Ch2

贝叶斯决策理论

任课教师：柳欣老师

email：starxliu@163.com

Bayes，Thomas(1702–1761)

贝叶斯是英国数学家.1702年生于伦敦；

1761年4月17日卒于坦布里奇韦尔斯.

贝叶斯是一位自学成才的数学家.曾助理宗教事务，后来长期担任坦布里奇韦尔斯地方教堂的牧师.1742年，贝叶斯被选为英国皇家学会会员.

如今在概率、数理统计学中以贝叶斯姓氏命名的有贝叶斯公式、贝叶斯风险、贝叶斯决策函数、贝叶斯决策规则、贝叶斯估计量、贝叶斯方法、贝叶斯统计等等.

2.1 引言

- 统计决策理论是处理模式分类问题的基本理论之一
- 贝叶斯（**Bayes**）决策理论是统计模式识别中一个基本方法

硬币识别
- 无观测：先验概率
- 有观测：后验概率

最小错误率

贝叶斯公式：$P(\omega_i \mid x) = \dfrac{p(x,\omega_i)}{p(x)} = \dfrac{P(\omega_i)p(x \mid \omega_i)}{p(x)}, i = 1,2$

2.1 引言

- **问题描述**
  - 样本与样本空间：d 维特征向量

$$x = [x_1, x_2, \cdots, x_d]^T \quad x \in R^d$$

  - 类别与类别空间： C 类

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_i \cdots, \omega_c\}$$

- **前提**

（1） 各类总体的概率分布是已知的；

（2）类别数是一定的

  - $P(\omega_i)$ 和$P(\mathbf{x}|\omega_i)$已知

- **现观察到一个特征向量x， x 分到哪一类最合理？？**

2.1 引言

贝叶斯推理就是在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对先验概率进行修正，最后再利用修正概率做出最优决策。

贝叶斯决策理论方法是统计决策中的一个基本方法，其基本思想是：

1、已知条件概率密度参数表达式和先验概率。

2、利用贝叶斯公式转换成后验概率。

3、根据后验概率大小进行决策分类。

2.1 引言

- **分类错误率**
  - 条件错误率（样本**x**上）为：

$$P(e|\mathbf{x}) = \begin{cases} P(\omega_2|\mathbf{x}) = 1 - P(\omega_1|\mathbf{x}) & 若决定 \mathbf{x} \in \omega_1 \\ P(\omega_1|\mathbf{x}) = 1 - P(\omega_2|\mathbf{x}) & 若决定 \mathbf{x} \in \omega_2 \end{cases}$$

  - （平均）错误率（所有独立样本上条件错误率的期望）：

$$P(e) = E(P(e|\mathbf{x})) = \int_{-\infty}^{\infty} P(e|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

2.2 几种常用的决策规则

2.2.1 基于最小错误率的贝叶斯决策

- 例子：癌细胞的识别
  - ω_1正常细胞，ω_2异常细胞（状态，随机变量）
  - 先验概率：$P(\omega_1),P(\omega_2)$， $P(\omega_1)+P(\omega_2)=1$
- 只使用先验概率决策的情况：

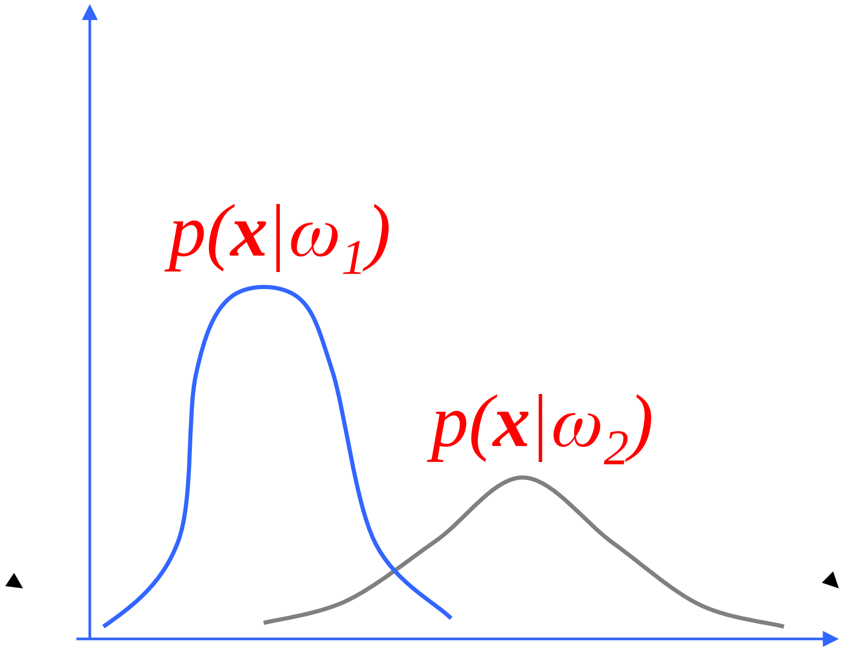
$$\left.\begin{array}{l} P(\omega_1) > P(\omega_2), x \in \omega_1 \\ P(\omega_1) < P(\omega_2), x \in \omega_2 \end{array}\right\}\text{这种分类器决策无意义}$$

奇怪的结果，需要利用观测信息（胞核总光密度，$d=1$）

$$P(\omega_1/x) \qquad P(\omega_2/x)$$

2.2.1 基于最小错误率的贝叶斯决策

- 类条件概率密度：$P(\mathbf{x}/\omega_i)$ i=1,2,…

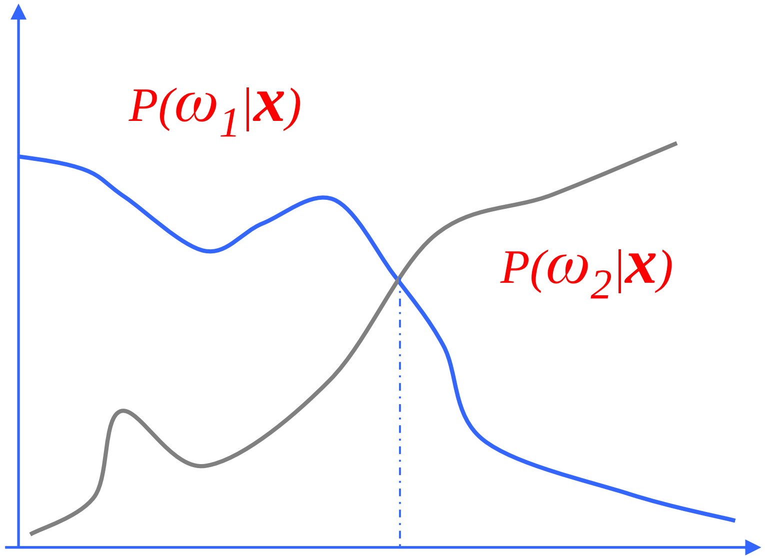


2.2.1 基于最小错误率的贝叶斯决策

- 利用**贝叶斯公式**

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

得到的条件概率称为状态的**后验概率**



2.2.1 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规则

$$\begin{cases} 若P(\omega_1/x) > P(\omega_2/x), 则x \in \omega_1 \\ 若P(\omega_1/x) < P(\omega_2/x), 则x \in \omega_2 \end{cases}$$

- 等价形式：

（1） If $P(\omega_i \mid x) = \max\limits_{j=1,2} P(\omega_j \mid x)$， then $x \in \omega_i$

（2） If $p(x \mid \omega_i) P(\omega_i) = \max\limits_{j=1,2} p(x \mid \omega_j) P(\omega_j)$, then $x \in \omega_i$

（3） If $l(x) = \dfrac{p(x \mid \omega_1)}{p(x \mid \omega_2)} \begin{matrix} > \\ < \end{matrix} \dfrac{P(\omega_2)}{P(\omega_1)}$, then $x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

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

If $h(x) \begin{matrix} < \\ > \end{matrix} \ln\left(\dfrac{P(\omega_1)}{P(\omega_2)}\right)$, then $x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

1， $l(x)$： 似然比， $\dfrac{P(\omega_2)}{P(\omega_1)}$： 似然比阈值， $h(x)$： 对数似然比

2.2.1 基于最小错误率的贝叶斯决策

有两个可选的假设：

病人有癌症(cancer)、病人无癌症(normal)

可用数据来自化验结果：正+和负-

有先验知识：在所有人口中，患病率是0.8%

对确实有病的患者的化验准确率为98%，

对确实无病的患者的化验准确率为97%，

总结如下

$P(cancer)=0.008$， $P(normal)=0.992$

$P(+|cancer)=0.98$， $P(-|cancer)=0.02$

$P(+|normal)=0.03$， $P(-|normal)=0.97$

问题：假定有一个新病人，化验结果为正，是否应将病人断定为有癌症？求后验概率$P(cancer|+)$和$P(normal|-)$

例子

因此极大后验假设计算如下：

$P(+|cancer)P(cancer)=0.00784$

$P(+|normal)P(normal)=0.02976$

$P(canner|+)=0.00784/(0.00784+0.02976)=0.21$

$P(-|cancer)P(cancer)=0.00016$

$P(-|normal)P(normal)=0.96224$

$P(normal|-)= 0.96224/(0.00016 + 0.96224)=0.99834$

贝叶斯推理的结果很大程度上依赖于先验概率，另外不是完全接受或拒绝假设，只是在观察到较多的数据后增大或减小了假设的可能性。

2.2.1 基于最小错误率的贝叶斯决策

- 例2.1 根据已有知识和经验，两类的先验概率为：
  - 正常(ω_1)：$P(\omega_1)=0.9$
  - 异常(ω_2)：$P(\omega_2)=0.1$
  - 对某一样本观察值$\mathbf{x}$，通过计算或查表得到：
    $p(\mathbf{x}|\omega_1)=0.2$，$p(\mathbf{x}|\omega_2)=0.4$

$$P(\omega_1|\mathbf{x})=\frac{P(\omega_1)p(\mathbf{x}|\omega_1)}{\sum_{j=1}^{2}P(\omega_j)p(\mathbf{x}|\omega_j)}=\frac{0.9\times0.2}{0.9\times0.2+0.1\times0.4}=0.818$$

$$P(\omega_2|\mathbf{x})=\frac{P(\omega_2)p(\mathbf{x}|\omega_2)}{\sum_{j=1}^{2}P(\omega_j)p(\mathbf{x}|\omega_j)}=\frac{0.4\times0.1}{0.2\times0.9+0.4\times0.1}=0.182$$

2.2.1 基于最小错误率的贝叶斯决策

- 分类错误率最小的证明（一维）
  - 条件错误率（样本**x**上）为：

$$P(e|\mathbf{x}) = \begin{cases} P(\omega_2|\mathbf{x}) = 1 - P(\omega_1|\mathbf{x}) & \text{若决定 } \mathbf{x} \in \omega_1 \\ P(\omega_1|\mathbf{x}) = 1 - P(\omega_2|\mathbf{x}) & \text{若决定 } \mathbf{x} \in \omega_2 \end{cases}$$

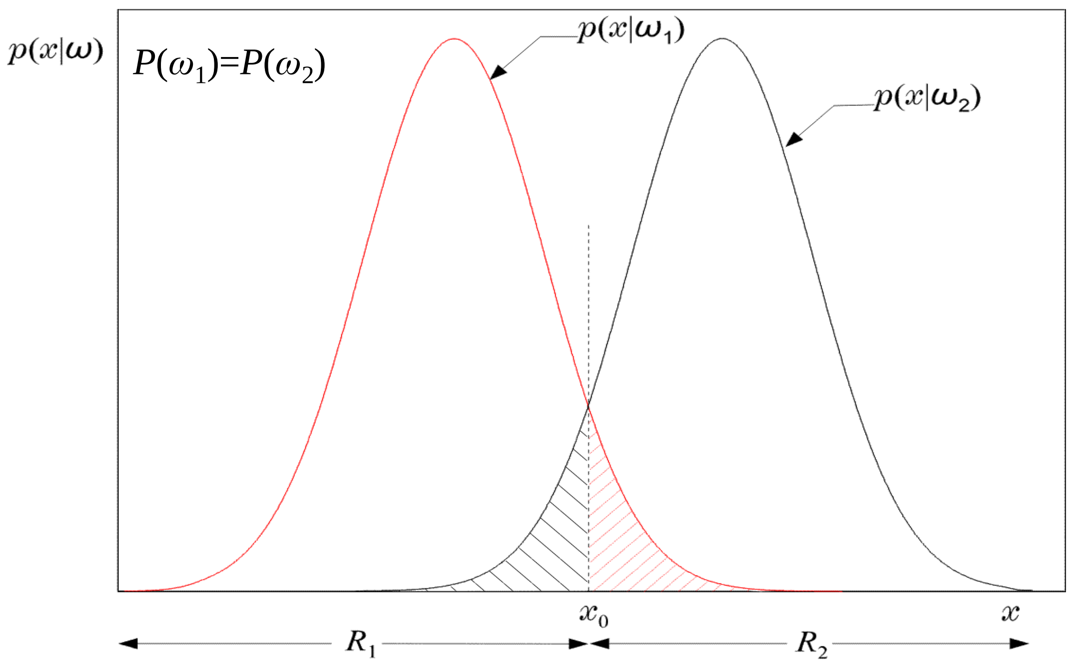
$$= 1 - \max_i P(\omega_i|\mathbf{x})$$

  - （平均）错误率（所有独立样本上条件错误率的期望）：

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

2.2.1 基于最小错误率的贝叶斯决策

- 设t为两类的分界面，则在特征向量x是一维时，$t=x_0$为x轴上的一点。形成两个决策区域：**R1~(-∞，t)和R2~(t，+∞)**



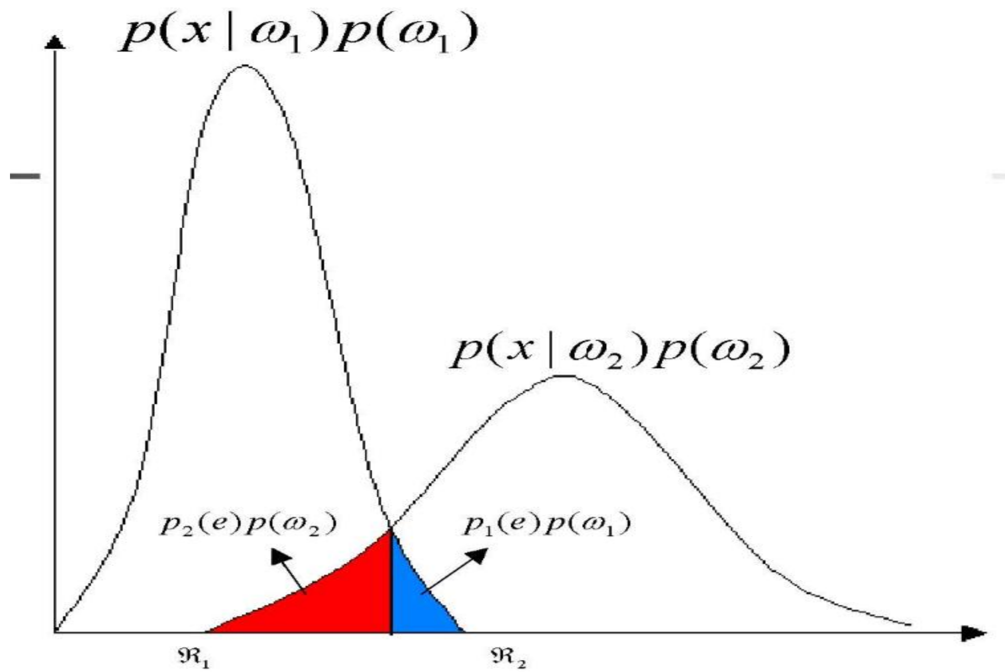
$$R_1(\to\omega_1)\ and\ R_2(\to\omega_2)$$

2.2.1 基于最小错误率的贝叶斯决策

- 设t为两类的分界面，则在特征向量x是一维时，$t=x_0$为x轴上的一点。形成两个决策区域: R1~(-∞，t)和R2~(t，+∞)

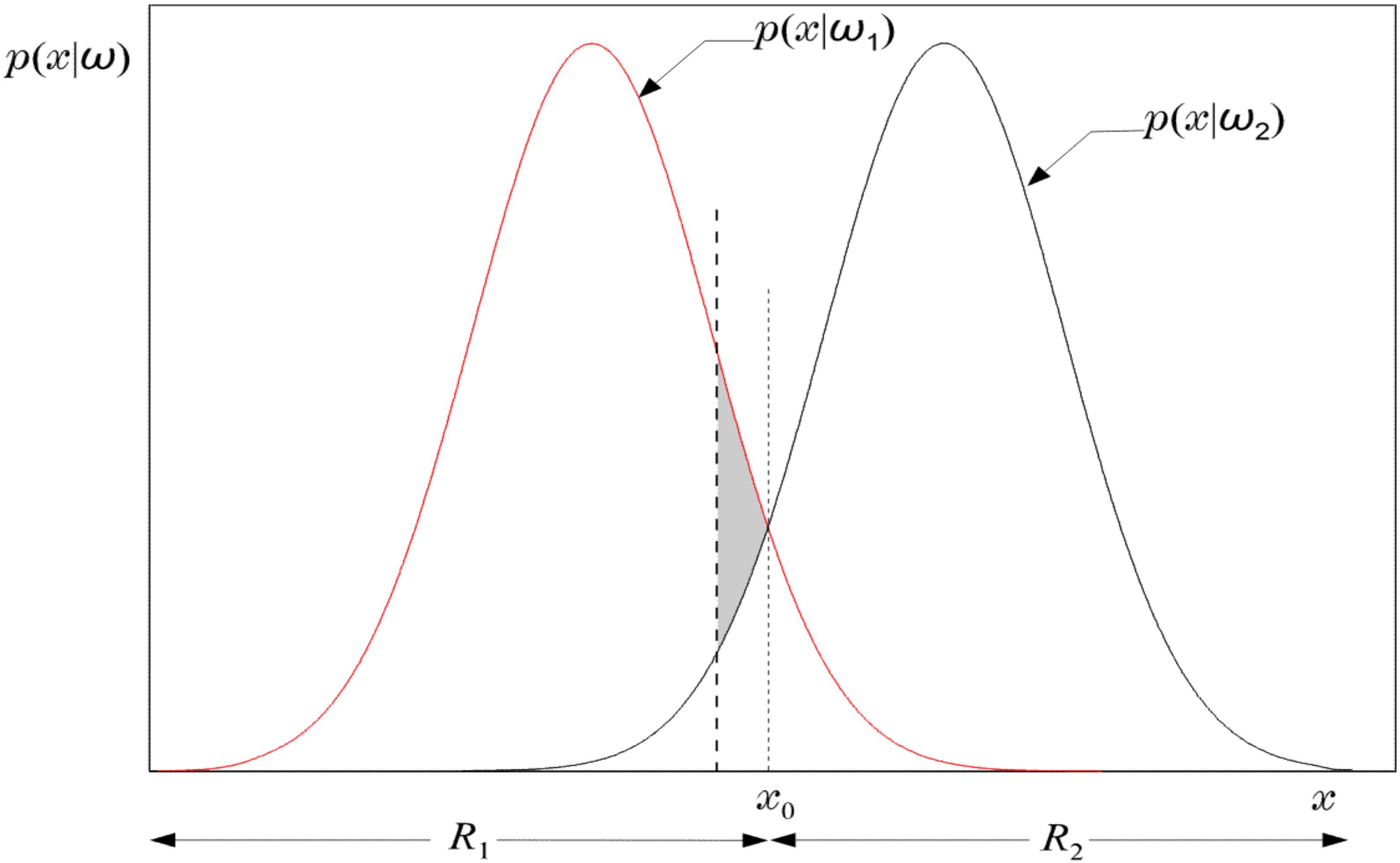
$$
\begin{aligned}
P(e) &= P(x \in R_1, \omega_2) + P(x \in R_2, \omega_1) \\
&= P(\omega_2)P(x \in R_1 \mid \omega_2) + P(\omega_1)P(x \in R_2 \mid \omega_1) \\
&= P(\omega_2)\int_{R_1} p(x \mid \omega_2)dx + P(\omega_1)\int_{R_2} p(x \mid \omega_1)dx \\
&= P(\omega_2)P_2(e) + P(\omega_1)P_1(e)
\end{aligned}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

2.2.1 基于最小错误率的贝叶斯决策



- Bayes最小错误率决策使得每个观测值下的条件错误率最小，因而保证了（平均）错误率最小。
- Bayes决策是一致最优决策

2.2.1 基于最小错误率的贝叶斯决策



2.2.1 基于最小错误率的贝叶斯决策

- 多类决策规则：

（1） If $P(\omega_i \mid x) = \max_{j=1,\cdots,c} P(\omega_j \mid x)$， then $x \in \omega_i$

（2） If $p(x \mid \omega_i)P(\omega_i) = \max_{j=1,\cdots,c} p(x \mid \omega_j)P(\omega_j)$， then $x \in \omega_i$

- 计算正确分类概率更简单

$$P(e) = 1 - P(c) = 1 - \sum_{j=1}^{c} \int_{\Re_j} p(x \mid \omega_j) dx$$

2.2 几种常用的决策规则

2.2.2 基于最小风险的贝叶斯决策

- **风险**：错误判断所造成的损失的严重程度
- **例子：癌细胞的识别**
  - 没病(正常ω_1)被判为有病(异常ω_2)，还可以做进一步检查，损失不大；
  - 有病(ω_2)被判为无病(ω_1)，错过诊治时机，损失严重
- **决策（行动）空间**

$$\mathcal{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_i,\cdots,\alpha_a\}$$

- **损失函数 —决策表或损失矩阵** $(\lambda_{i,j})_{N*N}$

$$\lambda(\alpha_i,\omega_j),\quad i=1,2,\cdots,a;\ j=1,2,\cdots,c$$

2.2.2 基于最小风险的贝叶斯决策

- 基于**最小风险**的贝叶斯决策：决策有代价，选择风险最小的决策

- 条件期望损失（条件风险）—特定x

$$R(\alpha_i \mid x) = E\left[\lambda(\alpha_i,\omega_j)\middle| x\right] = \sum_{j=1}^{c} \lambda(\alpha_i,\omega_j) P(\omega_j \mid x), \quad i = 1, 2, \cdots, a$$

- 期望风险 —整个特征空间

$$R(\alpha) = \int R[\alpha(x) \mid x] p(x) dx$$

2.2.2 基于最小风险的贝叶斯决策

- 贝叶斯最小风险决策，通过保证每个观测值下决策的条件风险最小，使得它的期望风险最小，是一致最优决策
- 最小风险的贝叶斯决策规则

$$\text{if}\ \ R(\alpha_k \mid x) = \min_{i=1,2,\cdots,a} R(\alpha_i \mid x),\ \ \text{then}\ \ \alpha = \alpha_k$$

- 计算步骤
  - （1） 计算后验概率 $P(\omega_j \mid x) = \dfrac{P(\omega_j)p(x \mid \omega_j)}{\sum_{i=1}^{c} P(\omega_i)p(x \mid \omega_i)}$
  - （2） 计算风险

$$R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i, \omega_j) P(\omega_j \mid x),\ \ i = 1, 2, \cdots, a$$

  - （3）决策

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

2.2.2 基于最小风险的贝叶斯决策

- **例2.2 细胞识别——最小风险决策**
  - 正常(ω_1)： $P(\omega_1)=0.9$
  - 异常(ω_2)： $P(\omega_2)=0.1$
  - 对某一样本观察值x，通过计算或查表得到：
    $p(x|\omega_1)=0.2$， $p(x|\omega_2)=0.4$
  - $\lambda_{11}=0$， $\lambda_{12}=6$， $\lambda_{21}=1$， $\lambda_{22}=0$
- 后验概率： $P(\omega_1|x)=0.818$， $P(\omega_2|x)=0.182$

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

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

$$j=\underset{i}{\operatorname{argmin}}\, R(\omega_i \mid \mathbf{x})=2 \qquad \mathbf{x}\in\omega_2$$

2.2.2 基于最小风险的贝叶斯决策

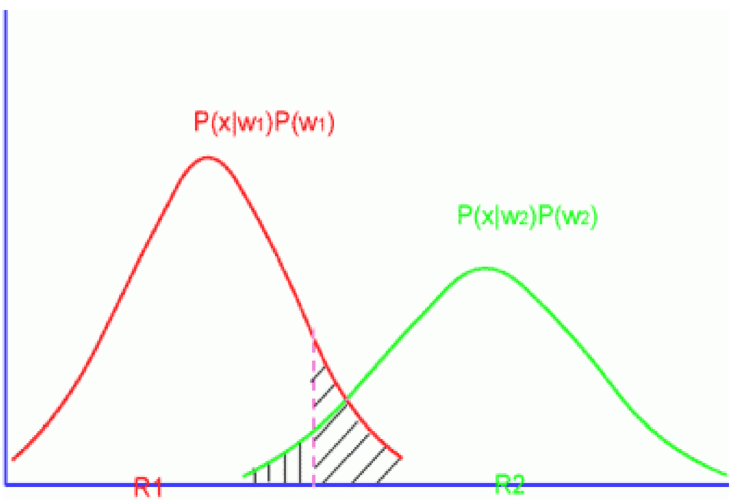
- 最小风险贝叶斯决策除须已知先验概率、类条件概率，还需要有合适的损失函数，要求更多先验知识

- 最小错误率贝叶斯决策就是在**0—1损失函数**条件下的最小风险贝叶斯决策（特例）

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

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

$$=\sum_{j=1,\,j\neq i}^{c}P(\omega_j\mid x)=1-P(\omega_i\mid x)$$



2.2.3 其它决策方法

- **损失函数无法确定，在限定一类错误率条件下，使另一类错误率为最小的两类决策（neyman-pearson决策）**

$$P(\omega_2)P_2(e)+P(\omega_1)P_1(e) \Rightarrow P_2(e)=\varepsilon_0$$

用 Lagrange 乘子法：

$$\min \quad L = P_1(e)+\lambda(P_2(e)-\varepsilon)$$

（对分类边界和 λ 求最小）

$$= \int_{R_2} p(x\mid\omega_1)dx+\lambda\left(\int_{R_1} p(x\mid\omega_2)dx-\varepsilon\right)$$

$$= 1-\int_{R_1} p(x\mid\omega_1)dx+\lambda\left(\int_{R_1} p(x\mid\omega_2)dx-\varepsilon\right)$$

$$= (1-\lambda\varepsilon)+\int_{R_1} \left[\lambda p(x\mid\omega_2)-p(x\mid\omega_1)\right]dx$$

为了求极值点，L对边界t和λ求偏导数，并令其为0.

$$\frac{\partial L}{\partial t}=0 \Rightarrow \lambda^*=\frac{p(t^*|w_1)}{p(t^*|w_2)}$$

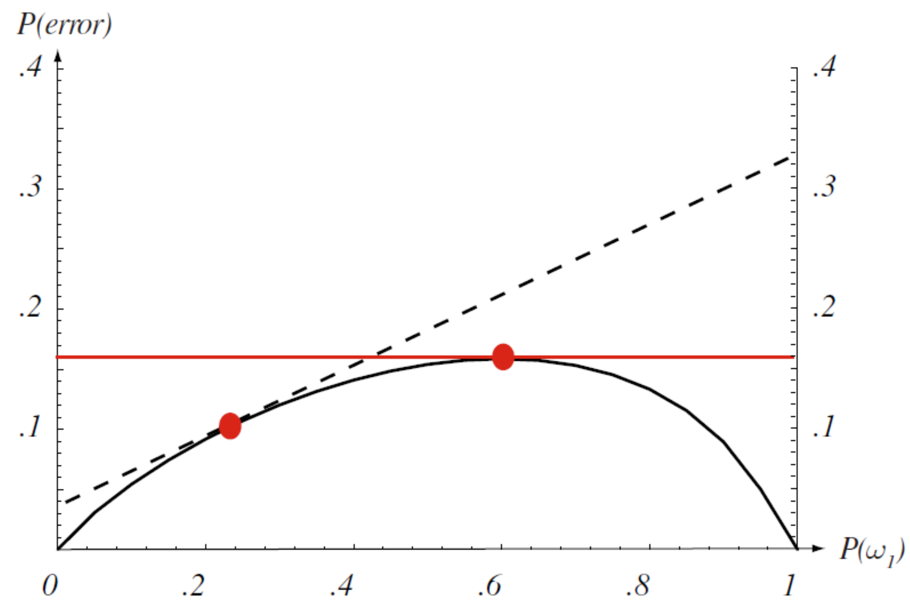
$$\frac{\partial L}{\partial \lambda}=0 \Rightarrow \int_{R_1^*} p(x|w_2)dx$$

- 决策规则 $\quad if \ \dfrac{p(x/\omega_1)}{p(x/\omega_2)} \begin{matrix} > \\ < \end{matrix} \lambda \ \text{then} \ x\in \begin{matrix}\omega_1\\ \omega_2\end{matrix}$

2.2.3 其它决策方法

- **最小最大决策**
  - 先验概率$P(\omega_i)$可变或未知情况下，风险最小

$$R = a + bP(\omega_1)$$



  - 选择使最小贝叶斯风险为最大值时的来设计分类器
  - 最小最大决策是偏于保守的方法，是最坏的贝叶斯风险

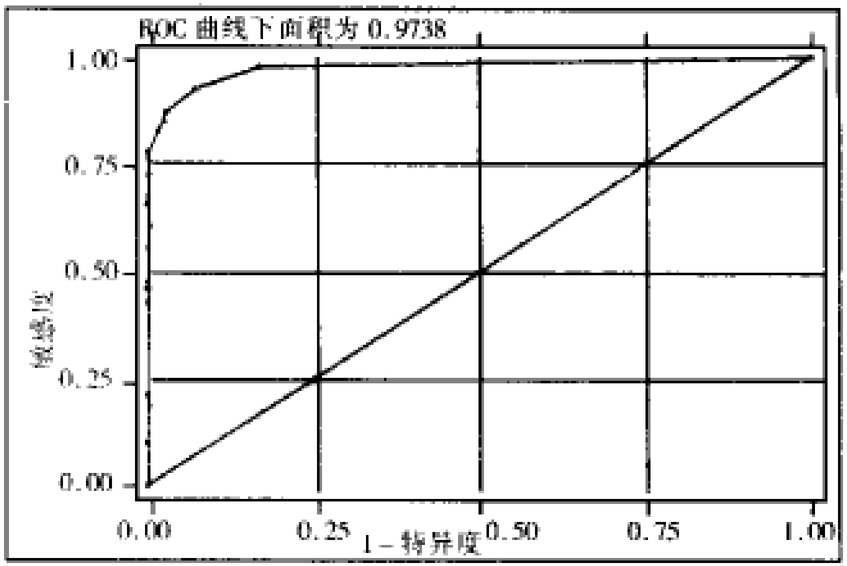
两类错误率与ROC曲线

		预测		
		1	0	合计
实际	1	True Postive **TP**	Frue Negative **FN**	Actual Postive(**TP+FN**)
	0	False Postive **FP**	True Negative **TN**	Actual Negative(**FP+TN**)
合计		Predicted Postive (**TP+FP**)	Predicted Negative (**FN+TN**)	**TP+FN+FP+TN**

- 两类错误率
  - 阳性（Positive）、阴性（Negative）
  - 真阳性（TP）、假阳性（FP、Type-I Error、误报、虚警）
  - 真阴性（TN）、假阴性（FN、Type-II Error、漏报）
  - 灵敏度 $Sn = \dfrac{TP}{TP + FN}$

    真正类率(True Postive Rate)
  - 特异度 $Sp = \dfrac{TN}{TN + FP}$

    负正类率(False Postive Rate)FPR

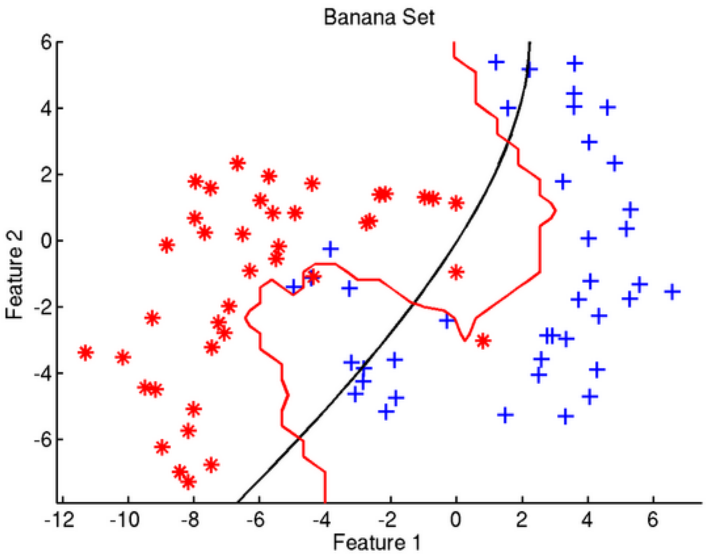


- ROC曲线 接收者操作特征(receiveroperating characteristic)
- AUC (Area under Curve)：Roc曲线下的面积，介于0.1和1之间。Auc作为数值可以直观的评价分类器的好坏，值越大越好。

2.2.4 分类器设计

- 特征空间 $\xrightarrow{\text{决策规则}}$ 决策区域
- 决策面：决策域的边界面
- 决策面方程：解析形式
- 判决函数：表达决策规则的一组函数

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

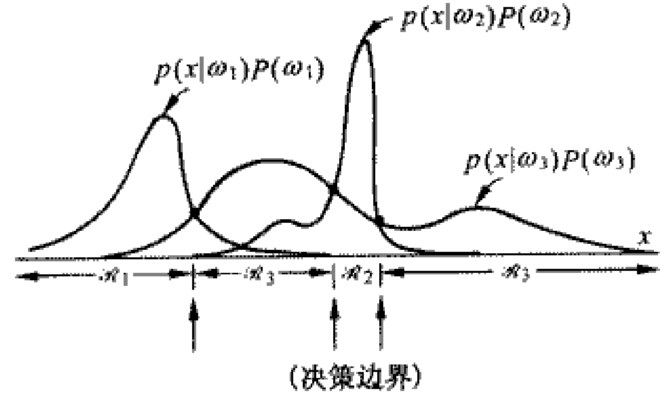


2.2.4 分类器设计

- 多类情况
  - 决策规则

$$\text{if} \quad g_j(x) = \max_i \{g_i(x)\} \quad \text{then} \quad x \in \omega_j$$

$$j = \operatorname*{argmax}_i g_i(x)$$

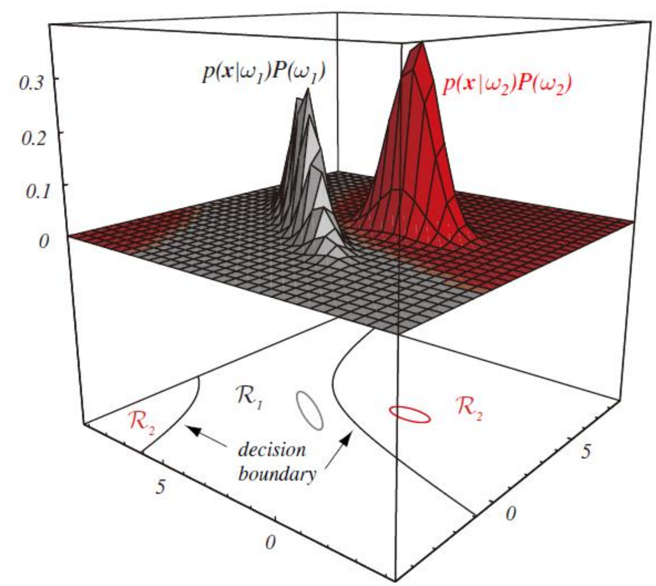


  - 判别函数

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

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

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

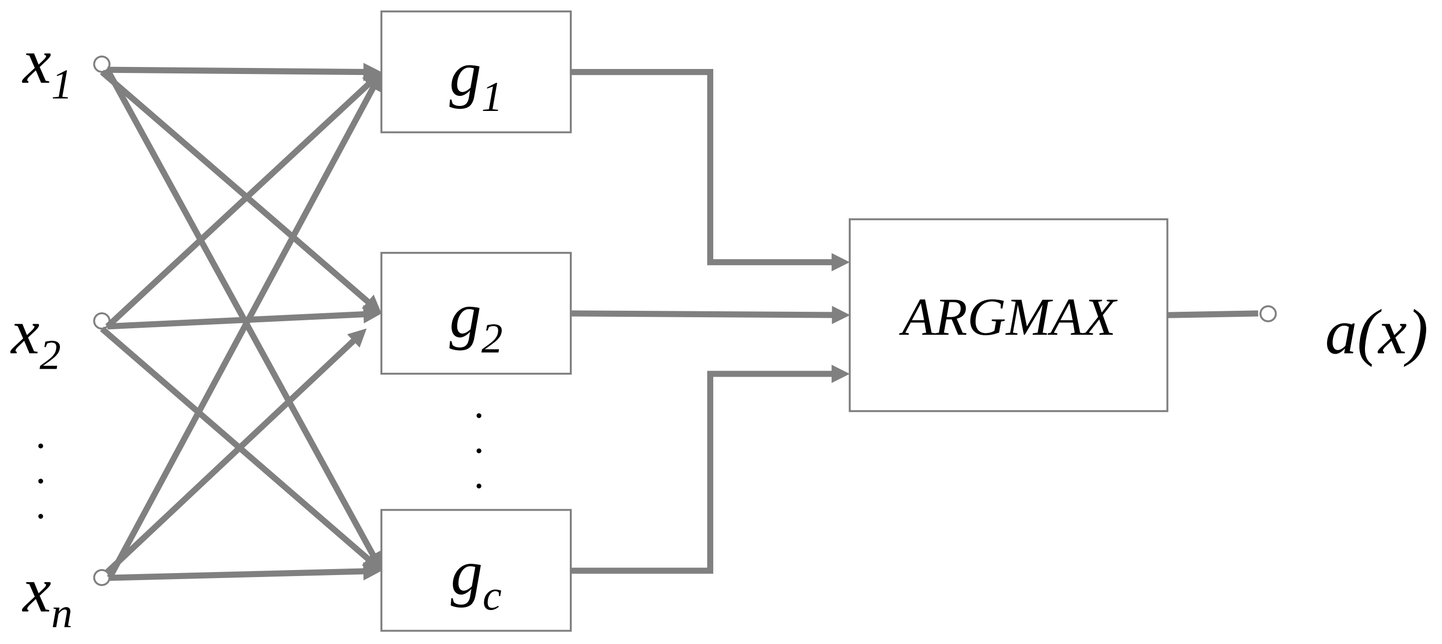


  - 决策面方程

$$g_i(x) = g_j(x)$$

2.2.4 分类器设计

- **分类器是某种由硬件或软件组成的“机器”：**
  - 计算c个判别函数$g_i(\mathbf{x})$
  - 最大值选择



2.2.4 分类器设计

- x为一维时，决策面为一点
- x为二维时决策面为曲线
- x为三维时，决策面为曲面
- x大于三维时决策面为超曲面

两类情况

- 决策规则

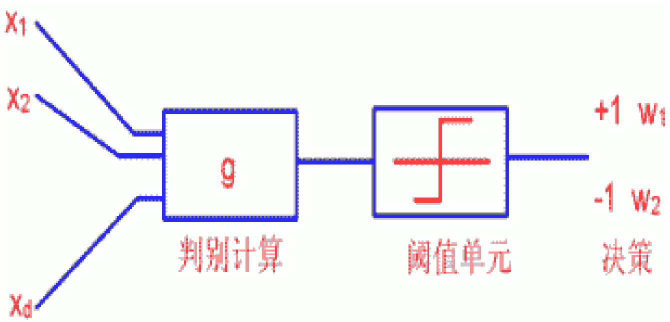
$$if \quad g(x)=g_1(x)-g_2(x) \begin{matrix} > \\ < \end{matrix} 0 \quad \text{then} \quad x \in \begin{matrix} \omega_1 \\ \omega_2 \end{matrix}$$

- 判别函数（仅定义一个）

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

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

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



- 决策面方程

$$g(x)=0$$

2.2.4 分类器设计

- **例2.1、2.2的判别函数和决策面**
  - 2.1

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

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

  - 2.2

判别函数 $$g(x) = \lambda_{21}p(x|\omega_1)P(\omega_1) - \lambda_{12}p(x|\omega_2)P(\omega_2)$$
$$= 0.9p(x|\omega_1) - 0.6p(x|\omega_2) \quad [\lambda_{21} = 1, \lambda_{12} = 6]$$

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

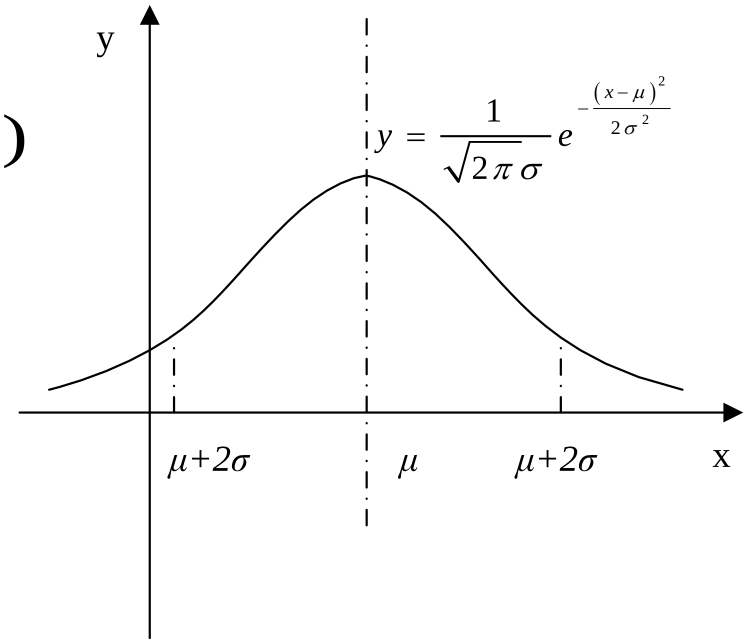
2.3 正态分布时的统计决策

- **Bayes决策的三个前提：**
  - 类别数确定
  - 各类的先验概率$P(\omega_i)$已知
  - 各类的条件概率密度函数$p(x|\omega_i)$已知
- **Bayes决策中，类条件概率密度的选择要求：**
  - 模型合理性
  - 计算可行性
- **最常用概率密度模型：正态分布**
  - 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，它们（近似）服从正态分布
  - 计算、分析最为简单的模型

2.3.1 关于正态分布的知识

- 单变量正态分布（一元正态分布）
  - 概率密度函数

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



  - 两个重要参数
    - 均值（中心）
    - 方差（分散度）

$$\mu=E\{x\}=\int_{-\infty}^{\infty}xp(x)dx$$

$$\sigma^2=E\left\{(x-\mu)^2\right\}=\int_{-\infty}^{\infty}(x-\mu)^2p(x)dx$$

2.3.1 关于正态分布的知识

- 多元正态分布
  - 概率密度函数

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

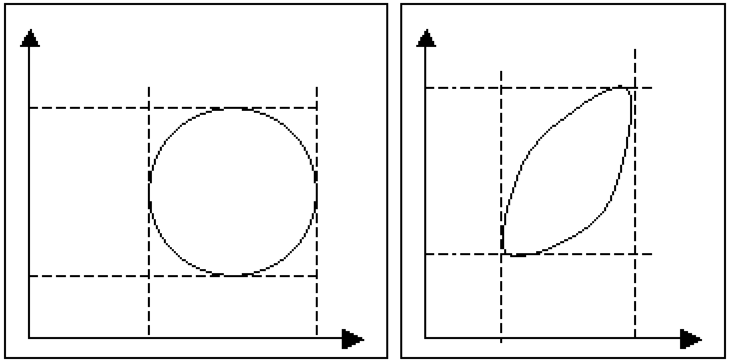
$$\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$$

均值向量 $$\boldsymbol{\mu} = E(\mathbf{x}) = (\mu_1, \mu_2, \ldots, \mu_d)^T,\ \mu_i = E(x_i)$$

协方差矩阵 $$\Sigma = E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\} = (\sigma_{ij}^2)_{d*d}$$

$$\sigma_{ij}^2 = E\left[(x_i-\mu_i)(x_j-\mu_j)\right]$$

协方差矩阵是对称矩阵



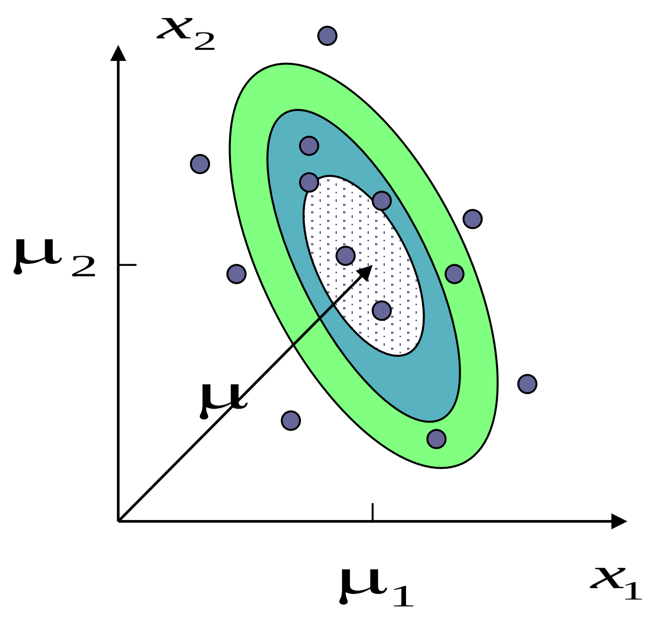
2.3.1 关于正态分布的知识

- **多元正态分布的性质**
  - 参数μ和Σ完全决定分布

    $$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

  - 等概率密度轨迹为超椭球面
    - Mahalanobis距离（马氏距离）

$$p(\mathbf{x}) = c \Rightarrow (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \gamma^2$$



  - 不相关性等价于独立性
  - 边缘分布和条件分布的正态性
  - 线性变换的正态性 $\mathbf{y} = A\mathbf{x}$
  - 线性组合的正态性 $y = \alpha^T \mathbf{x}$

    （一维）

2.3.2 正态分布下的贝叶斯决策

- **最小错误率贝叶斯判别函数和决策面**
  - 观测向量的类条件分布服从正态分布：

$$p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \Sigma_i), \quad i = 1, 2, \ldots, c$$

  - 判别函数

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

  - 决策面方程

$$g_i(x) = g_j(x)$$

2.3.2 正态分布下的贝叶斯决策

- **1.第一种情况：等协方差阵，独立，且同方差**

$$\Sigma_i = \sigma^2 I = \begin{bmatrix} \sigma_{11}^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_{nn}^2 \end{bmatrix}, \quad i = 1, 2, \dots, c$$

  - 判别函数简化为

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

$$(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) = \|\mathbf{x}-\boldsymbol{\mu}_i\|^2 = \sum_{j=1}^{d}(x_j-\mu_{ij})^2$$

欧氏距离

2.3.2 正态分布下的贝叶斯决策

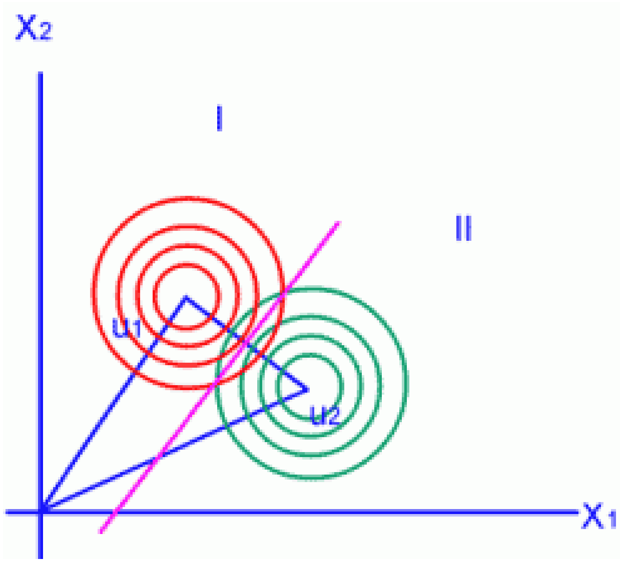
- **1.第一种情况：等协方差阵，独立，且同方差**
  - 特例：先验概率相等

$$\Sigma_i = \sigma^2 I, \quad P(\omega_i) = P(\omega_j) \quad i, j = 1, 2, \ldots, c$$

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu}_i)^T(\mathbf{x} - \boldsymbol{\mu}_i) = -\frac{1}{2\sigma^2}\|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

$$j = \arg\min_{i=1,\cdots,c} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

最小距离分类器
（模板匹配）



2.3.2 正态分布下的贝叶斯决策

- **1.第一种情况：等协方差阵，独立，且同方差**
  - 考虑 二次项$\mathbf{x}^T\mathbf{x}$与i无关

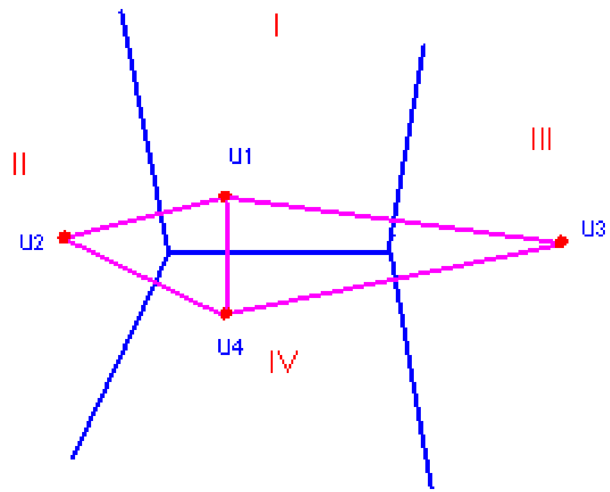
$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(-2\boldsymbol{\mu}_i^T\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i) + \ln P(\omega_i) = \mathbf{w}_i^T\mathbf{x} + w_{i0},$$

其中：$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i, w_{i0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i + \ln P(\omega_i)$

线性分类器

$$j = \arg\max_{i=1,\cdots,c} g_i(\mathbf{x})$$

2.3.2 正态分布下的贝叶斯决策



- 决策面方程（线性）

$$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$$

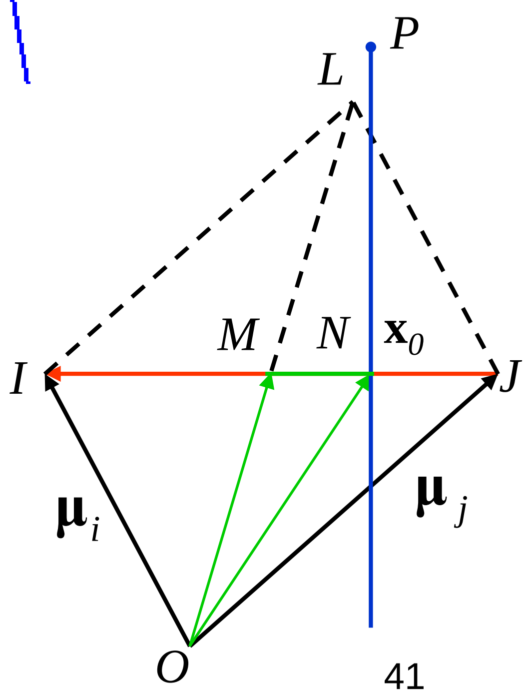
- 改写为

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

其中 $\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|} \ln\frac{P(\omega_i)}{P(\omega_j)}$$

- 决策面为一超平面，



2.3.2 正态分布下的贝叶斯决策

- **2.第二种情况：等协方差阵**

$$\Sigma_i = \Sigma, \quad i = 1, 2, \ldots, c$$

  - 判别函数简化为

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

$P(\omega_i) = P(\omega_j)$ 时

$$g_i(\mathbf{x}) = \gamma^2 = (\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$$

马氏距离

2.3.2 正态分布下的贝叶斯决策

- **2.第二种情况：等协方差阵**
  - 线性判别函数

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

$$\mathbf{w}_i = \Sigma^{-1} \boldsymbol{\mu}_i,$$

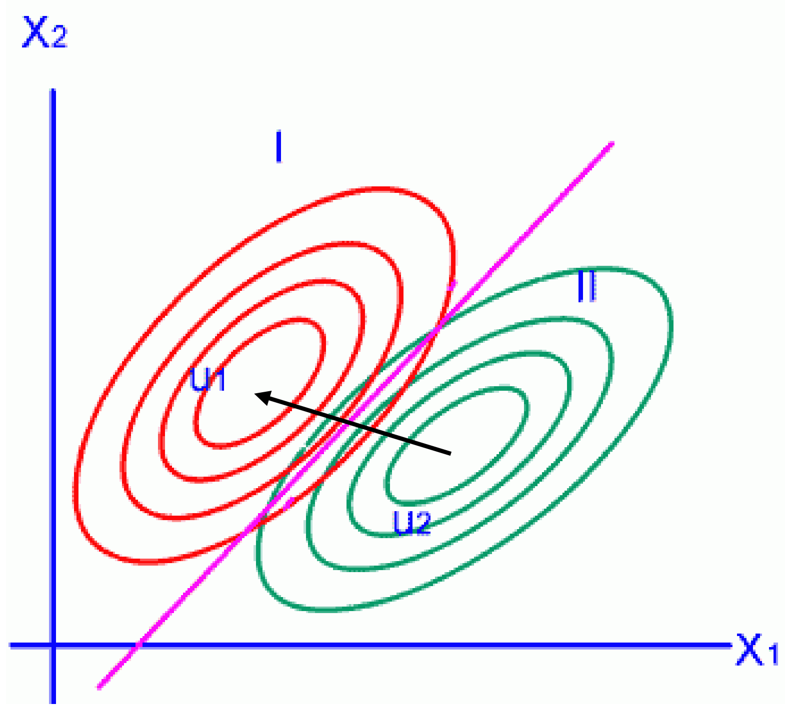
$$w_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i + \ln P(\omega_i)$$

  - 决策面方程

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

$$\mathbf{w} = \Sigma^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$\mathbf{x}_0 = \frac{1}{2} (\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$



2.3.2 正态分布下的贝叶斯决策

- **3.第三种情况：各类协方差阵不相等**

$$\Sigma_i \neq \Sigma_j, \quad i, j = 1, 2, \ldots, c$$

  - 判别函数为

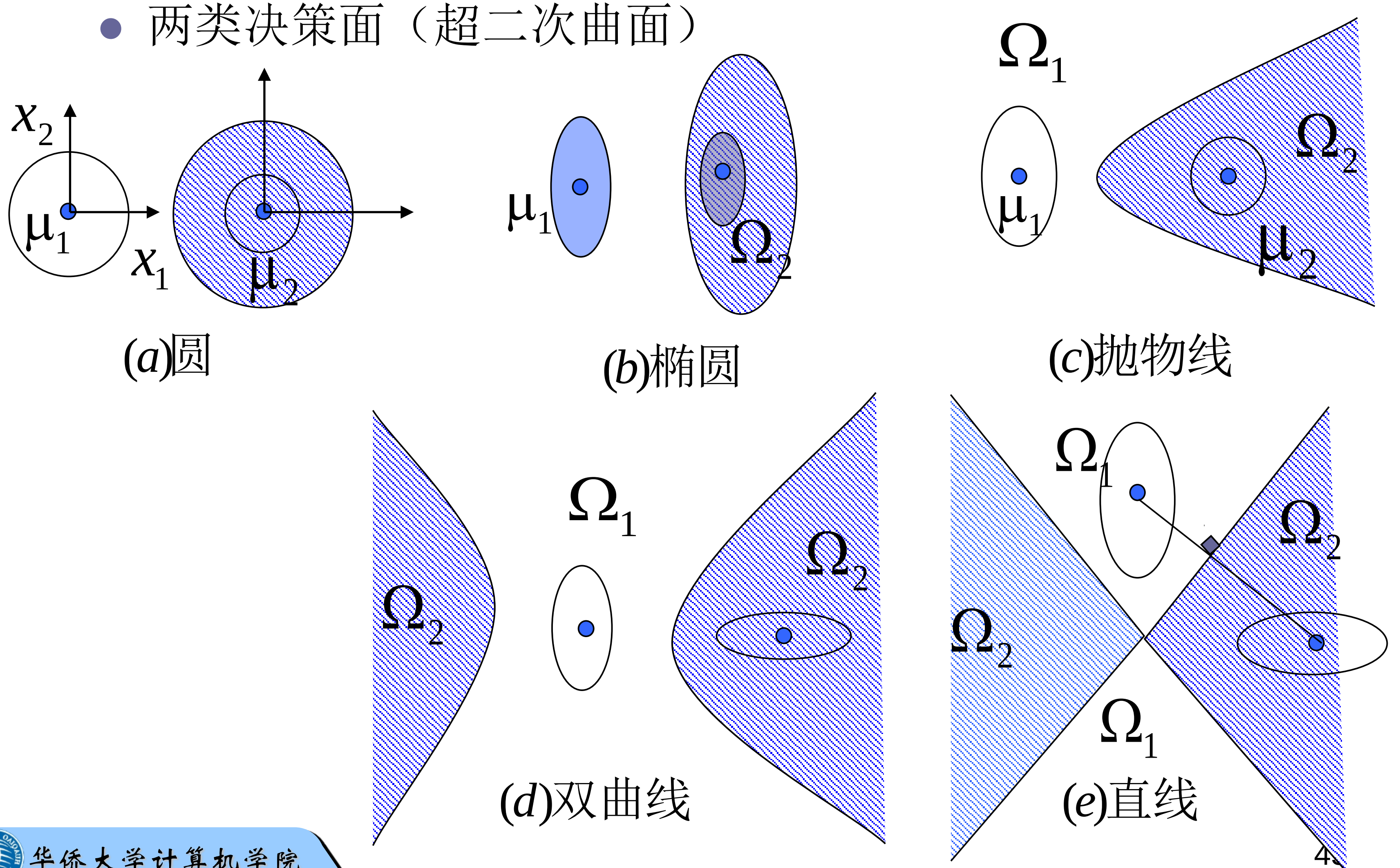
$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$
$$= \mathbf{x}^T W_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

  - 为二次型，决策面为超二次曲面

$$\mathbf{x}^T (W_i - W_j)\mathbf{x} + (\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{x} + w_{i0} - w_{j0} = 0$$

2.3.2 正态分布下的贝叶斯决策

- 两类决策面（超二次曲面）



(a)圆　(b)椭圆　(c)抛物线　(d)双曲线　(e)直线

2.3.2 正态分布下的贝叶斯决策

- **例2.3　两类的识别问题：医生要根据病人血液中白细胞的浓度来判断病人是否患血液病**
  - 根据医学知识和以往的经验，医生知道：
    - 患病的人，白细胞的浓度服从均值2000，标准差1000的正态分布；未患病的人，白细胞的浓度服从均值7000，标准差3000的正态分布；
    - 一般人群中，患病的人数比例为0.5%。
    - 一个人的白细胞浓度是3100
  - 医生应该做出怎样的判断？

2.3.2 正态分布下的贝叶斯决策

- 用 ω 表示“类别”这一随机变量，ω_1表示患病，ω_2表示正常；x表示“白细胞浓度”这个随机变量

1) 类别的先验分布：

$P(\omega_1) = 0.5\%$

$P(\omega_2) = 99.5\%$

2)观测数据白细胞浓度分别在两种情况下的类条件分布：

$P(x|\omega_1) \sim N(2000,1000^2)$

$P(x|\omega_2) \sim N(7000,3000^2)$

$$\boldsymbol{P(3100|\omega_1) = 2.1785\text{e-}004}$$
$$\boldsymbol{P(3100|\omega_2) = 5.7123\text{e-}005}$$

3)计算后验概率

$\boldsymbol{P(\omega_1|3100)=1.9\%}$ $\boldsymbol{P(\omega_2|3100)=98.1\%}$

- 医生的判断：**正常**

2.4 关于分类器的错误率问题

- 研究错误的意义
  - 对于同一规则，错误率反映问题的固有复杂程度（bayes问题）
  - 对于同一问题，错误率反映不同方法的优劣
- 错误率计算

$$P(e) = P(\omega_1)\int_{R_2} p(x|\omega_1)dx + P(\omega_2)\int_{R_1} p(x|\omega_2)dx$$

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

  - 1) 按理论公式计算
  - 2) 计算错误率上界
  - 3) 实验估计：C法、交叉验证（划分、留一）

2.5 讨论

- **基于Bayes决策的最优分类器**
  - 最优的条件
  - 评价指标（错误率、风险）

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i) P(\omega_i)}{\sum_j p(\mathbf{x} \mid \omega_j) P(\omega_j)}$$

Bayes决策的三个前提：

1）类别数确定

2）各类的先验概率$P(\omega_i)$已知

3）各类的条件概率密度函数$p(\mathbf{x}|\omega_i)$已知

→ **基于样本估计概率密度**

→ **基于样本直接确定判别函数**

（直接设计分类器）

正态分布的Bayes决策例解

- 两类的识别问题：医生要根据病人血液中白细胞的浓度来判断病人是否患血液病。
- 根据医学知识和以往的经验，医生知道：
  - 患病的人，白细胞的浓度服从均值2000，标准差1000的正态分布；未患病的人，白细胞的浓度服从均值7000，标准差3000的正态分布；
  - 一般人群中，患病的人数比例为0.5%。
  - 一个人的白细胞浓度是3100，医生应该做出怎样的判断？

正态分布的Bayes决策例解

- **数学表示：**
  - 用ω表示"类别"这一随机变量，ω_1表示患病，ω_2表示正常；
  - x表示"白细胞浓度"这个随机变量。
- **本例医生掌握的知识非常充分，包括：**

1) 类别的先验分布：$P(\omega_1) = 0.5\%$，$P(\omega_2) = 99.5\%$

先验分布：没有获得观测数据（病人白细胞浓度）之前，已知的关于类别的分布（某类事物出现的比例）。

正态分布的Bayes决策例解

2) 观测数据白细胞浓度分别在两种情况下的类条件分布：

$$P(x|\omega_1) \sim N(2000, 1000^2)$$

$$P(x|\omega_2) \sim N(7000, 3000^2)$$

- 样本观测值： $x = 3100$

  $P(3100|\omega_1) = 2.1785\text{e-}004$, $P(3100|\omega_2) = 5.7123\text{e-}005$

- 计算后验概率：

  $P(\omega_1|3100)=1.9\%$, $P(\omega_2|3100)=98.1\%$

- 医生的判断：正常

习题

1. 对一个 c 类分类问题，假设各类先验概率为 $P(\omega_i), i=1,\dots,c$，条件概率密度为 $P(\mathbf{x}\mid\omega_i), i=1,\dots,c$（这里 x 表示特征向量），将第 j 类模式判别为第 i 类的损失为 λ_{ij}。

（1）请写出贝叶斯最小风险决策和最小错误率决策的决策规则；

（2）引入拒识（表示为第 c+1 类），假设决策损失为

$$\lambda_{ij}=\begin{cases}0, & i=j\\ \lambda_r, & i=c+1\\ \lambda_s, & \text{otherwise}\end{cases}$$

请写出最小损失决策的决策规则（包括分类规则和拒识规则）。

2. 表示模式的特征向量 $\mathbf{x} \in R^d$，对一个 c 类分类问题，假设各类先验概率相等，每一类条件概率密度为高斯分布。

（1）请写出类条件概率密度函数的数学形式；

（2）请写出在下面两种情况下的最小错误率决策判别函数：(a)类协方差矩阵不等；(b)所有类协方差矩阵相等。

3 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ 是来源于已知协方差矩阵、未知均值的正态分布的向量，即

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

求出未知均值向量的最大似然估计。

假设在 上例 中的未知均值向量 $\boldsymbol{\mu}$ 现在是已知的，且服从的正态分布是

$$p(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{l/2}\sigma_{\mu}^{l}} \exp\left(-\frac{1}{2}\frac{\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2}{\sigma_{\mu}^2}\right)$$

通过下式得到最大后验概率估计：

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln\left(\prod_{k=1}^{N} p(\boldsymbol{x}_k|\boldsymbol{\mu}) p(\boldsymbol{\mu})\right) = 0$$

每位同学严格命名

学号_姓名

作业缴交地址：
ftp://210.34.246.224/