{\mu^{-1}(\boldsymbol{X})} p(\boldsymbol{X} \mid \omega_2)\mathrm{d}\boldsymbol{X} \tag{1.29}$$

式中，$\mu^{-1}(\boldsymbol{X})$为$\mu(\boldsymbol{X})$的逆函数。

8. 正态分布中的Bayes分类方法

由式（1.5）的最小错误率的判决准则，可得其对应的判别函数为

$$g_i(\boldsymbol{X}) = p(\boldsymbol{X} \mid \omega_i) \cdot P(\omega_i) \qquad i = 1,2,\cdots,c \tag{1.30}$$

对c类问题，其判决规则为：

$$g_i(\boldsymbol{X}) > g_j(\boldsymbol{X}), i = 1,2,\cdots,c, j \neq i \Rightarrow \boldsymbol{X} \in \omega_i \tag{1.31}$$

此时任两个类别的决策面方程为

$$g_i(\boldsymbol{X}) = g_j(\boldsymbol{X}) \tag{1.32}$$

设$\boldsymbol{X}$为n维特征向量，且$p(\boldsymbol{X} \mid \omega_i)$服从正态分布的，即$p(\boldsymbol{X} \mid \omega_i) \sim N(\boldsymbol{\mu}_i, \Sigma_i)$，则

$$g_i(\boldsymbol{X}) = \frac{P(\omega_i)}{(2\pi)^{n/2} \mid \Sigma_i \mid^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_i)\right] \tag{1.33}$$

为了方便计算，对原判别函数取对数，$g_i(x)$可写为如下形式

$$g_i(\boldsymbol{X})=-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_i)-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i) \tag{1.34}$$

式中，$\frac{n}{2}\ln 2\pi$ 与类别无关，不影响分类决策，可以去掉。因此 $g_i(\boldsymbol{X})$ 可进一步简化为

$$g_i(\boldsymbol{X})=-\frac{1}{2}(\boldsymbol{X}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{X}-\mu_i)-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i) \tag{1.35}$$

将式（1.35）代入式（1.32），得

$$-\frac{1}{2}\left(\ln|\Sigma_i|-\ln|\Sigma_j|\right)-\frac{1}{2}\left[(\boldsymbol{X}-\boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_i)-(\boldsymbol{X}-\boldsymbol{\mu}_j)^{\mathrm{T}}\Sigma_j^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_j)\right]+\ln\frac{P(\omega_i)}{P(\omega_j)}=0 \tag{1.36}$$

式中，Σ_i 为 ω_i 类的 $n\times n$ 维协方差矩阵，$\boldsymbol{\mu}_i=(\mu_1,\mu_2,\cdots,\mu_n)^{\mathrm{T}}$ 为 ω_i 类的 n 维均值向量，$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 为 n 维的特征向量，Σ_i^{-1} 为 Σ_i 的逆阵，$|\Sigma_i|$ 为 Σ_i 的行列式。

设 $\Sigma_i=\sigma^2 I$

即每类的协方差矩阵都相等，类内各特征维度间相互独立，且方差相同。

$$\boldsymbol{\Sigma}_i=\sigma^2\boldsymbol{I}=\begin{bmatrix}\sigma^2 & \dots & 0\\ \dots & \dots & \dots\\ 0 & \dots & \sigma^2\end{bmatrix}$$

式（1.34）的判别函数重写为

$$g_i(\boldsymbol{X})=-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_i)^{\mathrm{T}}\Sigma_i^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_i)-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$$

将式中与类别无关的项 $\Sigma_i=\sigma^2 I$，$\Sigma_i^{-1}=I/\sigma^2$，$|\Sigma_i|=\sigma^{2n}$，$\frac{n}{2}\ln 2\pi$ 去掉，判别函数可简化为

$$g_i(\boldsymbol{X})=-\frac{\|\boldsymbol{X}-\boldsymbol{\mu}_i\|^2}{2\sigma^2}+\ln P(\omega_i) \tag{1.37}$$

其中，$\|\boldsymbol{X}-\boldsymbol{\mu}_i\|^2=(\boldsymbol{X}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu}_i)$。

1.2　实验指导

1.2.1　基于最小错误率的贝叶斯决策

1．实验内容

（1）学习、理解基于最小错误率贝叶斯决策的基本原理。

（2）设计贝叶斯决策的 MATLAB 算法，对一定数据进行分析，验证算法的正确性。

（3）对实验结果进行总结分析。

2．实验原理

本次实验所采用的 iris 数据样本具有 $d=4$ 个特征，样本数据其正态分布的概率密度函

数为

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right\}$$

该式中，$\boldsymbol{x}=[x_1,x_2,\cdots,x_d]$ 是 d 维行向量，$\boldsymbol{\mu}=[\mu_1,\mu_2,\cdots,\mu_d]$ 是 d 维行向量，$\boldsymbol{\Sigma}$ 是 $d\times d$ 维协方差矩阵，$\boldsymbol{\Sigma}^{-1}$ 是 $\boldsymbol{\Sigma}$ 的逆矩阵，$|\boldsymbol{\Sigma}|$ 是 $\boldsymbol{\Sigma}$ 的行列式。使用如下的函数作为判别函数

$$g_i(\boldsymbol{x})=p(\boldsymbol{x}\mid\omega_i)P(\omega_i),\qquad i=1,2,3\text{（3 个类别）}$$

其中，$P(\omega_i)$ 为类别 ω_i 发生的先验概率，$p(\boldsymbol{x}\mid\omega_i)$ 为类别 ω_i 的类条件概率密度函数。如果使 $g_i(\boldsymbol{x})>g_j(\boldsymbol{x})$ 对一切 $j\neq i$ 成立，则将 $\boldsymbol{x}$ 归为 ω_i 类。根据假设：类别 ω_i，i=1,2,⋯,N 的类条件概率密度函数 $p(\boldsymbol{x}\mid\omega_i)$，$i$=1,2,⋯,$N$ 服从正态分布，即有 $p(\boldsymbol{x}\mid\omega_i)\sim N(\boldsymbol{\mu}_i,\Sigma_i)$，则上式可以写为

$$g_i(\boldsymbol{x})=\frac{P(\omega_i)}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right\},\quad i=1,2,3$$

对上式右端取对数，可得

$$g_i(\boldsymbol{x})=-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{T}}+\ln P(\omega_i)-\frac{1}{2}\ln|\boldsymbol{\Sigma}_i|-\frac{d}{2}\ln(2\pi)$$

上式中的第二项与样本所属类别无关，将其从判别函数中消去，不会改变分类结果。则判别函数 $g_i(\boldsymbol{x})$ 可简化为以下形式

$$g_i(\boldsymbol{x})=-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{T}}+\ln P(\omega_i)-\frac{1}{2}\ln|\boldsymbol{\Sigma}_i|$$

3. 实验方法及程序

（1）从 Iris.txt 文件中读取估计参数用的样本，每一类样本抽出前 40 个，分别求其均值，代码如下：

```
clear
%原始数据导入
iris = load('C:\Matlab7\work\模式识别\iris.txt');
N=40;%每组取 N=40 个样本

%求第一类样本均值
for i = 1:N
    for j = 1:4
        w1(i,j) = iris(i,j+1);
    end
end
sumx1 = sum(w1,1);
for i=1:4
    meanx1(1,i)=sumx1(1,i)/N;
end
```

```
%求第二类样本均值
for i = 1:N
    for j = 1:4
        w2(i,j) = iris(i+50,j+1);
    end
end
sumx2 = sum(w2,1);
for i=1:4
    meanx2(1,i)=sumx2(1,i)/N;
end

%求第三类样本均值
for i = 1:N
    for j = 1:4
        w3(i,j) = iris(i+100,j+1);
    end
end
sumx3 = sum(w3,1);
for i=1:4
    meanx3(1,i)=sumx3(1,i)/N;
end
```

（2）求每一类样本的协方差矩阵、逆矩阵 $\mathbf{\Sigma}_i^{-1}$ 以及协方差矩阵的行列式 $|\mathbf{\Sigma}_i|$，代码如下：

```
%求第一类样本协方差矩阵
z1 (4,4)  = 0;
var1 (4,4)  = 0;
for i=1:4
for j=1:4
      for k=1:N
          z1 (i,j) =z1 (i,j) + (w1 (k,i) -
      meanx1 (1,i)) * (w1 (k,j) -meanx1 (1,j)) ;
       end
     var1 (i,j)  = z1 (i,j) / (N-1) ;
end
end

%求第二类样本协方差矩阵
z2 (4,4)  = 0;
var2 (4,4)  = 0;
for i=1:4
```

```
 for j=1:4
       for k=1:N
           z2 (i,j) =z2 (i,j) +
      (w2 (k,i) -meanx2 (1,i)) * (w2 (k,j) -meanx2 (1,j)) ;
        end
         ar2 (i,j) = z2 (i,j) / (N-1) ;
  end
end

%求第三类样本协方差矩阵
z3 (4,4) = 0;
var3 (4,4) = 0;
for i=1:4
 for j=1:4
    for k=1:N
          z3 (i,j) =z3 (i,j) + (w3 (k,i) -
      meanx3 (1,i)) * (w3 (k,j) -meanx3 (1,j)) ;
        end
         var3 (i,j) = z3 (i,j) / ( N-1) ;
  end
end

%求各类的协方差矩阵逆矩阵及行列式
var1_inv = [];var1_det = [];
var2_inv = [];var2_det = [];
var3_inv = [];var3_det = [];
var1_inv = inv (var1)
var2_inv = inv (var2)
var3_inv = inv (var3)
var1_det = det (var1)
var2_det = det (var2)
var3_det = det (var3)
```

（3）对三个类别，分别取每组剩下的 10 个样本，每两组进行分类。由于每一类样本都相等，且每一类选取用作训练的样本也相等，在每两组进行分类时，待分类样本的类先验概率 $P(\omega_i)=0.5$。将各个样本代入判别函数

$$g_i(\boldsymbol{x})=-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})\Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{T}}+\ln P(\omega_i)-\frac{1}{2}\ln|\Sigma_i|$$

根据判决规则，如果使 $g_i(\boldsymbol{x})>g_j(\boldsymbol{x})$ 对一切 $j\neq i$ 成立，则将 $\boldsymbol{x}$ 归为 ω_i 类。若取第一类后 10 个数据和第二类进行分类，代码如下：