厦门大学信息学院 本科选修课
2021-2022 第二学期

模 式 识 别
Pattern Recognition

主 讲：王 程

第四章　贝叶斯决策理论

4.1引言

4.2基于最小错误率的贝叶斯决策

4.3基于最小风险的贝叶斯决策

4.4分类器、判别函数及决策面

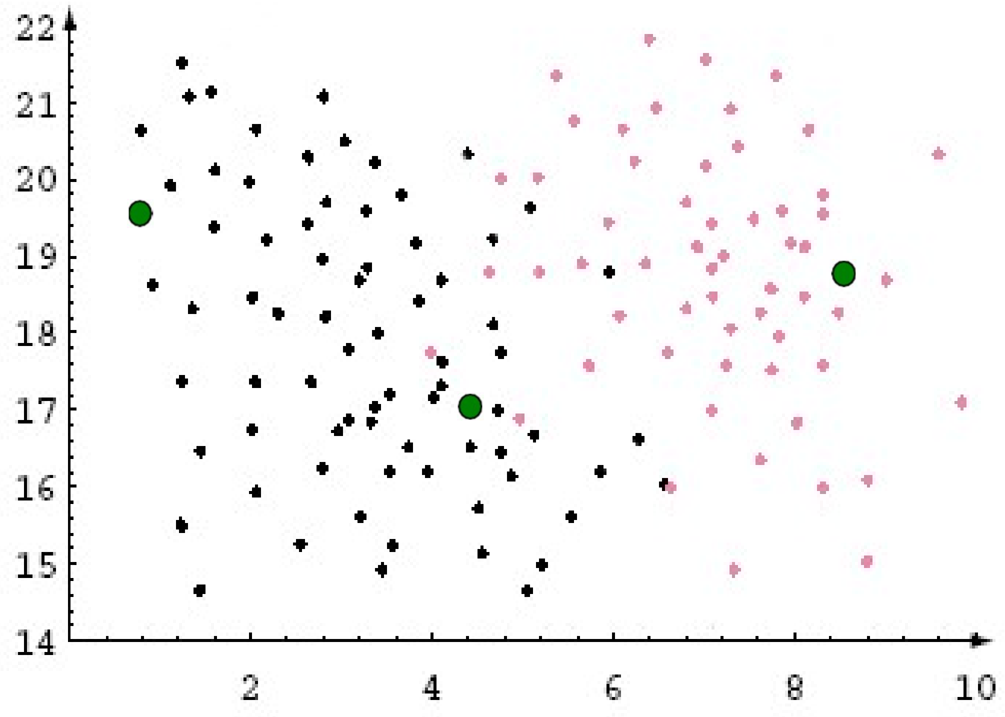
4.5正态分布下的统计决策

4.1 引 言

- 统计决策理论
  - 是模式分类问题的基本理论之一
- 贝叶斯决策理论
  - 是统计决策理论中的一个基本方法
- 贝叶斯决策的两个要求
  - 各个类别的总体概率分布(先验概率和类条件概率密度) 是已知的
  - 要决策分类的类别数是一定的

4.1 引 言

决 策



黑色：第一类

粉色：第二类

绿色：哪一类？

统计决策理论就是根据每一类总体的概率分布决定未知类别的样本属于哪一类！

4.1 引 言

决策准则

- 评价决策有多种标准，对于同一个问题，采用不同的标准会得到不同意义下"最优"的决策
- 贝叶斯决策常用的准则：
  - 最小错误率准则
  - 最小风险准则
  - **Neyman-Pearson**(黎曼皮尔逊)准则
  - 最小最大决策准则

4.1 引 言

基本概念

- 在连续情况下，假设对要识别的物理对象有d种特 征观察量$x1,x2,\ldots xd$，这些特征的所有可能的取值 范围构成了d维特征空间。

- 称向量 $\boldsymbol{x}=\left[x_1,x_2,\cdots,x_d\right]^T$ $\boldsymbol{x}\in\boldsymbol{R}^d$ 为d维特征向量。

- 假设要研究的分类问题有c个类别，类型空间表示为：

$$\Omega=\left\{\omega_1,\omega_2,\cdots,\omega_c\right\}$$

4.1 引 言

几个重要概念

- 先验概率: $P(\omega_i)$ 未获得观测数据之前类别的分布
- 类条件概率: $p(\boldsymbol{x}|\omega_i)$ 表示在ω_i类条件下$\boldsymbol{x}$的概率分布密度
- 后验概率: $P(\omega_i|\boldsymbol{x})$ 在$\boldsymbol{x}$出现条件下ω_i类出现的概率

第五章　贝叶斯决策理论

4.1引言

4.2基于最小错误率的贝叶斯决策

4.3基于最小风险的贝叶斯决策

4.4分类器、判别函数及决策面

4.5正态分布下的统计决策

4.2 基于最小错误率的贝叶斯决策

- 鲈鱼/鲑鱼例子
  - 自然状态下，先验的类别状态，ω_i,　　i=1,2
    - ω_i类别状态是一个随机变量, $P(\omega_i)$ 表示为先验概率。
    - 捕获鲈鱼和鲑鱼的几率相等。

$$P(\omega_1) = P(\omega_2) \quad (\text{先验})$$

$$P(\omega_1) + P(\omega_2) = 1 \ (\text{排除其它鱼的种类})$$

4.2 基于最小错误率的贝叶斯决策

- 仅含先验信息的判别规则

$$\begin{cases} P(\omega_1) > P(\omega_2), & \omega_1 \\ P(\omega_1) < P(\omega_2), & \omega_2 \end{cases}$$

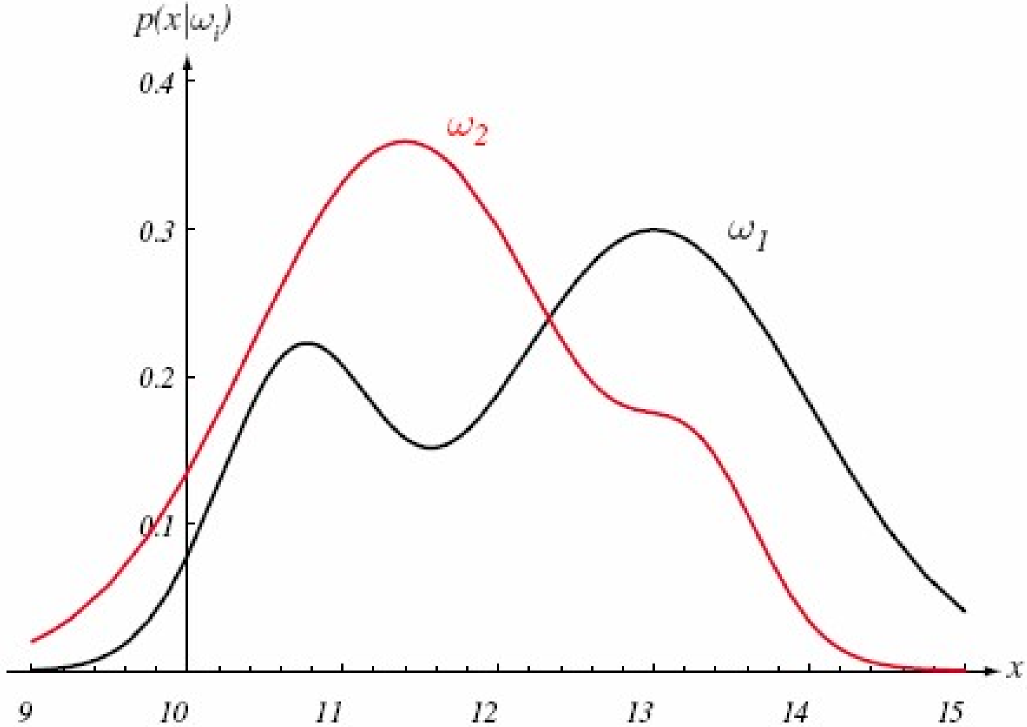
  - 这种分类决策没有意义
  - 由先验概率所提供的信息太少

- 采用类条件信息——类条件概率密度函数

  $p(x|\omega_1)$：鲈鱼的属性分布

  $p(x|\omega 2)$：鲑鱼的属性分布。

4.2 基于最小错误率的贝叶斯决策



鲈鱼和鲑鱼判别中的类条件概率密度函数
（以光泽度为例）

4.2 基于最小错误率的贝叶斯决策

- 贝叶斯公式
  - 先验概率，后验概率，概率密度函数之间关系

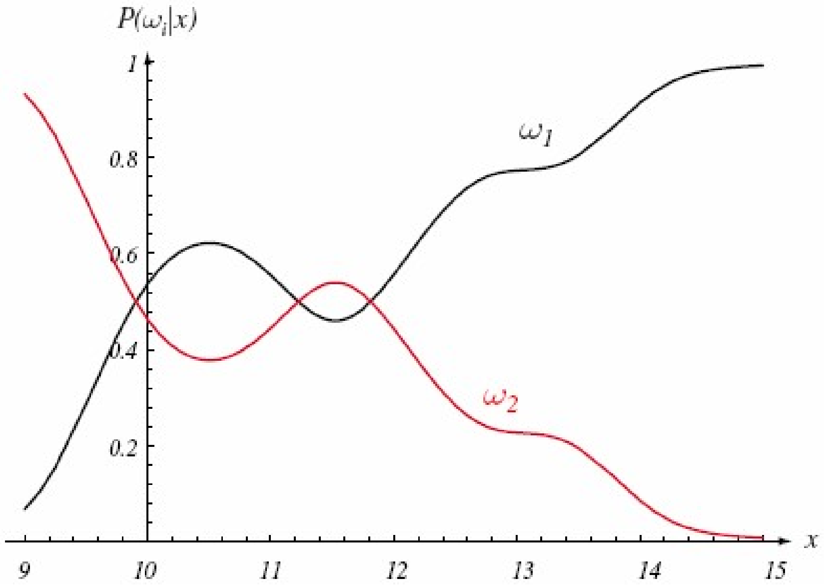
$$P(\omega_i \mid x) = \frac{p(x \mid \omega_i)P(\omega_i)}{\sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)} \qquad i = 1,2$$

其中，$p(x) = \sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)$为一因子

贝叶斯公式通过类条件概率密度形式的观察值， 将先验概率转化为后验概率。

4.2 基于最小错误率的贝叶斯决策

- 后验概率含义
  - $P(\omega_1|x)$：当观测向量为x值时, 是鲈鱼的概率。
  - $P(\omega_2|x)$：当观测向量为x值时, 是鲑鱼的概率。



4.2 基于最小错误率的贝叶斯决策

◆基于后验概率的决策规则：

存在一个观察值x(特征)

如果$P(\omega_1 \mid x) > P(\omega_2 \mid x)$ 类别状态= ω_1

如果$P(\omega_1 \mid x) < P(\omega_2 \mid x)$ 类别状态 = ω_2

因此，无论何时观测到某一个特定值x，概率误差为：

$P(error|x)=P(\omega_1|x)$ 判定为ω_2 (错误选择ω_1);

$P(error|x)=P(\omega_2|x)$ 判定为ω_1(错误选择ω_2);

4.2 基于最小错误率的贝叶斯决策

◆基于后验分布的判别规则：

错误概率的最小化判定规则：

如果 $P(\omega_1|x)>P(\omega_2|x)$，判定为$\omega_1$否则，判定为$\omega_2$。

因此， $P(error \mid x) = \min [P(\omega_1 \mid x), P(\omega_2 \mid x)]$

(最大后验概率准则可以保证最小错误率， 所以又称最小错误率准则)

4.2 基于最小错误率的贝叶斯决策

◆基于最小错误率的贝叶斯决策：

$$\boxed{\begin{array}{ll} P(\omega_1|x) > P(\omega_2|x) & x\in\omega_1 \\ P(\omega_1|x) < P(\omega_2|x) & x\in\omega_2 \end{array}}$$

等价形式

$$(1) P(\omega_i|x) = \max_{j=1,2} P(\omega_j|x) \qquad x\in\omega_i$$

$$(2) p(x|\omega_i)P(\omega_i) = \max_{j=1,2} p(x|\omega_j)P(\omega_j) \qquad x\in\omega_i$$

$$(3) l(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} \gtrless \frac{P(\omega_2)}{P(\omega_1)} \qquad x\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

似然比形式

$$(4) \ln[l(x)] = \ln p(x|\omega_1) - \ln p(x|\omega_2) \gtrless \ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) \qquad x\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

4.2 基于最小错误率的贝叶斯决策

例：假设在某个局部地区细胞识别中正常和异常两类的先验概率分别为

正常状态：$P(\omega_1)=0.9$

异常状态：$P(\omega_2)=0.1$

现有一待识别的细胞，其观察值为x，类条件概率密度分别为$p(x\mid\omega_1)=0.2, p(x\mid\omega_2)=0.4$，试对该细胞$x$进行分类。

解：

$$P(\omega_1\mid x)=\frac{p(x\mid\omega_1)P(\omega_1)}{\sum_{j=1}^{2}p(x\mid\omega_j)P(\omega_j)}=\frac{0.2\times0.9}{0.2\times0.9+0.4\times0.1}=0.818$$

$$P(\omega_2\mid x)=1-P(\omega_1\mid x)=0.182$$

$$P(\omega_1\mid x)=0.818>P(\omega_2\mid x)=0.182 \qquad \therefore x\in\omega_1$$

4.2 基于最小错误率的贝叶斯决策

最小错误率的讨论

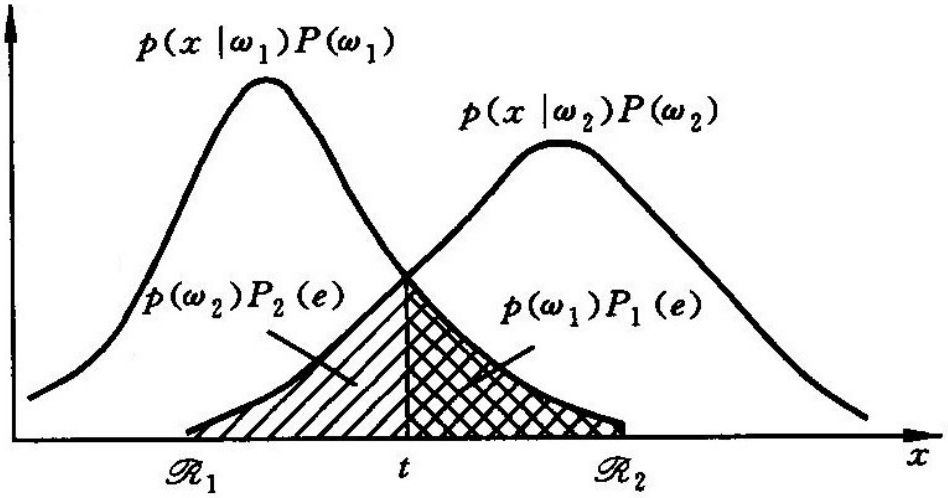
- 以一维情况为例讨论基于最小错误率的贝叶斯 决策确实对应最小错误率
  - 统计意义上的错误率，即平均错误率，用***P*(e)**表示

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

其中，$P(e|x)=\begin{cases}P(\omega_1|x) & \text{当} P(\omega_2|x)>P(\omega_1|x)\\ P(\omega_2|x) & \text{当} P(\omega_1|x)>P(\omega_2|x)\end{cases}$

4.2 基于最小错误率的贝叶斯决策

最小错误率的讨论



$$P(e)=\int_{-\infty}^{\infty}P(e\mid x)p(x)dx$$

$$=\int_{-\infty}^{t}P(\omega_2\mid x)p(x)dx+\int_{t}^{\infty}P(\omega_1\mid x)p(x)dx$$

$$=\int_{-\infty}^{t}p(x\mid\omega_2)P(\omega_2)dx+\int_{t}^{\infty}p(x\mid\omega_1)P(\omega_1)dx$$

$$=P(\omega_2)P_2(e)+P(\omega_1)P_1(e)$$

4.2 基于最小错误率的贝叶斯决策

最小错误率的讨论

- 最大后验概率保证了选取到t*，同时t*保证了 最小错误率；
- 每个样本的误差最小，保证了整个空间误差最 小；
- 最小错误率所对应的分类器是最优分类器；
- 为什么存在基于最小错误率的最优分类器，还 需要设计其他分类器呢？

样本有限情况下，类条件概率密度形式、参数估 计都不准确。

4.2 基于最小错误率的贝叶斯决策

C类别情况下最下错误率

- 在C类别情况下最小错误率贝叶斯决策规则的后验 概率形式：

$$P(\omega_i|x) = \max_{j=1,2,\cdots,c} P(\omega_j|x)$$

- 先验概率与类条件概率密度相联系的形式 ：

$$P(x|\omega_i)P(\omega_i) = \max_{j=1,2,\cdots,c} P(x|\omega_j)P(\omega_j), \text{则} \quad x \in \omega_i$$

4.2 基于最小错误率的贝叶斯决策

小结

贝叶斯公式：

$$P(\omega_i \mid x) = \frac{p(x \mid \omega_i) P(\omega_i)}{\sum_{j=1}^{2} p(x \mid \omega_j) P(\omega_j)} \qquad i = 1,2$$

基于最小错误率的贝叶斯决策规则：

$$\boxed{\begin{array}{ll} P(\omega_1 \mid x) > P(\omega_2 \mid x) & x \in \omega_1 \\ P(\omega_1 \mid x) < P(\omega_2 \mid x) & x \in \omega_2 \end{array}}$$

第四章　贝叶斯决策理论

4.1引言

4.2基于最小错误率的贝叶斯决策

4.3基于最小风险的贝叶斯决策

4.4分类器、判别函数及决策面

4.5正态分布下的统计决策

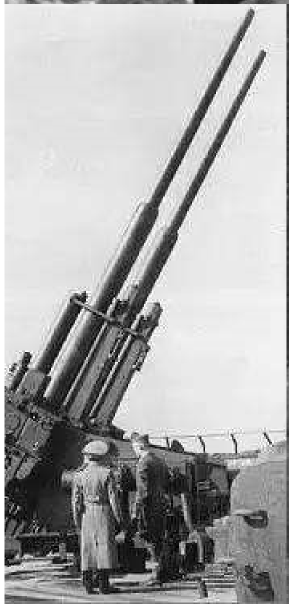
4.3 基于最小风险的贝叶斯决策

主要思想：

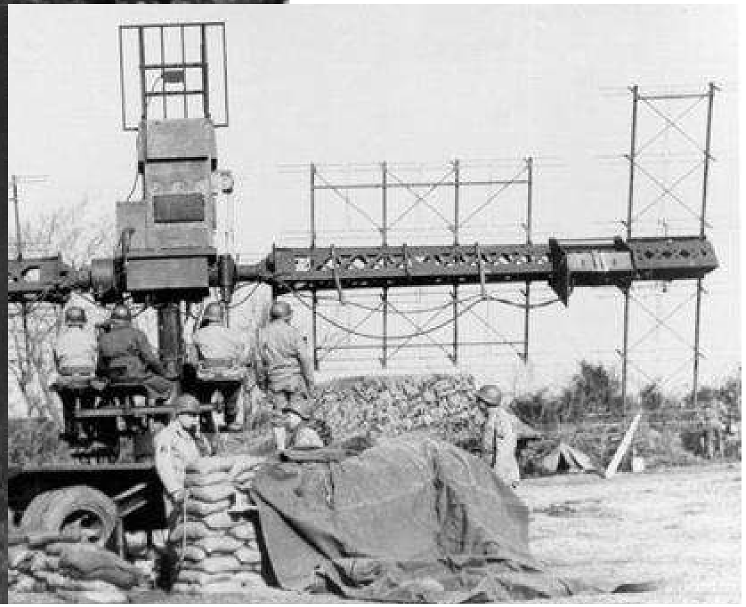
◆在最小错误率决策中，使错误率达到最小是 重要的。但实际上，有时候需要考虑一个比错 误率更广泛的概念—风险，而风险又是和损失 紧密相连的。

◆我们对样本的分类不仅要考虑到尽可能作出 正确的判断，而且还要考虑到作出错误判断时 会带来什么后果。

◆最小风险贝叶斯决策正是考虑各种错误造成损失不同而提出的一种决策规则。




LuftArchiv.de



4.3 基于最小风险的贝叶斯决策

两分类问题下的最小风险准则

决策行动：α_1 ：对应于类别判别ω_1； α_2:对应于类别判别ω_2

损失：$\lambda(\alpha_i,\omega_j)=\lambda_{ij}$ 表示当实际类别为ω_j时误判为ωi所引起的损失。

条件风险（条件期望损失）：

$$R(\alpha_1\,|\,x)=\lambda_{11}P(\omega_1\,|\,x)+\lambda_{12}P(\omega_2\,|\,x)$$

$$R(\alpha_2\,|\,x)=\lambda_{21}P(\omega_1\,|\,x)+\lambda_{22}P(\omega_2\,|\,x)$$

最小风险决策规则：

如果 $R(\alpha_1\,|\,x)<R(\alpha_2\,|\,x)$ ，则根据决策行动α_1，判决类别ω_1。

4.3 基于最小风险的贝叶斯决策

似然比形式

$$R(\alpha_1 \mid x) = \lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x) \qquad R(\alpha_2 \mid x) = \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

$R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$ 等价于：

$$\lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x) < \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

$$(\lambda_{12} - \lambda_{22}) P(\omega_2 \mid x) < (\lambda_{21} - \lambda_{11}) P(\omega_1 \mid x)$$

$$(\lambda_{12} - \lambda_{22}) \frac{p(x \mid \omega_2) P(\omega_2)}{p(x)} < (\lambda_{21} - \lambda_{11}) \frac{p(x \mid \omega_1) P(\omega_1)}{p(x)}$$

$$(\lambda_{12} - \lambda_{22}) p(x \mid \omega_2) P(\omega_2) < (\lambda_{21} - \lambda_{11}) p(x \mid \omega_1) P(\omega_1)$$

$$\frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} > \boxed{\frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \times \frac{P(\omega_2)}{P(\omega_1)}}$$

与x无关，对于某个问题，是个可以事先计算的常量。

似然比大于某个阈值，则采取行动决策α_1(判决ω_1)；否则为：ω_2

4.3 基于最小风险的贝叶斯决策

两分类问题下的最小风险准则

$$(\lambda_{12}-\lambda_{22})P(\omega_2\mid x)<(\lambda_{21}-\lambda_{11})P(\omega_1\mid x) \qquad \omega_1$$

$$(\lambda_{12}-\lambda_{22})P(\omega_2\mid x)>(\lambda_{21}-\lambda_{11})P(\omega_1\mid x) \qquad \omega_2$$

在两类问题中，若有 $\boxed{\lambda_{12}-\lambda_{22}=\lambda_{21}-\lambda_{11}}$ ，决策规则变为

$$P(\omega_2\mid x)<P(\omega_1\mid x) \qquad \omega_1$$

$$P(\omega_2\mid x)>P(\omega_1\mid x) \qquad \omega_2$$

4.3 基于最小风险的贝叶斯决策

多类问题下的最小风险准则

在c个类别的问题中，如果损失函数为“0-1”损 失函数：

$$\lambda(\alpha_i/\omega_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases} \qquad i,j=1,\ldots,c \qquad \alpha_i,\ \ i=1,2,\cdots,c$$

“0-1”损失函数：

1　对于c类问题只有c个决策，

2　实际类别 ω_j 正确判定为第j类时，损失为0。

3　实际类别 ω_j 误判为第 $i\neq j$ 类时，损失均为1。

4.3 基于最小风险的贝叶斯决策

“0-1” 损失函数下的最小风险准则

$$R(\alpha_i|\mathbf{x})=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j|\mathbf{x})=\sum_{\substack{j=1\\i\neq j}}^{c}P(\omega_j|\mathbf{x})$$

$$\sum_{j=1}^{c}P(\omega_j|\mathbf{x})=\frac{\sum_{j=1}^{c}p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{i=1}^{c}p(\mathbf{x}|\omega_i)P(\omega_i)}=1$$

$$R(\alpha_k|\mathbf{x})=\boxed{\min_{i=1,2,\cdots,c}R(\alpha_i|\mathbf{x})}$$

$$=\min_{i=1,2,\cdots,c}\sum_{\substack{j=1\\j\neq i}}^{c}P(\omega_j|\mathbf{x})$$

$$=\min_{i=1,2,\cdots,c}\left[1-P(\omega_i|\mathbf{x})\right]=\boxed{\max_{i=1,2,\cdots,c}P(\omega_i|\mathbf{x})}$$

最小错误率贝叶斯决策是在0-1损失函数条件下的最小风险贝叶斯决策，最小错误率贝叶斯决策是最小风险贝叶斯决策的特例。

第四章　贝叶斯决策理论

4.4 分类器、判别函数及决策面

判别函数（Discriminant Function）：

◆用于表示决策规则的某些函数$g_i(x)$称为判别函数。每个类别对应一个判别函数

$$g_i(x),\ i=1,2,\cdots,c$$ 。

◆通常定义一组判别函数， $g_i(x),\quad i=1,2,\cdots,c$ 用于表示多类决策规则。

◆如果使 $g_i(x)>g_j(x)$ 对于一切$i\neq j$ 成立，则将**x**归于 ω_i 类。

4.4 分类器、判别函数及决策面

决策面（Decision Surface）：

对于c类分类问题，按照决策规则可以把d维特征空间分成c个决策域，将划分决策域的边界面称为决策面，在数学上用解析形式可以表示成决策面方程。

◆　判决区域R_i是特征空间中的一个子空间，判决规则将所有落入R_i的样本x分类为类别ω_i；

◆判决边界是特征空间中划分判决区域的（超）平面；

◆　在判决边界上，通常有两类或多类的判别函数值相等。

4.4 分类器、判别函数及决策面

分类器设计（Classifier）：

分类器的功能是先计算出c个判别函数，再从中选择出对应于判别函数为最大值的类作为决策结果。

分类器设计核心设计判别函数，求出判定面方程！

分类器最常用的表述方式为判别函数：

$$g_i(x), \quad i=1,2,\cdots,c$$

每个类别对应一个判别函数。

基于判别函数的判决：

如果：$g_i(x) > g_j(x), \quad \forall i,j$ ，则属于 ω_i

决策面方程： $g_i(x) = g_j(x)$

4.4 分类器、判别函数及决策面

基于最小错误率的判决函数

$$g_i(x) = p(\omega_i | x)$$

$$g_i(x) = p(x | \omega_i) p(\omega_i)$$

$$g_i(x) = \ln p(x | \omega_i) + \ln p(\omega_i)$$

基于最小风险的判决函数

$$g_i(x) = -R(\alpha_i | x)$$

4.4 分类器、判别函数及决策面

两分类下的判别函数

- 特殊的，对于两分类问题，也可以只用一个判别函数

令： $g(x) = g_1(x) - g_2(x)$

- **判决规则**

如果： $g(x) > 0$ 则模式为 ω_1 否则为 ω_2

- **例如：**

$$g(x) = p(\omega_1|x) - p(\omega_2|x) \qquad g(x) = \ln\frac{p(x|\omega_1)}{p(x|\omega_2)} + \ln\frac{p(\omega_1)}{p(\omega_2)}$$

$$g(x) = p(x|\omega_1)p(\omega_1) - p(x|\omega_2)p(\omega_2)$$

- **决策面：** $g(x) = 0$

4.4 分类器、判别函数及决策面

两分类下的判别函数

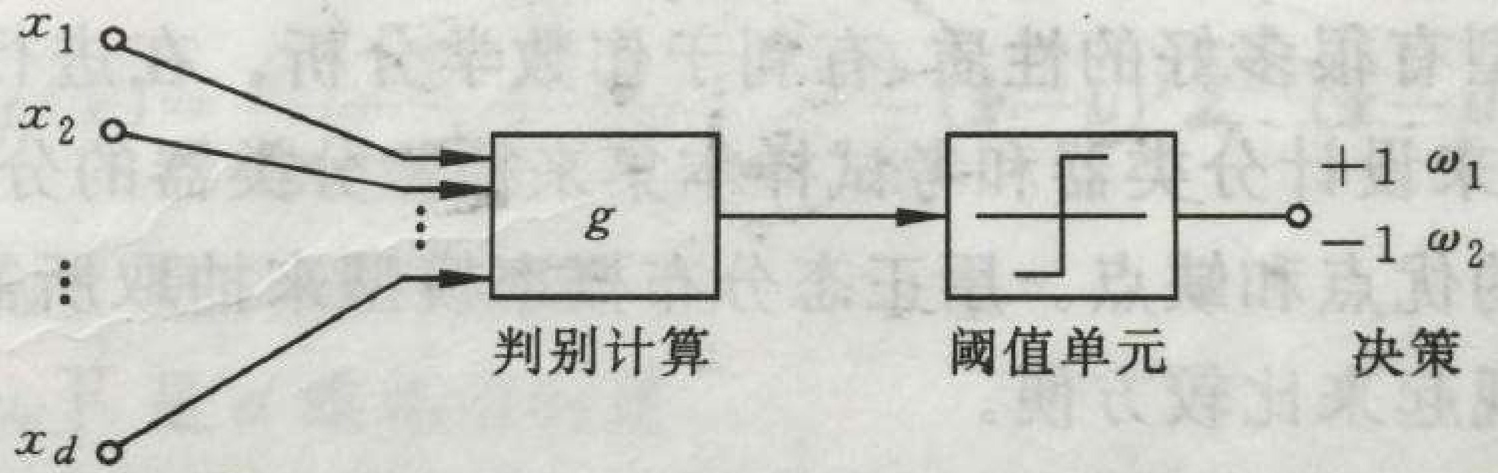


图 2.7　两类分类器的构成

4.4 分类器、判别函数及决策面

例子

例 2.1 假设细胞识别中正常(ω_1)和异常(ω_2)两类的先验概率分别为 0.9 和 0.1。现有一待识别的细胞，其观测值为 $\boldsymbol{X}$，从类条件概率密度分布曲线上查得 $P(\boldsymbol{X}|\omega_1)=0.2$，$P(\boldsymbol{X}|\omega_2)=0.4$，并且有表 2.2 的决策表。试利用最小错误率贝叶斯决策和最小风险贝叶斯决策进行分类。

表 2.2 决策表

决策 \ 损失 \ 状态	ω_1	ω_2
α_1	0	6
α_2	1	0

求：利用最小错误率和最小风险决策分别写出判别函数和决策面方程。

利用最小错误率决策，其对应的判别函数为：

$$g(x)=p(x|\omega_1)P(\omega_1)-p(x|\omega_2)P(\omega_2)$$
$$=0.9p(x|\omega_1)-0.1p(x|\omega_2)$$

决策面方程为： $9p(x|\omega_1)-p(x|\omega_2)=0$

利用最小风险决策，其对应的判别函数为：

$$g(x)=R(\alpha_2|x)-R(\alpha_1|x)=\lambda_{21}P(\omega_1|x)-\lambda_{12}P(\omega_2|x)$$
$$=\lambda_{21}p(x|\omega_1)P(\omega_1)-\lambda_{12}p(x|\omega_2)P(\omega_2)$$
$$=0.9p(x|\omega_1)-0.6p(x|\omega_2)$$

决策面方程为： $9p(x|\omega_1)-6p(x|\omega_2)=0$

4.4 分类器、判别函数及决策面

多分类下的判别函数

判决函数：

$$g_i(x) = p(\omega_i \mid x)$$

$$g_i(x) = p(x \mid \omega_i) p(\omega_i)$$

$$g_i(x) = \ln p(x \mid \omega_i) + \ln p(\omega_i)$$

决策面：

$$g_i(x) > g_j(x), \quad \forall j \neq i$$

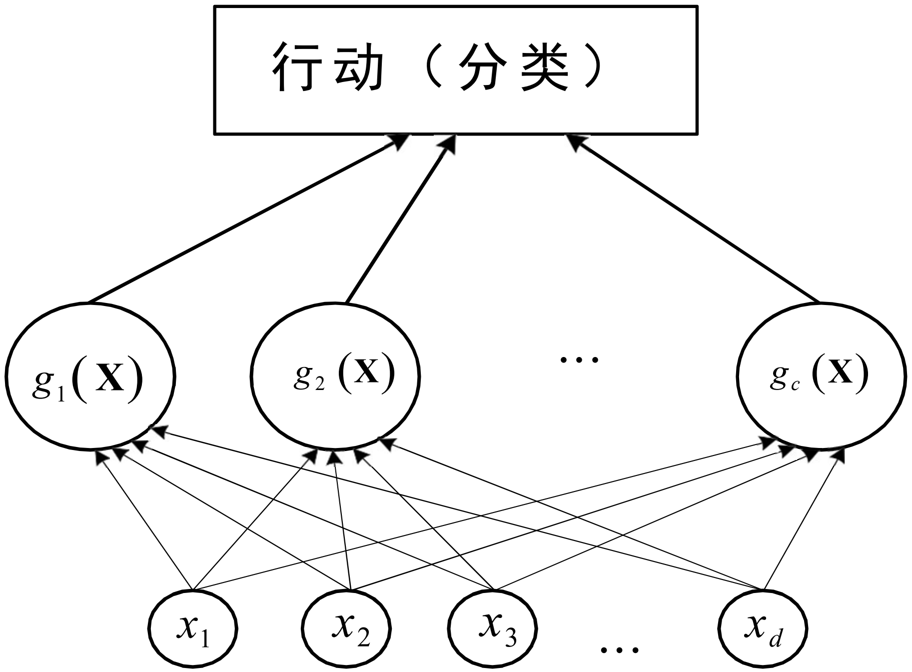
则模式为： ω_i

X	决策面
一维	分界点
二维	曲线
三维	曲面
d维（d>3）	超曲面

4.4 分类器、判别函数及决策面

多分类下的判别函数

分类器设计：它的功能是先计算出c个判别函数g_i，再从中选出对应于判别函数为最大值的类作为决策结果。



4.4 分类器、判别函数及决策面

判别函数、决策面

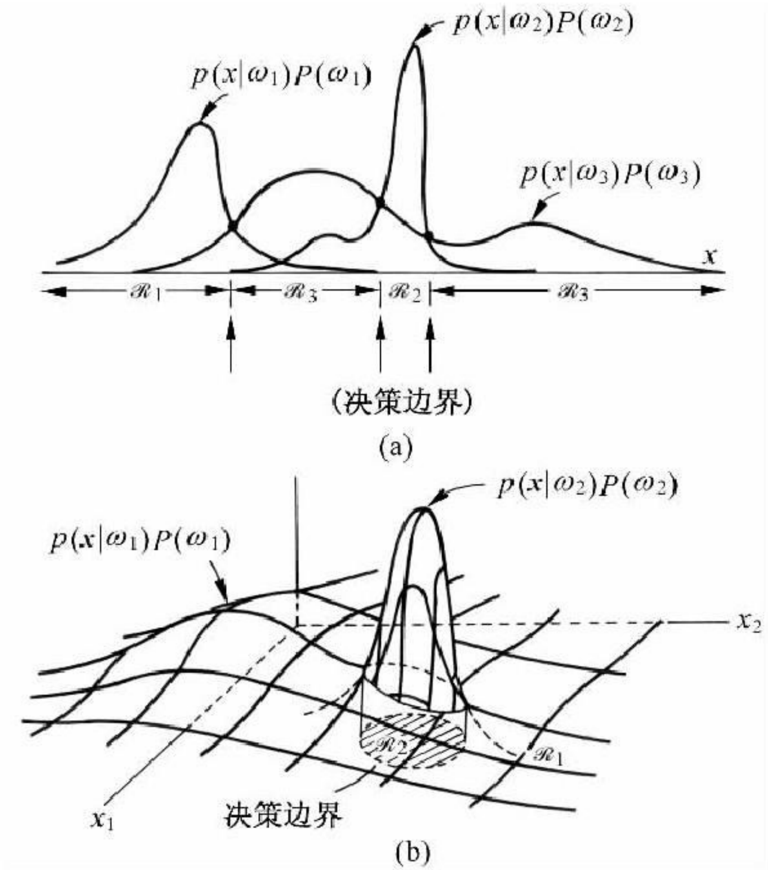
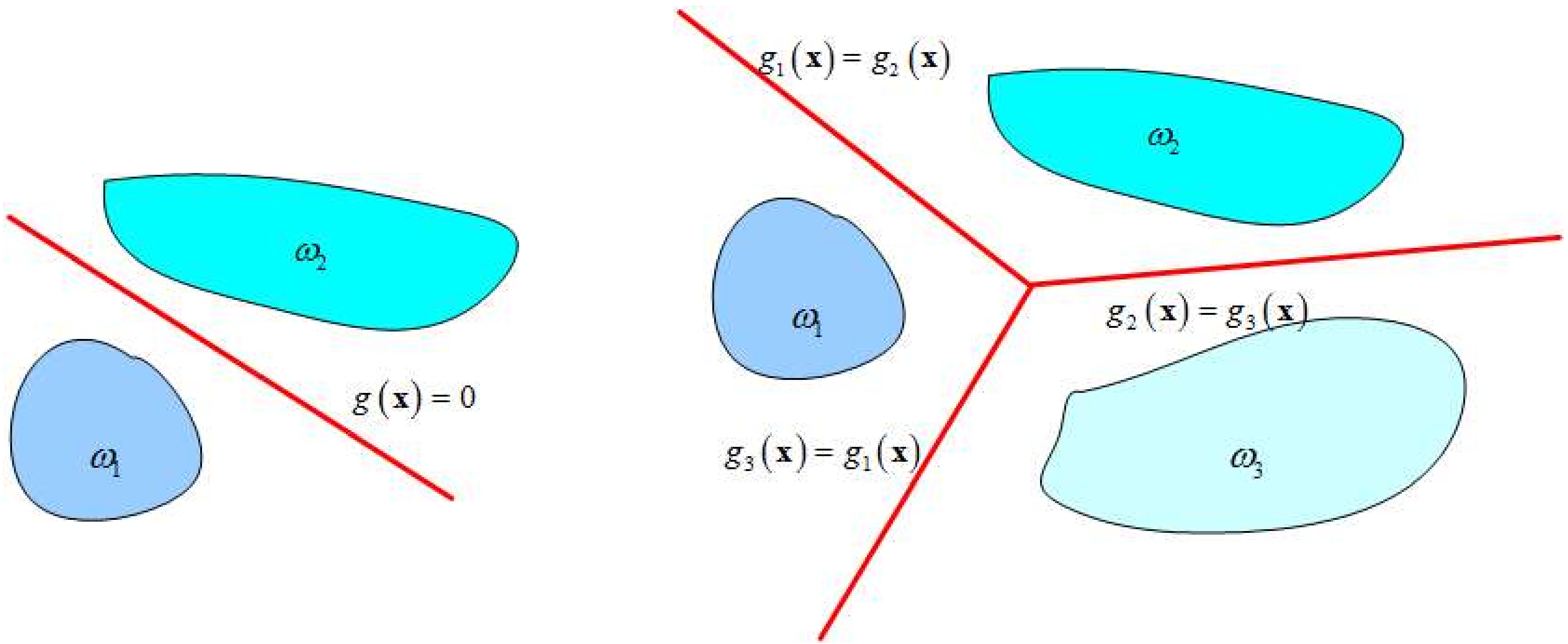


图 2.5 决策面

4.4 分类器、判别函数及决策面

判别函数，决策面



第四章　贝叶斯决策理论

4.5 正态分布下的统计决策

◆为什么研究正态分布？

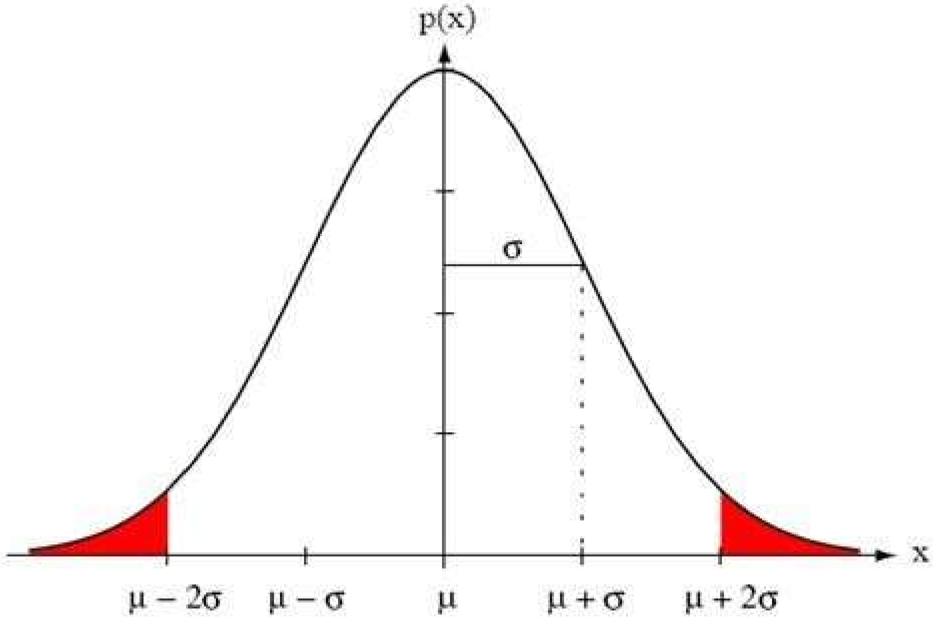
- 物理上的合理性：较符合很多实际情况，观测值通常是很多种因素共同作用的结果；
- 根据中心极限定理（这组定理是数理统计学和误差分析的理论基础，指出了大量随机变量累积分布函数逐点收敛到正态分布的积分布函数的条件），服从正态分布。
- 数学上比较简单：参数个数少

◆单变量正态分布

◆多元正态分布

4.5 正态分布下的统计决策

单变量正态分布



$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$ 记作 $N(\mu,\sigma^2)$

$$\mu = E\{x\} = \int x\, p(x)dx \qquad \sigma^2 = E\left\{(x-\mu)^2\right\} = \int (x-\mu)^2 p(x)dx$$

4.5 正态分布下的统计决策

多变量正态分布

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

$$\mathbf{x}=\left(x_1,\ \ x_2,\cdots,\ \ x_d\right)^T$$

期望(均值向量)

$$\mu=E(\mathbf{x})=\left(\mu_1,\ \ \mu_2,\cdots,\ \ \mu_d\right)^T$$

协方差矩阵

(对称非负定)

二次型$\mathbf{x}^T\Sigma\mathbf{x}\geq 0$

$$\Sigma=E\left((x-\mu)(x-\mu)^T\right)=\left[\sigma_{ij}^2\right]_{d\times d}$$

$$\sigma_{ij}=E\left(\left(\mathbf{x}_i-E(\mathbf{x}_i)\right)\left(\mathbf{x}_j-E(\mathbf{x}_j)\right)\right)$$

4.5 正态分布下的统计决策

多元正态分布的性质

- 参数个数：**d+d(d+1)/2**
  - 均值向量：**d**个参数
  - 协方差矩阵：对称的**d**维矩阵，**d(d+1)/2**个参 数
- 等密度点的轨迹为一超椭球面

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

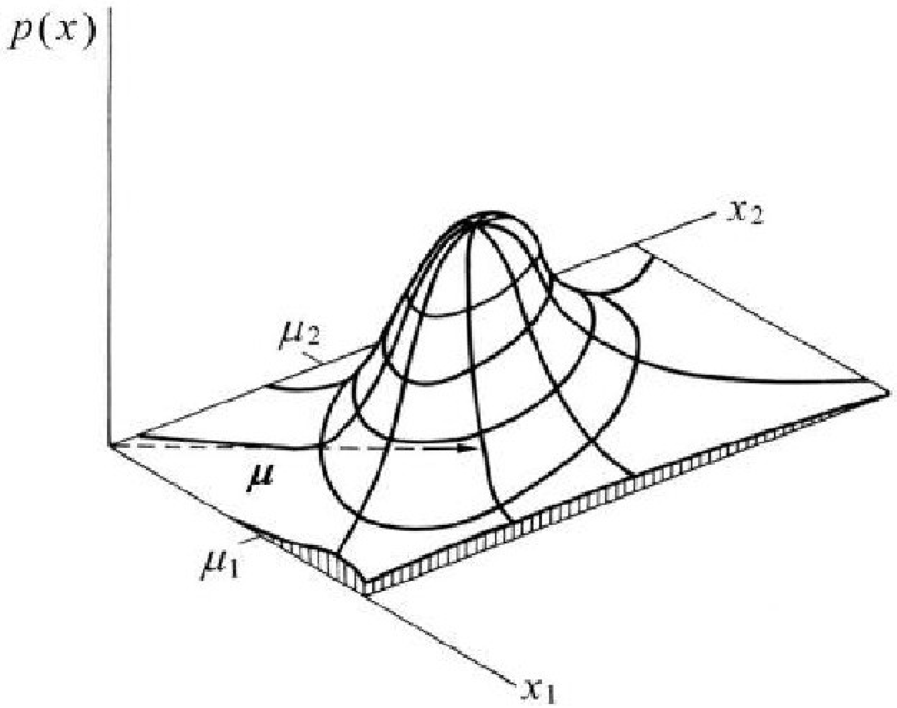
要使密度**p(x)**值不变，需指数项为常数，即：

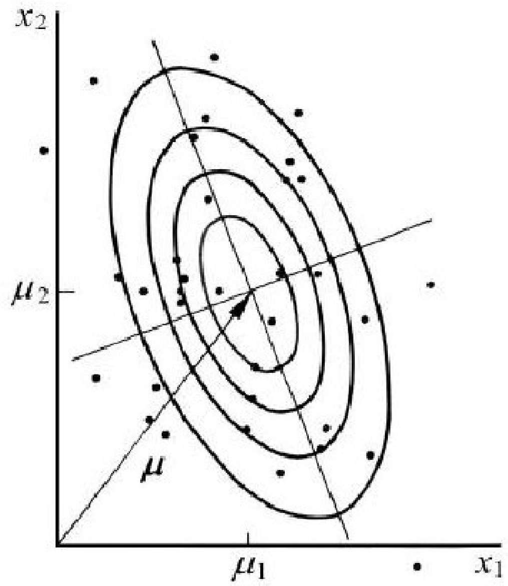
$$(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu) = \text{常数}$$

超椭球面

4.5 正态分布下的统计决策

多元正态分布的性质





Σ 的特征向量决定了主轴方向，主轴长度与 Σ 的本征值成正比。

4.5 正态分布下的统计决策

多元正态分布的性质

- 马氏距离：

$$\gamma^2 = (X-\mu)^T \Sigma^{-1}(X-\mu) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}(x_i-\mu_i)(x_j-\mu_j) \geq 0$$

与 欧式距离：

$$(x-\mu)^T(x-\mu)$$

马氏距离考虑数据各个维度间的相关性，x到μ的马 氏距离为常数时，所组成的超椭球面为等密度点。

4.5 正态分布下的统计决策

- 根据最小错误率贝叶斯判别函数，在多元正态概型 $(p(\mathbf{x}|\omega_i)\sim N(\boldsymbol{\mu}_i, \Sigma_i), i=1,\ldots, c)$ 下就可以立即写出其相应的表达式。

$$p\left(x|\omega_i\right)=\frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right)$$

◆判别函数为： $g_i\left(x\right)=\ln p\left(x|\omega_i\right)+\ln p\left(\omega_i\right)$

$$g_i(\mathbf{x})=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\sum\nolimits_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\sum\nolimits_i|+\ln P(\omega_i) \qquad \textbf{(1)}$$

◆决策面方程为： $\boldsymbol{g_i}(\mathbf{x})-\boldsymbol{g_j}(\mathbf{x})=0$ 即

$$-\frac{1}{2}[(\mathbf{x}-\boldsymbol{\mu}_i)^T\sum\nolimits_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)-(\mathbf{x}-\boldsymbol{\mu}_j)^T\sum\nolimits_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)]-\frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|}+\ln\frac{P(\omega_i)}{P(\omega_j)}=0$$

4.5 正态分布下的统计决策

- 情况一：各类协方差阵相等，且每类各特征 独立，方差相等 （对角矩阵）

$$\Sigma_1 = \cdots = \Sigma_c = \sigma^2 \mathbf{I}$$

- 情况二：各类协方差阵相等

$$\Sigma_1 = \ldots = \Sigma_c = \Sigma$$

- 情况三：各类协方差阵不相等

任意的 Σ

4.5 正态分布下的统计决策

情况一：$\Sigma_1 = \cdots = \Sigma_c = \sigma^2 \mathrm{I}$

将 $\boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma^2}\mathbf{I}$ $|\Sigma_i| = \sigma^{2d}$ 代入

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

得到决策函数

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}\|\mathbf{x}-\mu_i\|^2 + \ln P(\omega_i) + \text{const}$$

展开决策函数

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x} + \frac{1}{\sigma^2}\mu_i^T\mathbf{x} - \frac{1}{2\sigma^2}\mu_i^T\mu_i + \ln P(\omega_i)$$

其中，二次项 $\mathbf{x}^T\mathbf{x}$ 对所有的 i 是相等的

4.5 正态分布下的统计决策

因此，等价的判决函数为：

$$g_i(\mathbf{x}) = \frac{1}{\sigma^2}\mu_i^T\mathbf{x} - \frac{1}{2\sigma^2}\mu_i^T\mu_i + \ln P(\omega_i) = w_{i1}^T\mathbf{x} + w_{i0}$$

其中： $w_{i1}^T = \dfrac{1}{\sigma^2}\mu_i$ $\qquad w_{i0} = -\dfrac{1}{2\sigma^2}\mu_i^T\mu_i + \ln P(\omega_i)$

此时，写成了一个线性判别函数的形式。

决策面 $g_i(\mathbf{x}) = g_j(\mathbf{x})$ 可以写成：

$$w^T(\mathbf{x} - \mathbf{x}_0) = 0$$

过$\mathbf{x}_0$与w正交的超平面

其中： $w = \mu_i - \mu_j$

$$\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2}\ln\frac{p(\omega_i)}{p(\omega_j)}(\mu_i - \mu_j)$$

4.5 正态分布下的统计决策

当 $p(\omega_i)=p(\omega_j)$ ，$\mathbf{x}_0=\frac{1}{2}(\mu_i+\mu_j)$ 位于两中心的中点；

当 $p(\omega_i)\neq p(\omega_j)$ ，$\mathbf{x}_0$ 向先验概率小的方向偏移。

在先验概率相等的情况下，最优判决的规则为：

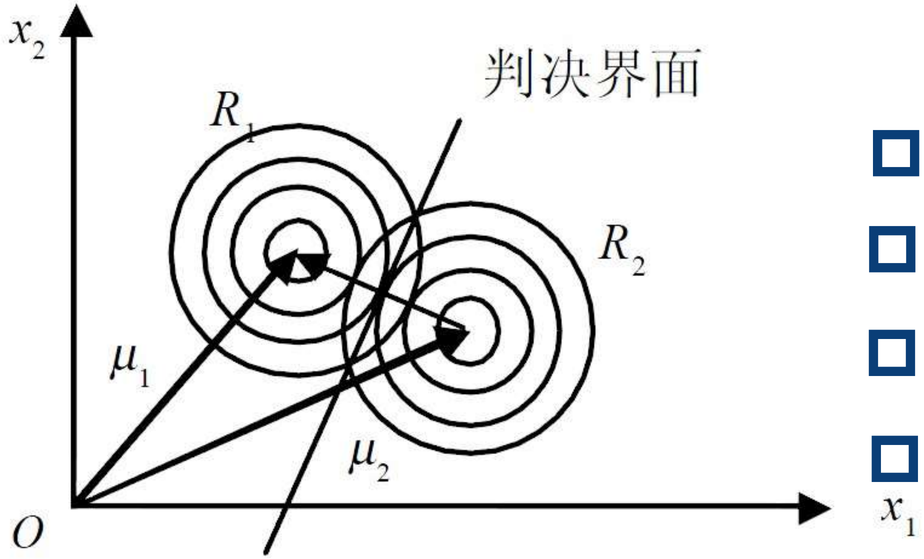
为将某特征向量x归类，通过测量每一x到c个均值向量中心的每一个欧氏距离，并将x归为离它最近的那一类。这样的分类器称为“最小距离分类器”。

4.5 正态分布下的统计决策

最小距离分类器

$$w^T(\mathbf{x}-\mathbf{x}_0)=0 \qquad w=\mu_i-\mu_j$$

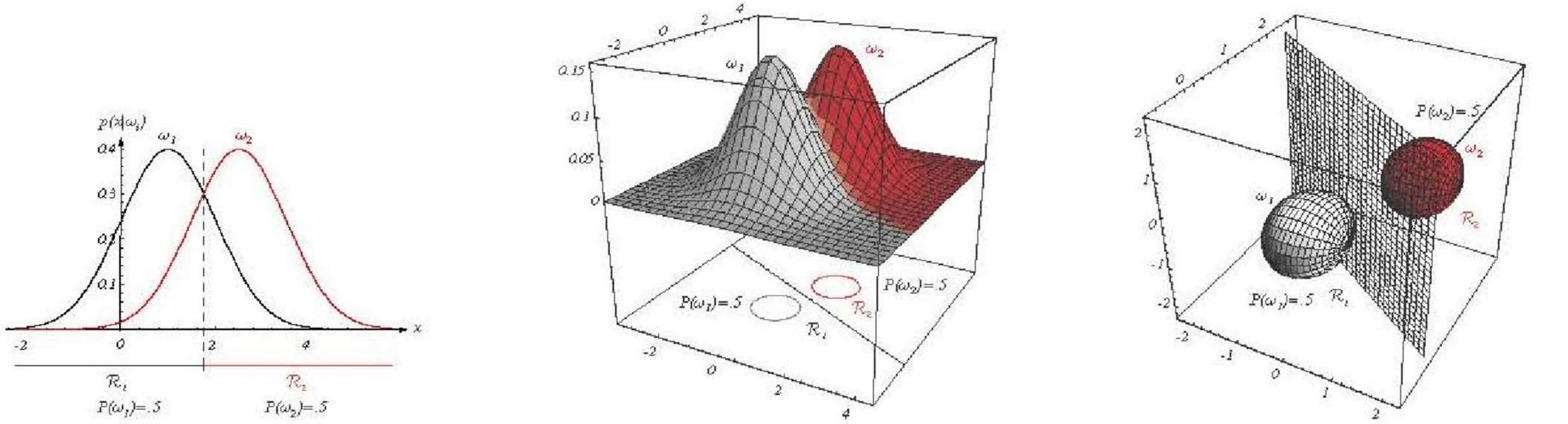
- 上述结果表示在二维特征空间里，如下图所示：



- 两类判决面与 $\mu_1-\mu_2$ 垂直，其交点为 x_0
- $P(\omega_1)=P(\omega_2)$ 时 x_0 为 $\mu_1-\mu_2$
- 的中点
- $P(\omega_1)\neq P(\omega_2)$ 时 x_0 向先验概率
- 较小类型的均值点偏移。
- 先验概率大，样本分布多，远离先验概率大的区域。

4.5 正态分布下的统计决策

□ 最小距离分类器



最小距离分类器 $p(\omega_i) = p(\omega_j)$

判决边界是**d-1**维超平面，垂直于两类中心的连线

4.5 正态分布下的统计决策

情况二： $\Sigma_1 = \cdots = \Sigma_c = \Sigma$

各类的协方差矩阵相等，在几何上，相当于各类样本 集中在以该类均值μ_i为中心的同样大小和形状的超椭 球内。

决策函数

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Σ 不变，与 $\boldsymbol{i}$ 无关：

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + \ln P(\omega_i)$$

4.5 正态分布下的统计决策

一个特例：当 $P(\omega_i) = P$ 时，各样本先验概率相等。

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$$

其中：

$$\gamma^2 = (x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$$

γ^2 为x到均值点 μ_i 的“马 氏 距 离”（**Mahalanobis**）的平方。

进一步简化： $g_i(x) = -\gamma^2 = -(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$

对于每类样本x，只要计算出x到每类的均值点μ_i的马氏距离平方，最后把x归于最小的类别。

4.5 正态分布下的统计决策

一般地，决策函数

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma^{-1} (\mathbf{x}-\mu_i) + \ln P(\omega_i)$$

展开决策函数

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} + \mu_i^T\Sigma^{-1}\mathbf{x} - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \ln P(\omega_i)$$

$\mathbf{x}^T\Sigma^{-1}\mathbf{x}$ 对所有的 ***i*** 是相等的，则

$$g_i(\mathbf{x}) = \mu_i^T\Sigma^{-1}\mathbf{x} - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \ln P(\omega_i) = w_{i1}^T\mathbf{x} + w_{i0}$$

其中： $w_{i1} = \Sigma^{-1}\mu_i$ $\quad w_{i0} = -\frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \ln P(\omega_i)$

4.5 正态分布下的统计决策

决策面 $g_i(\mathbf{x}) = g_j(\mathbf{x})$ 可以写成：

$$w^T(\mathbf{x}-\mathbf{x}_0)=0$$

过 $\mathbf{x}_0$ 与 w 正交的超平面

其中：

$$w = \Sigma^{-1}(\mu_i - \mu_j)$$

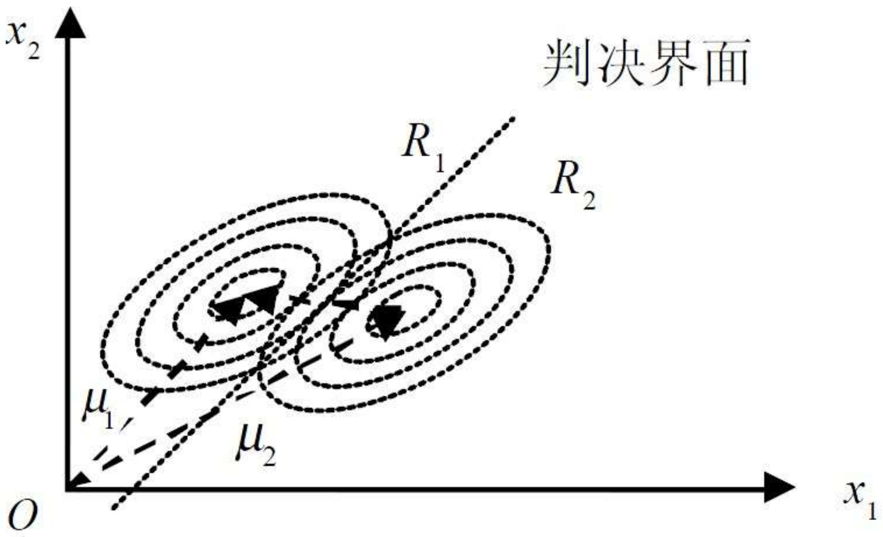
$$\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{1}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)} \ln \frac{P(\omega_i)}{P(\omega_j)} (\mu_i - \mu_j)$$

由于 w 并非沿着 $\mu_i - \mu_j$ 方向，因此分界面并非与均值间的连线垂直正交。

4.5 正态分布下的统计决策

- 上述结果表示在二维特征空间里，如下图所示：

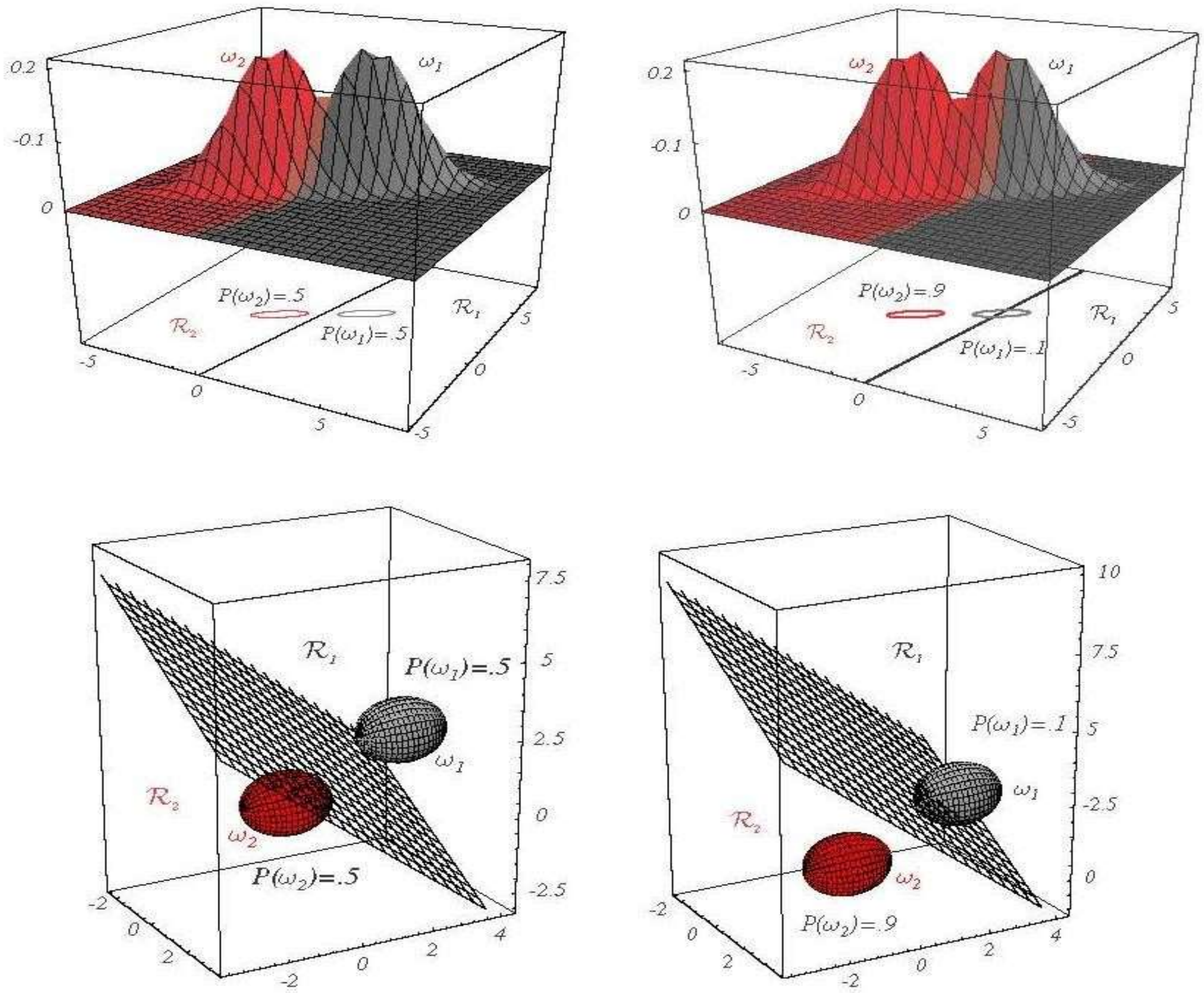
当各类先验概率相等时，判决面与 $\mu_i - \mu_j$ 的交点



$$x_0 = \frac{1}{2}(\mu_i + \mu_j)$$

当各类先验概率不相等 时，x_0 不在 $\mu_i - \mu_j$ 的中 点上，而是偏向先验概 率较小的均值点。

4.5 正态分布下的统计决策



$p(\omega_i) \neq p(\omega_j)$ 时

决策面向先验概率小的方向偏移

4.5 正态分布下的统计决策

情况三：任意的 Σ $\Sigma_i \neq \Sigma_j$

由于： $$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

去掉与i无关的项：

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

可以写为： $$g_i(x) = x^T w_{i2} x + w_{i1}^T x + w_{i0}$$ （二次型）

其中二次项，一次项系数和常数项分别为：

$$w_{i2} = -\frac{1}{2}\Sigma_i^{-1} \qquad w_{i1} = \Sigma_i^{-1}\mu_i$$

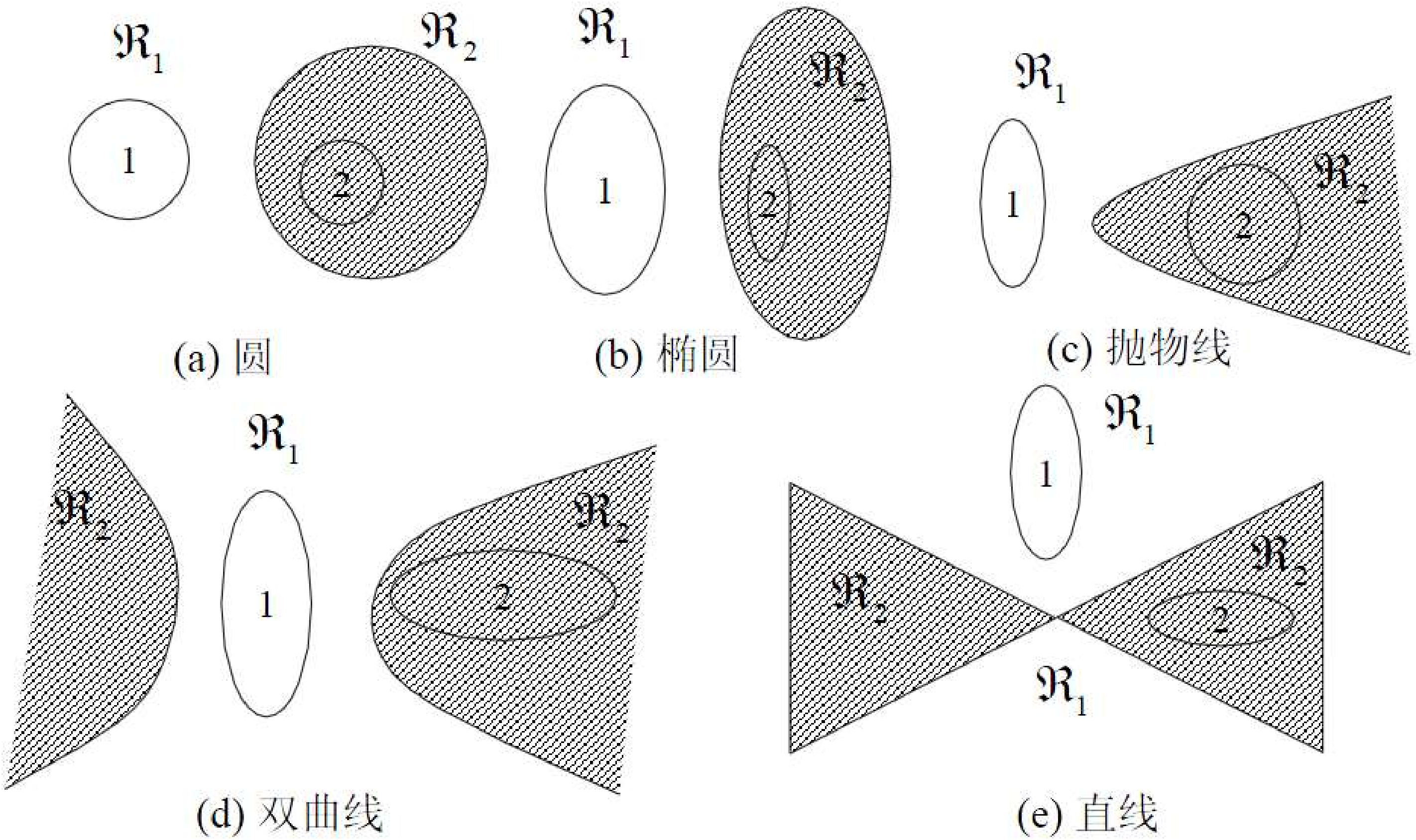
$$w_{i0} = -\frac{1}{2}\mu_i^T \Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

4.5 正态分布下的统计决策

- 判别函数$g_i(\mathbf{x})$表示为x的二次型。
- 若决策域R_i与R_j相邻，则决策面应满足
  - $g_i(\mathbf{x})-g_j(\mathbf{x})=0$
- 即 $\mathbf{x}^T(W_i-W_j)\mathbf{x}+(\mathbf{w}_i-\mathbf{w}_j)^T\mathbf{x}+w_{i0}-w_{j0}=0$
- 由上式所决定的决策面为超二次曲面，随着$\sum_i$， μ_i，$P(\omega_i)$的不同而呈现为某种超二次曲面，即 超球面、超椭球面、超抛物面、超双曲面或超 平面。

4.5 正态分布下的统计决策

❖ 上述结果表示在二维特征空间里，如下图所示：



各类协方差不同，决策面为超二次曲面。

小结

4.3 基于最小风险的贝叶斯决策

(1) x是d维随机向量

损失 \ 状态 / 决策	自然状态		
	ω_1	…	ω_c
α_1	$\lambda(\alpha_1, \omega_1)$	…	$\lambda(\alpha_1, \omega_c)$
⋮	⋮	$\lambda(\alpha_i, \omega_j)$	⋮
α_a	$\lambda(\alpha_a, \omega_1)$	…	$\lambda(\alpha_a, \omega_c)$

$$x = [x_1, x_2, \ldots, x_d]^T$$

(2) 状态空间Ω由c个自然状态(c类)组成：

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$$

(3) 决策/行动 α_1 指将模式x判定为ω_i或者是拒判。

决策空间是由a个决策组成 $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$

(4) 损失函数为 表示当真实状态为ω_j而所采取的决策为α_1时所带来的损失。

4.3 基于最小风险的贝叶斯决策

条件风险：

◆由于引入了“ 损失” 的概念，在考虑错判所 造成的损失时，就不能只根据后验概率的大小 来做决策，而必须考虑所采取的决策是否使损 失最小。

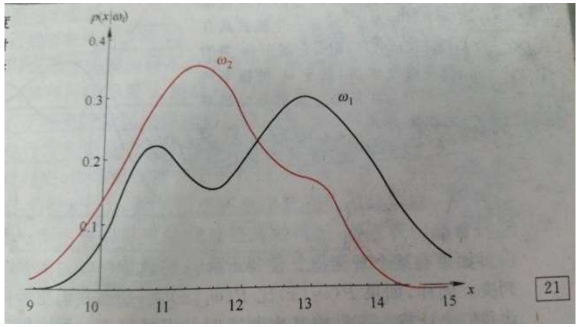
◆对于给定的x ，如果采取决策α_i ，从决策表 可见，λ可以在c个$\lambda(\alpha_i,\omega_j)$,j=1,2,…,c值中任取　一个,其相应概率为$P(\omega_j|x)$。 因此在采取决策 α_i 情况下的条件期望损失(也称为条件风险） $R(\alpha_i|x)$为：

$$R(\alpha_i / \mathrm{x}) = E[\lambda(\alpha_i, \omega_j)] = \sum_{j=1}^{c} \lambda(\alpha_i, \omega_j) P(\omega_j / x), \quad i = 1, 2, \ldots, C$$

4.3 基于最小风险的贝叶斯决策

- **期望风险：** 对于x的不同观察值，采取决策α_i时，其 条件风险大小是不同的。所以究竟采取哪一种决策 将随x的取值而定。这样，决策α可以看成随机向量x 的函数，记为α(x)。
- 可以定义期望风险R_{exp}为：

$$R_{exp} = \int R\left(\alpha(\mathbf{x})|\mathbf{x}\right) p(\mathbf{x}) d\mathbf{x}$$



- 期望风险反映对整个空间上所有x的取值采取相应的 决策α(x)所带来的平均风险。

4.3 基于最小风险的贝叶斯决策

决策规则：

◆在考虑错判带来的损失时，总是希望损失最 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。这就是最小风险贝叶斯 决策。

◆ 最小风险贝叶斯决策规则为：

$$If \quad R(a_k|x) = \min_{i=1,\ldots,a} \{R(a_k|x)\}$$

$$then \quad \alpha = \alpha_k$$

4.3 基于最小风险的贝叶斯决策

实现过程：

◆（1）已知先验概率和类条件概率，根据贝叶斯公式计算出后验概率；

$$P(\omega_i/x)=\frac{p(x\mid\omega_i)P(\omega_i)}{\sum_{i=1}^{c}p(x\mid\omega_i)P(\omega_i)}$$

◆（2）利用后验概率和决策表，计算采取每种决策的条件风险；

$$R(\alpha_i\mid \mathrm{x})=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j/x)$$

◆（3）比较各个条件风险的值，找出条件风险最小的决策。

$$R(\alpha_i\mid\mathrm{x})=\min_{i=1,2,\cdots,a}R(\alpha_i\mid\mathrm{x})$$

4.3 基于最小风险的贝叶斯决策

举例

决策	状态	
	ω_1	ω_2
α_1	0	6
α_2	1	0

例：在某个局部地区细胞识别中正常（ω_1）和异常（ω_2）两类的先验概率为：

$P(\omega_1)=0.9$，　$P(\omega_2)=0.1$，

满足：

$P(x/\omega_1)=0.2$，$P(x/\omega_2)=0.4$

对于未知细胞x，利用最小风险贝叶斯决策和最小错误率贝叶斯决策，问该细胞属于正常细胞还是异常细胞？

解：计算出后验概率

$$P(\omega_1/x)=0.818,\quad P(\omega_2/x)=0.182$$

4.3 基于最小风险的贝叶斯决策

举例

计算条件风险：

$$R(\alpha_1/x)=\sum_{j=1}^{2}\lambda_{1j}P(\omega_j/x)=\lambda_{12}P(\omega_2/x)=1.092$$

$$R(\alpha_2/x)=\sum_{j=1}^{2}\lambda_{2j}P(\omega_j/x)=\lambda_{21}P(\omega_1/x)=0.818$$

◆因为 $R(\alpha_1/x)>R(\alpha_2/x)$ ，决策为ω_2，即判别待识别细胞为异常细胞。

◆利用基于最小错误率的准则，判定为ω_1，这里损失函数起了决定性作用。

◆各种错误造成的损失不同，正常细胞判定为异常细胞的损失远大于异常判定为正常的损失。

分析：最小风险决策必须要有合适的损失函数λ，实际中要列出合适的决策表很不容易，往往要根据所研究的具体问题，分析错误决策造成损失的严重程度，与有关专家共同商讨来确定，才能做出更有效的决策。

答疑

1. 全概率是不是为1?

1.设某工厂有两个车间生产同型号家用电器，第一车间的次品率为 0.15，第二车间的次品率为 0.12，两个车间的成品都混合堆放在一个仓库，假设第 1,2 车间生产的成品比例为 2:3，今有一客户从成品仓库中随机提一台产品，求该产品合格的概率。

解：设 B={从仓库中随机提出的一台是合格品}

A_i={提出的一台是第 i 车间生产的}，i=1,2

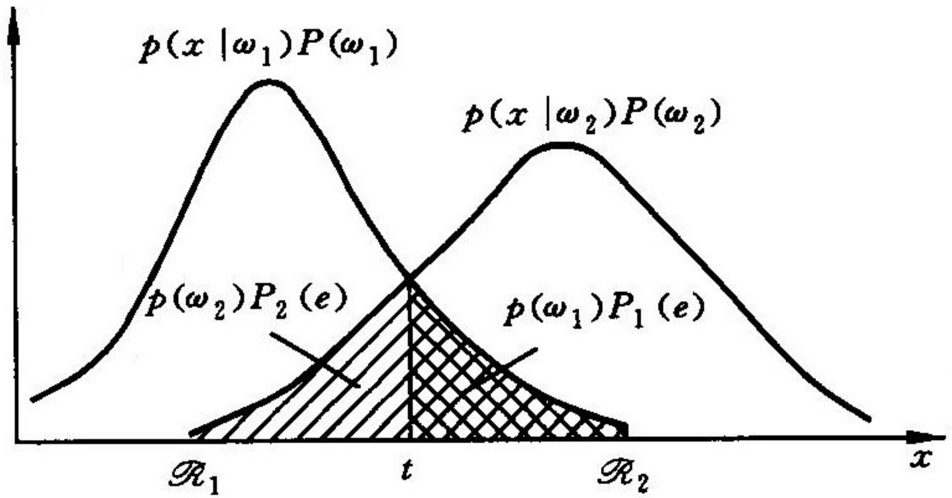
则有分解 $B=A_1B\cup A_2B$

由题意 P(A1)=2/5,P(A2)=3/5,P(B|A1)=0.85,P(B|A2)=0.88

由全概率公式 $P(B)= P(A_1)\ P(B|A_1)+ P(A_2)\ P(B|A_2)=0.4*0.85+0.6*0.88=0.868.$

答疑

2. 错误率图示及含义?



4.4 分类器、判别函数及决策面

判别函数（Discriminant Function）：

◆判别函数与决策面方程密切相关，且都由相应的决策规则所确定。

◆表达同样的判决规则可能采用不同的判别函数，只要满足如下条件：

例如：

$g_i(x)$ $\qquad$ $k\, g_i(x)$, k为正常 $g_i(x)+k$,

$g_i(x)$ $\qquad$ k为任意常数 $\ln(g_i(x))$

$g_i(x)$

用$f(g_i(x))$替换$g_i(x)$，其中$f(*)$为单调递增函数

End