第2章 统计/Bayes决策方法

Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

引言

- ✓机器自动识别分类，能不能避免错分类，如汉字识别能不能做到百分之百正确？怎样才能减少错误？
- ✓错分类往往难以避免，因此就要考虑减小因错分类造成的危害损失，譬如对病理切片进行分析，有可能将正确切片误判为癌症切片，反过来也可能将癌症病人误判为正常人，这两种错误造成的损失一样吗？看来后一种错误更可怕，那么有没有可能对后一种错误严格控制？
- ✓ 概率论中的先验概率，后验概率与概率密度函数？什么是贝叶斯公式？
- ✓什么叫正态分布？什么叫期望值？什么叫方差？为什么说正态分布是最重要的分布之一？

引言

- 统计模式识别：

    用概率统计的观点和方法来解决模式识别问题

引言

- 基本概念：
  - ✓样本(sample) $\boldsymbol{x} \in \boldsymbol{R}^d$
  - ✓状态(state) 第一类：$x \in \omega_1$ ，第二类：$x \in \omega_2$
  - ✓先验概率 (a priori probablity or prior)
    $P(\omega_1),\quad P(\omega_2)$
  - ✓样本分布密度(sample distribution density) $p(x)$
    （总体概率密度）
  - ✓类条件概率密度(class-conditional probablity density) $p(x|\omega_1),\quad p(x|\omega_2)$

类条件概率密度函数 $p(x|\omega)$

- (Class-conditional probability density)函数：类别状态为ω时的x的概率密度函数(It is also sometimes called state-conditional probability density)

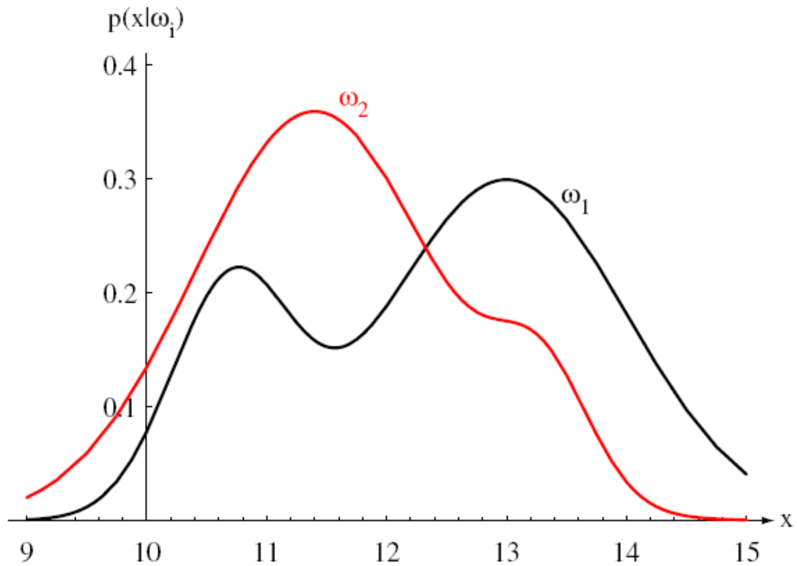


Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the length of a fish, the two curves might describe the difference in length of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0.

引言

贝叶斯决策理论方法所讨论的问题是：已知总共有c类物体，各类别ω_i，i=1,2,…,c的先验概率$P(\omega_i)$及类条件概率密度函数$p(x|\omega_i)$已知的条件下，如何对某一样本按其特征向量分类的问题。由于属于不同类的待识别对象存在着呈现相同观察值的可能，即所观察到的某一样本的特征向量为X，而在c个类别中又有不止一类可能呈现这一X值，这种可能性可用$P(\omega_i|X)$表示。如何作出合理的判决就是贝叶斯决策理论所要讨论的问题。

引言

- 贝叶斯决策（统计决策理论）
  - ✓是统计模式识别的基本方法和基础
  - ✓是“最优分类器”：使平均错误率最小
- 条件:
  - ✓类别数一定（已知）$\omega_i,\ i=1,\cdots,c$
  - ✓已知类先验概率和类条件概率密度

$$P(\omega_i),\ p(x|\omega_i),\quad i=1,\cdots,c$$

Outline:

- 引言
- Bayes决策
  - 基于最小错误率的Bayes决策
  - 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - 正态分布及性质
  - 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

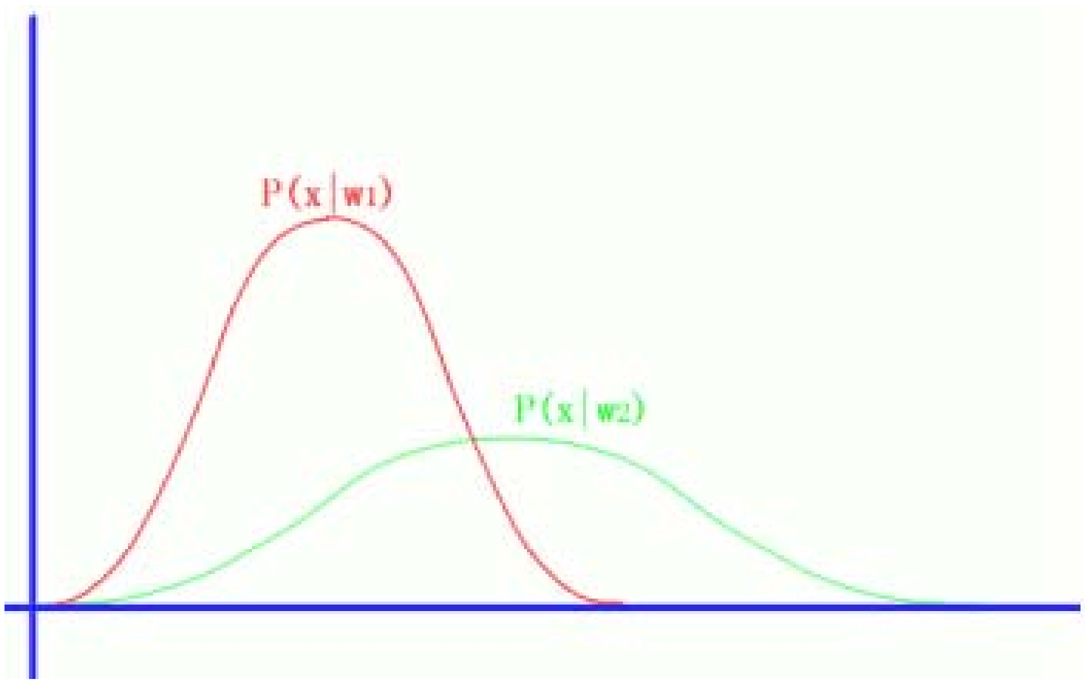
基于最小错误率的Bayes决策规则

- $p(*|\#)$ 的意义：

- $p(x|\omega_i)$ 及 $P(\omega_i|x)$

- 一个事物在某条件下出现的概率P(* | #)与该事件在不带任何条件下出现的概率(写成P(*))是不相同的。

- 假设每个要识别的细胞已作过预处理，并抽取出了d个特征描述量，用一个d维的特征向量X表示，识别的目的是要依据该X向量将细胞划分为正常细胞或者异常细胞。这里我们用ω_1表示是正常细胞，而ω_2则属于异常细胞。



➢如何计算后验概率?

✓已知 $P(\omega_i)$, $p(x|\omega_i)$, $i=1,2$

✓贝叶斯公式：（Bayesian Theory）

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

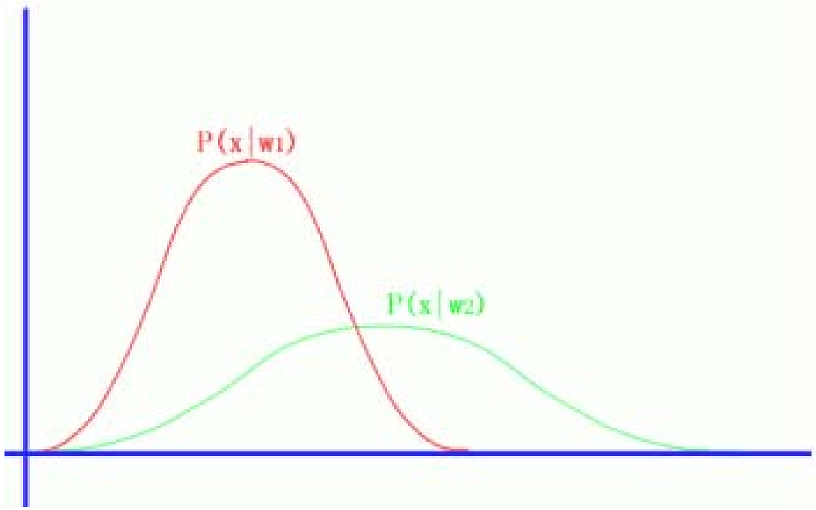
Bayes公式表明通过观测x的值，可将先验概率$P(\omega_i)$转换为后验概率$P(\omega_i|x)$，即假设特征值x已知的条件下类别属于ω_i的概率。

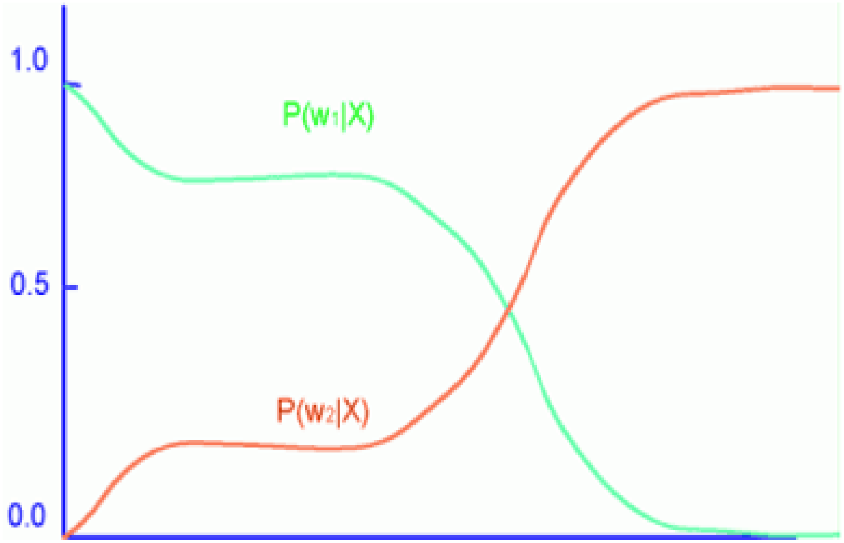
Bayes公式是根据联合概率这一概念推出的，同时出现两个事件X及ω_i的概率为$p(x,\omega_i)$。它是某个条件出现的概率(如 $P(\omega_i)$)以及在此条件下某事件出现概率($p(x|\omega_i)$)的乘积，在此写为：

$$p(x,\omega_i)=p(x|\omega_i)P(\omega_i)=P(\omega_i|x)p(x)$$

先验概率是针对ω_i，i = 1,2,···,c,这c个事件出现的可能性而言的，不考虑其它任何条件。例如世界上有60亿人口，而中国人口12亿，因此不管其它条件，应有20%的可能是中国人。

类条件概率密度函数$p(x|\omega_i)$是指ω_i条件下在一个连续的函数空间出现 x 的概率密度 $p(x|\omega_i)$在这里指第 ω_i 类样本，它的属性 x 是如何分布的。


P(x|w1)
P(x|w2)


1.0
0.5
0.0
P(w1|X)
P(w2|X)

✓错误概率(probablity of error)

$$P(e|\mathbf{x})=\begin{cases}P(\omega_2|\mathbf{x}) & \text{if } \mathbf{x} \text{ is assigned to } \omega_1\\ P(\omega_1|\mathbf{x}) & \text{if } \mathbf{x} \text{ is assigned to } \omega_2\end{cases}$$

✓错误率(average probablity of error) P(e)：

定义为所有服从同样分布的独立样本上错误概率的期望:

$$P(e)=\int P(e|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

✓正确率(proabality of correctness)

$$P(c)=1-P(e)$$

最小错误率贝叶斯决策规则的几种等价表达形式：

（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 $x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

- 例：课本(张)P17----例2.1

- 条件概率:

$$P(\omega_1|x) \text{ 和 } P(\omega_2|x)$$

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

$$p(x|\omega_1) \text{ 和 } p(x|\omega_2)$$

两者没有联系！

- 后验概率为什么要利用Bayes公式从先验概率和类条件概率密度函数导出?

➢贝叶斯决策之错误率（一维为例）

$$\min \quad P(e)=\int P(e|x)p(x)dx$$

$$\because \quad P(e|x)\geq 0, \; p(x)\geq 0$$

$\therefore$ 等价于：$\min P(e|x)$ for all x.

而 $$P(e|x)=\begin{cases} P(\omega_2|x) & \text{if assign } x\in\omega_1 \\ P(\omega_1|x) & \text{if assign } x\in\omega_2 \end{cases}$$

➢贝叶斯决策之错误率（一维为例）

✓如果把作出 ω_1决策的所有观测值区域称为 $\mathbf{R}_1$，则在 $\mathbf{R}_1$区内的每个x值，条件错误概率为$P(\omega_2|x)$。另一个区 $\mathbf{R}_2$中的x,条件错误概率为 $P(\omega_1|x)$。因此平均错误率P(e)可表示成

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

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

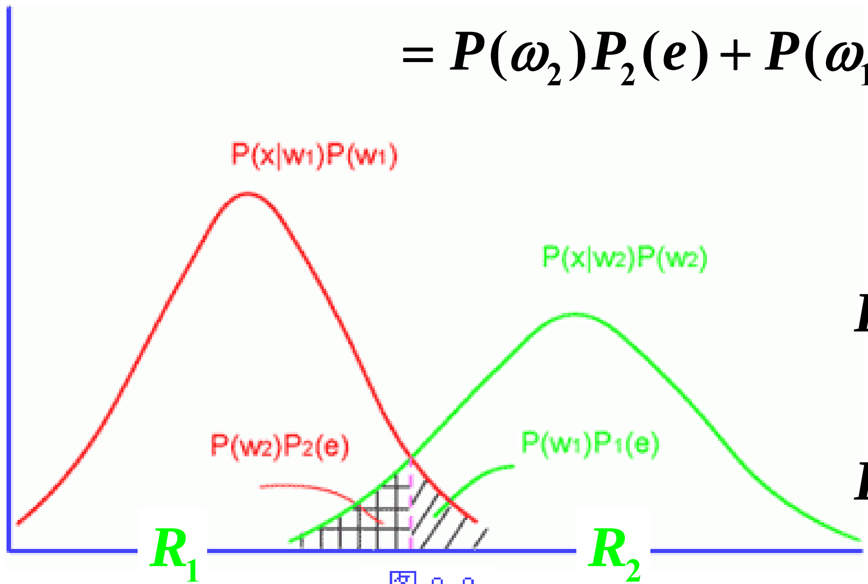


图 2.3

where：

$$P_1(e)=\int_{R_2}P(x|\omega_1)dx$$

$$P_2(e)=\int_{R_1}P(x|\omega_2)dx$$

多类情况:

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

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

$$\text{then}\ x\in\omega_i$$

$$P(c)=\sum_{j=1}^{c}P(x\in\Re_j|\omega_j)P(\omega_j)=\sum_{j=1}^{c}\int_{\Re_j}p(x|\omega_j)P(\omega_j)dx$$

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

Outline:

- 引言
- Bayes决策
  - 基于最小错误率的Bayes决策
  - 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - 正态分布及性质
  - 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

最小风险Bayes决策

- 根据不同性质的错误会引起不同程度的损失这一考虑出发，有些情况下，宁肯扩大一些总的错误率，但也要使总的损失减少。这会引进一个与损失有关联的，更为广泛的概念——风险。在作出决策时，要考虑所承担的风险。基于最小风险的贝叶斯决策规则正是为了体现这一点而产生的。

如果将作出判决的依据从单纯考虑后验概率最大值，改为对该观测值X条件下各状态后验概率求加权和的方式，表示成

$$R_i(X)=\sum_{j=1}^{c}\lambda_j^{(i)}P(\omega_j \mid X)$$

其中$\lambda_j^{(i)}$表示观测样本X实属类别ω_j,而被判为状态ω_i时所造成的损失，R_i则表示了观测值X被判为ω_i类时损失的均值。如果我们希望尽可能避免将某状态ω_j错判为状态ω_i,则可将相应的$\lambda_j^{(i)}$值选择得大些，以表明损失的严重性。加权和R_i用来衡量观测样本X被判为状态ω_i所需承担的风险。

病理切片X: ω_1表示正常，ω_2表示异常，$P(\omega_1|X)$与$P(\omega_2|X)$分别表示两种可能性的大小

$$\lambda_1^{(1)} \qquad \lambda_2^{(1)}$$

$$\lambda_1^{(2)} \qquad \lambda_2^{(2)}$$

$$R_1(X)=\lambda_1^{(1)}P(\omega_1|X)+\lambda_2^{(1)}P(\omega_2|X)$$

$$R_2(X)=\lambda_1^{(2)}P(\omega_1|X)+\lambda_2^{(2)}P(\omega_2|X)$$

此时作出哪一种决策就要看是$R_1(X)$小还是$R_2(X)$小了，这就是基于最小风险的贝叶斯决策的基本出发点。

- 最小错误率只考虑了错误
- 进一步可考虑不同错误所带来的损失（代价）
- 用决策论方法把问题表述如下：
  - ✓把样本 x 看作 d 维随机向量 $x=[x_1,x_2,\cdots,x_d]^T$
  - ✓状态空间Ω： $\Omega=\{\omega_1,\omega_2,\cdots,\omega_c\}$
  - ✓决策空间由 k 个决策组成：

$$A=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$$

✓ 对于实际状态为 ω_j 的向量 x，采取决策 α_i 带来的损失为

$$\lambda\left(\alpha_i, \omega_j\right), \quad i = 1, \cdots, k, \quad j = 1, \cdots, c$$

λ_j^i 表示当真实状态为 ω_j 而所采取的决策为 α_i 时所带来的损失，这可通过决策表来给出。

- 条件期望损失：对于特定的 x 采取决策 α_i 的期望损失：

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

$$i = 1, 2, \cdots, k$$

- 期望风险：对所有可能的 x采取决策 $\alpha(x)$（一般的判别规则是一个函数）所造成的期望损失之和

$$R(\alpha) = \int R(\alpha(x) | x) p(x) dx$$

在考虑错判带来的损失时，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 做出决策时，其期望风险也必然最小。这样的决策就是最小风险Bayes决策。

- 最小风险贝叶斯决策规则：

$$\text{if } R(\alpha_i | x) = \min_{j=1,\cdots,k} R(\alpha_j | x), \text{ then } \alpha = \alpha_i$$

最小风险贝叶斯决策规则计算步骤

（1）计算后验概率：

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

（2）计算风险：

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

（3）决策：

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

两类情况：λ_{11}，λ_{12}，λ_{21}，λ_{22}

$$\lambda_{11}P(\omega_1 \mid x)+\lambda_{12}P(\omega_2 \mid x) \begin{array}{c} < \\ > \end{array} \lambda_{21}P(\omega_1 \mid x)+\lambda_{22}P(\omega_2 \mid x),$$

$$\text{则 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

或

$$(\lambda_{11}-\lambda_{21})P(\omega_1 \mid x) \begin{array}{c} < \\ > \end{array} (\lambda_{22}-\lambda_{12})P(\omega_2 \mid x)$$

$$\frac{P(\omega_1 \mid x)}{P(\omega_2 \mid x)} = \frac{p(x \mid \omega_1)P(\omega_1)}{p(x \mid \omega_2)P(\omega_2)} \begin{array}{c} > \\ < \end{array} \frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}} = \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

$$l(x) = \frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} \begin{array}{c} > \\ < \end{array} \frac{P(\omega_2)}{P(\omega_1)} \cdot \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

$$\text{则 } x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

显然，当 $\lambda_{11} = \lambda_{22} = 0$，$\lambda_{12} = \lambda_{21} = 1$ 时，最小风险就是最小错误率。

- 请认真学习课本例2.1和2.2，体会同样数据情况下，不同的损失会导致不同的决策

➢两种决策之关系：

损失函数：

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

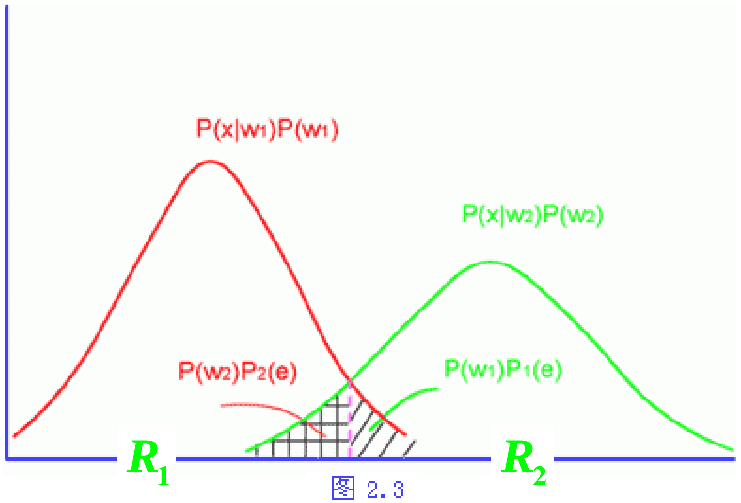
条件风险：

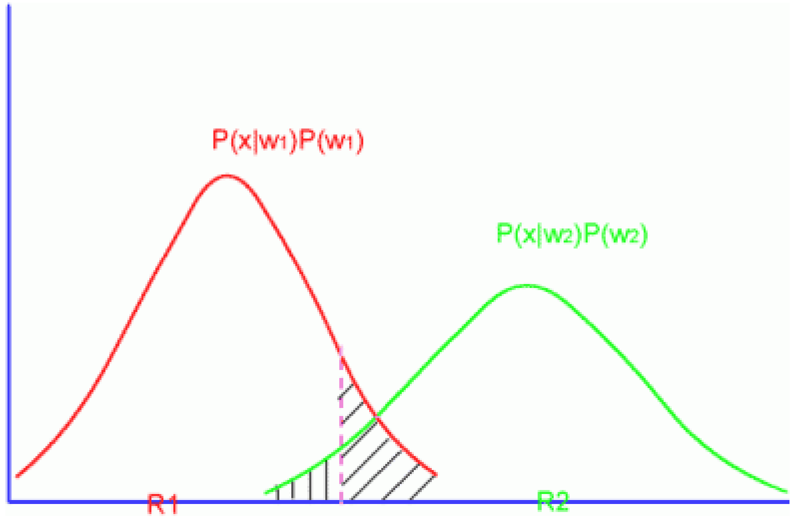
$$R\left(\alpha_i\middle|X\right)=\sum_{j=1}^{c}\lambda(\alpha_i|\omega_j)P\left(\omega_j\middle|X\right)=\sum_{j=1,j\neq i}^{c}P\left(\omega_j\middle|X\right)$$

而 $\sum_{j=1,j\neq i}^{c}P(\omega_j|X)$ 也恰恰是将X判为 ω_i 时的错误概率。因此基于最小风险的贝叶斯决策结果在0—1损失函数情况下，也就是基于最小错误概率的贝叶斯决策结果。由此可见，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。换句话说，前者是后者的特例。

实际上 $\sum_{i=1,\, j\neq i}^{c} P(\omega_j | X) = 1 - P(\omega_i | X)$

因此，当 $P(\omega_i | X)$ 最大时，$R_i(X)$ 最小。这与基于最小错误率的贝叶斯决策的判据是一样的。





- 用来表示最小风险贝叶斯决策方法的分类结果。与图2.3不同的是，R_1与R_2两个区域的分界线不再是t,而是向左移了一段距离，这是由于损失函数λ_{12}比λ_{21}大所造成(可以假设$\lambda_{11}=\lambda_{22}=0$)，在发生位移这一区域内，尽管$P\left(x|\omega_1\right)P(\omega_1)>P\left(x|\omega_2\right)P(\omega_2)$ ，但能够减少将ω_2错判为ω_1所带来的严重损失。当然平均错误率则明显增大了。

Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

限定一类错误率，使另一类错误率最小

- 某些实际情况，可能要求一类错误率控制在很小（不大于某数），在满足此条件的前提下再使另一类错误率尽可能小。

比如： $\min\ P_1(e)$ （对分类边界求最小）

$s.t.\ P_2(e) = \varepsilon,\qquad \varepsilon > 0$ 是个很小的常数

➢用Lagrange 乘子法：

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

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

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

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

R_1, R_2为两类的决策域

$$\left.\begin{array}{l}\dfrac{\partial L}{\partial t}=0\\[2ex]\dfrac{\partial L}{\partial \lambda}=0\end{array}\right\}\Rightarrow\left\{\begin{array}{l}\lambda=\dfrac{p(x|\omega_1)}{p(x|\omega_2)}\\[2ex]\int_{R_1}p(x|\omega_2)dx=\varepsilon\end{array}\right.$$

决策规则：

$$\lambda p(x\mid\omega_2)\ \begin{matrix}<\\>\end{matrix}\ p(x\mid\omega_1)\ 则\ x\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

or

$$\frac{p(x\mid\omega_1)}{p(x\mid\omega_2)}\ \begin{matrix}<\\>\end{matrix}\ \lambda\quad 则\ x\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

——也称**Neyman-Pearson**决策规则

通常 λ 在d较大时不易求取，可用似然比密度函数来确定 λ 值。

似然比：

$$l(x)=p(x|\omega_1)/p(x|\omega_2)$$

似然比密度函数：

$$p(l|\omega_2)$$

求解 $P_2(e)=1-\int_0^{\lambda} p(l|\omega_2)dl=\varepsilon$

当 $\lambda=0$ 时，$\Re_2=\phi$，$\Re_1=\Re$，$P_2(e)=1$（所有 ω_2 类均分到了 ω_1 类）

当 $\lambda\to\infty$ 时，$\Re_2=\Re$，$P_2(e)=0$，（所有样本都分到 ω_2，因此 ω_2 不会有错）

ROC曲线

状态与决策的可能关系

决策 \ 状态	阳性	阴性
阳性	真阳性（TP）	假阳性（FP）
阴性	假阴性（FN）	真阴性（TN）

阳性：表示某一症状存在，或者检测到某一指标的异常；

阴性：表示所要考察的症状不存在或者所监测的指标没有异常。

第一类错误率（Type-I error rate）= 假阳性率 = 假阳性样本数 / 总阴性样本数

第二类错误率（Type-II error rate）= 假阴性率 = 假阴性样本数 / 总阳性样本数

灵敏度（sensitivity） $Sn = \frac{TP}{TP + FN}$ 表示真正的阳性样本中有多少比例能被正确检测出来；

特异度（specificity） $Sp = \frac{TN}{TN + FP}$ 表示真正的阴性样本中有多少比例能被正确检测出来；

在医学应用背景下，一种诊断方法灵敏度高表示它能把有病的人都诊断出来；而特异性高则表示它不易把无病的人诊断为有病；

样本是100个人，其中90个人无病，10个人有病
诊断结果：
90个无病人群中，检测（有病，无病）=（2，88）
10个有病人群中，检测（有病，无病）=（9，1）

结果	真实阴性	真实阳性
测试阴性	88	1
测试阳性	2	9

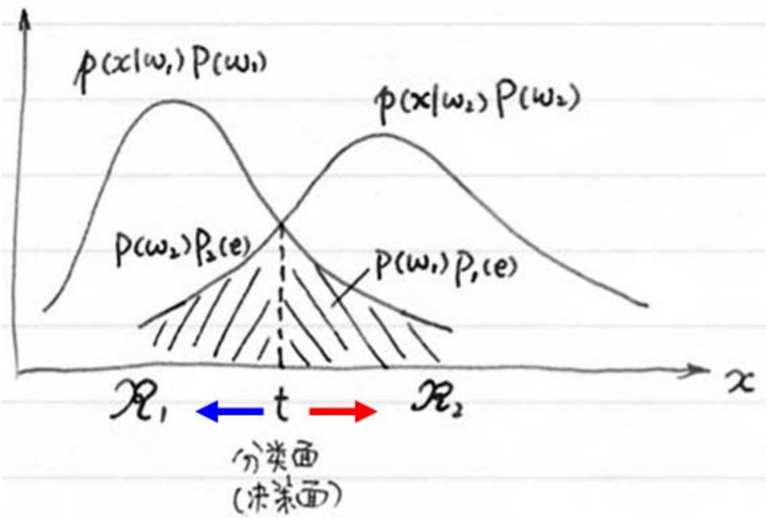
真阳性率（敏感性）=9/(1+9)=0.9

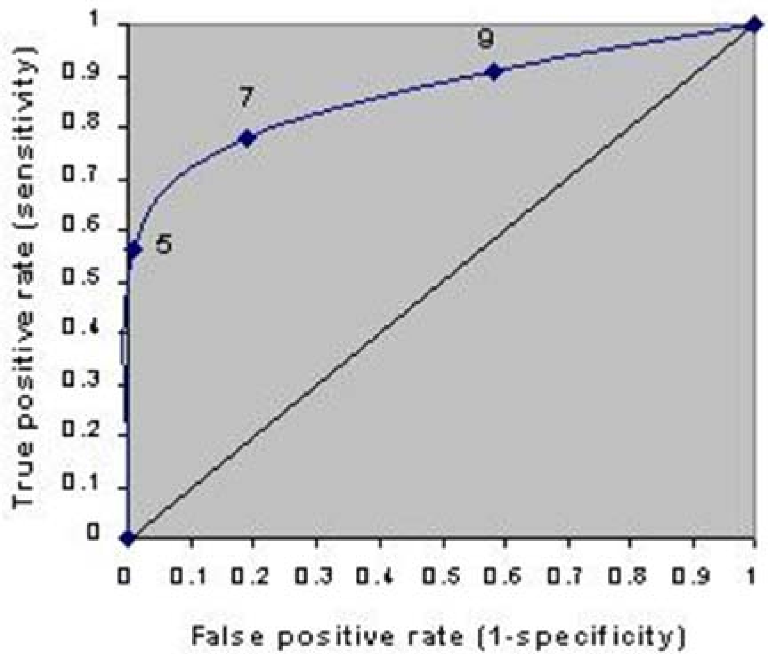
真阴性率（特异性）=88/(88+2)=0.9777

➢ 回顾Neyman-Pearson决策规则:

$$l(x)=\frac{p(x\,|\,\omega_1)}{p(x\,|\,\omega_2)}\ \begin{matrix} > \\ < \end{matrix}\ \lambda，\text{则}\ \ x\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

显然，变动λ会导致识别结果的变化，导致从一类错误变到另一类错误。





Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

判别函数、决策面与分类器设计

- 分类决策实质上是在描述待识别对象的d维特征所组成的特征空间内，将其划分为c个决策域，待识别的特征向量落在哪个决策域，该样本就被判为哪一类。因此决策域的**边界面**就是**决策面**，在数学上用解析形式表示成决策面方程。用于表达**决策规则**的某些函数则称为**判别函数**。显然判别函数与决策面方程是密切相关的，并且都是由相应决策规则所确定的。

➢例 最小错误率决策规则（两类）：

$$P(\omega_1 | X) > P(\omega_2 | X) \Rightarrow X \in \omega_1$$

- 则相应的判别函数就是：

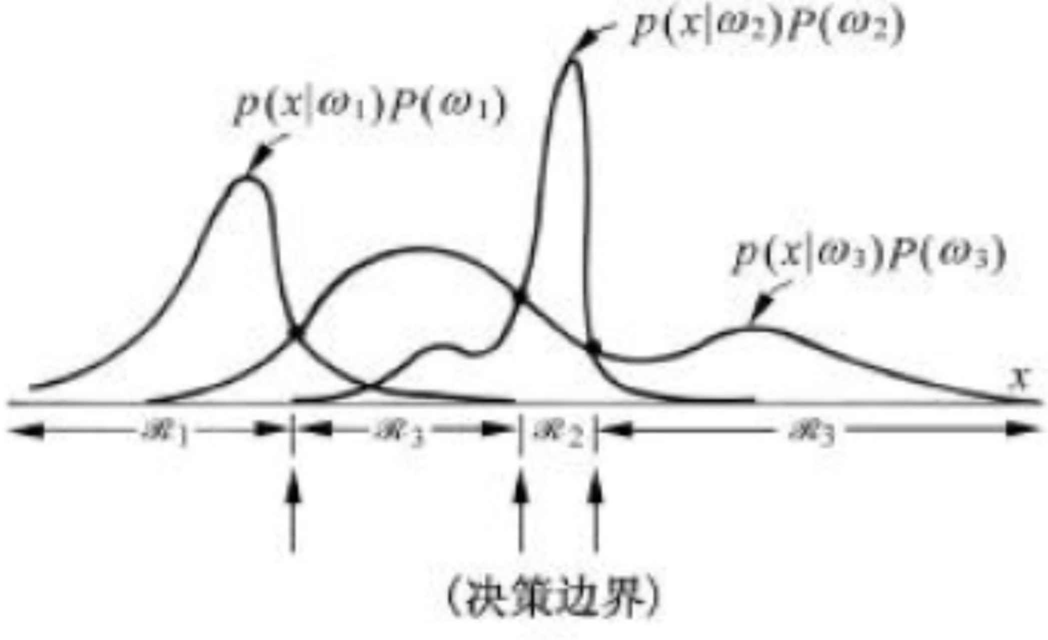
$$g_i(X) = P(\omega_1 | X), \quad i = 1, 2$$

- 而决策面方程则可写成：

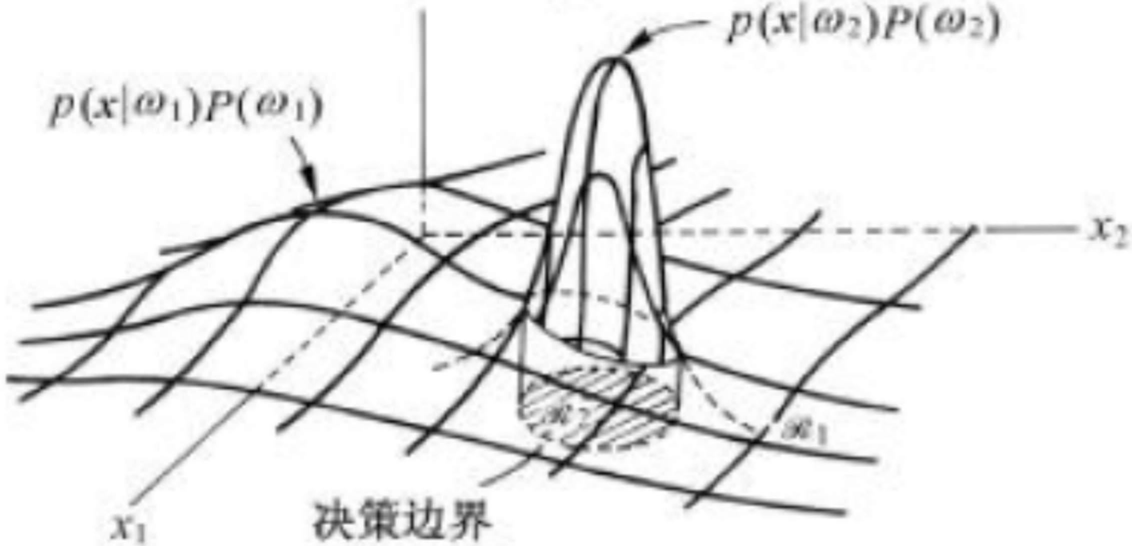
$$g_1(X) = g_2(X)$$

- 多类情况： $g_\mathrm{i}(X), \quad \mathrm{i} = 1, 2, \cdots, \mathrm{c}$
- 决策规则可表示成

$$\text{if } g_\mathrm{i}(X) = \max_\mathrm{j} g_\mathrm{j}(X) \Rightarrow X \in \omega_\mathrm{i}$$


p(x|ω1)P(ω1)
p(x|ω2)P(ω2)
p(x|ω3)P(ω3)
x
ℛ1
ℛ3
ℛ2
ℛ3
（决策边界）

(a)


p(x|ω1)P(ω1)
p(x|ω2)P(ω2)
x2
x1
ℛ2
ℛ1
决策边界

(b)

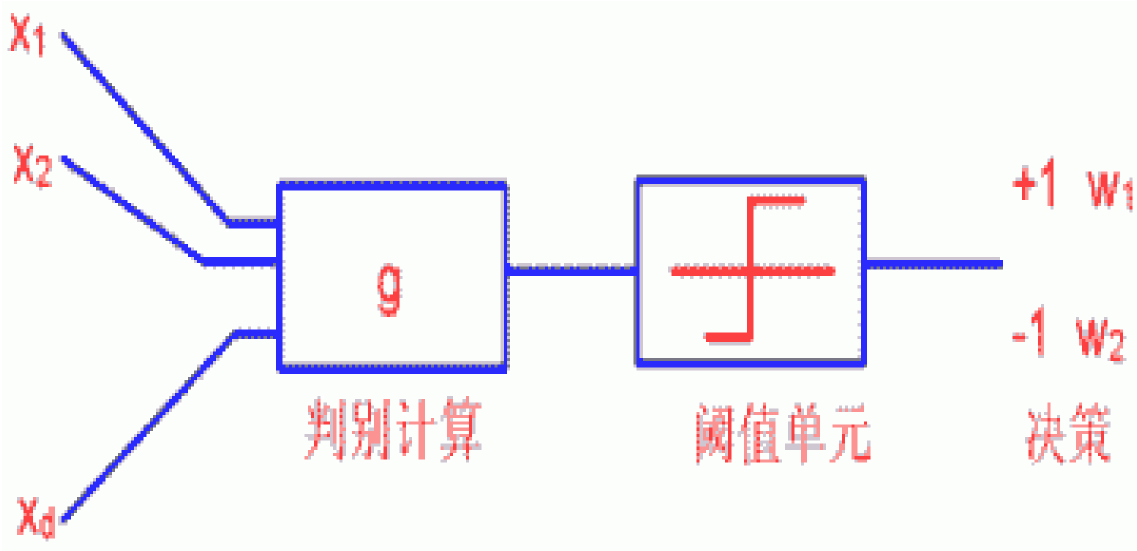
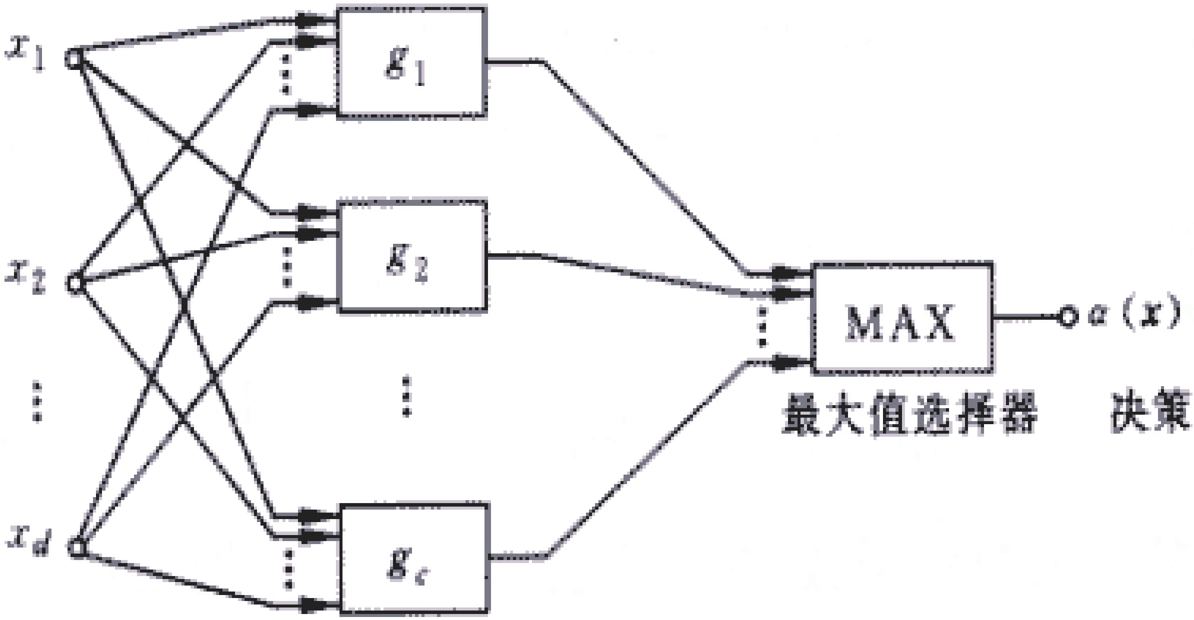


图 2.6



Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

正态分布时的统计决策

- (1)什么叫正态分布，或高斯分布，它是哪一种概率定义说的?
- (2)高斯分布的表达式：均值、协方差矩阵；
- (3)如何将正态分布与基于最小错误率的贝叶斯决策结合起来；
- (4)不同分类器的定义，如最小距离分类器、线性分类器等，这些定义也是比较重要的。

关于正态分布的知识

➢单变量

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

$$\mu=\boldsymbol{E}\{\boldsymbol{x}\}=\int \boldsymbol{p}(\boldsymbol{x})\boldsymbol{dx},$$

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

记作： $\boldsymbol{N}(\mu,\sigma^2)$

关于正态分布的知识

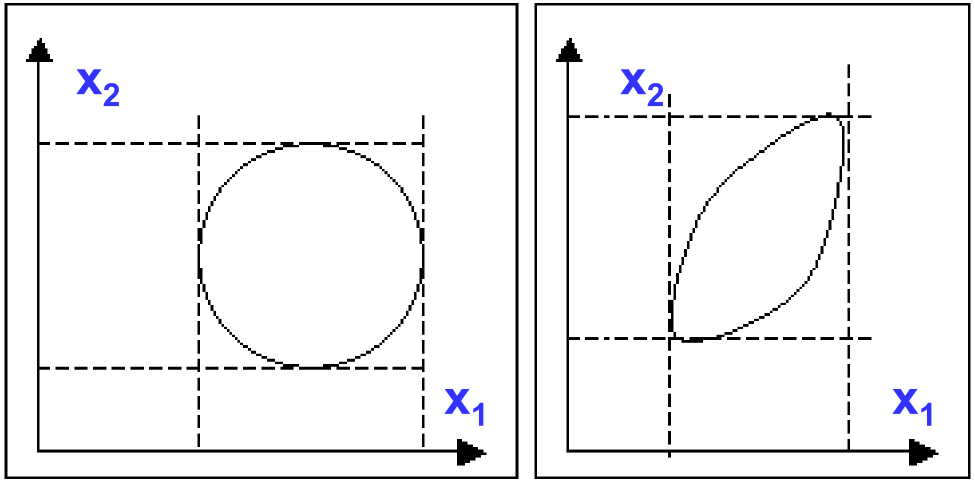
- 多变量

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

均值向量： $\mu = E[x]$

协方差矩阵： $\Sigma = E[(x-\mu)(x-\mu)^T]$

记作： $N(\mu, \Sigma)$



对右图来说，x_1和x_2有很大的相关性，而对左图来说，随机变量x_1与x_2之间的相关性很小。这可以从两者的区别看出来。对于右图可以看出一个随机变量的x_1分量较小时，另一分量x_2也必然较小。而当随机变量的x_1较大时，则其相应的x_2分量也较大。

$E\left[\left(x_2-\mu_2\right)\left(x_1-\mu_1\right)\right]$衡量两种向量的相关性，称为协方差

第二章 统计决策方法

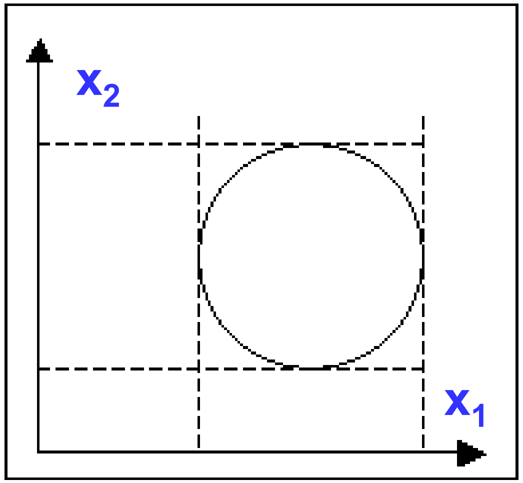
如X表示一个二维向量:

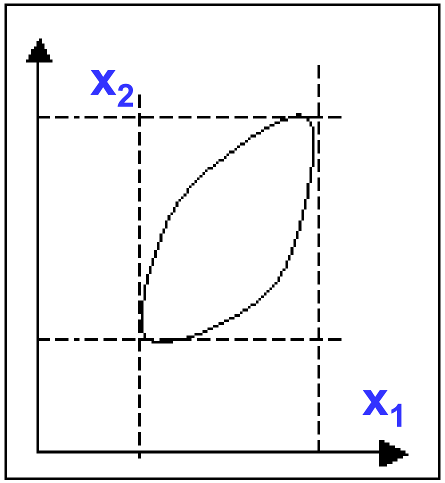
$$\Sigma=\begin{bmatrix} E[(x_1-\mu_1)^2] & E[(x_1-\mu_1)(x_2-\mu_2)] \\ E[(x_2-\mu_2)(x_1-\mu_1)] & E[(x_2-\mu_2)^2] \end{bmatrix}=\begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{bmatrix}$$

Σ称为协方差矩阵。

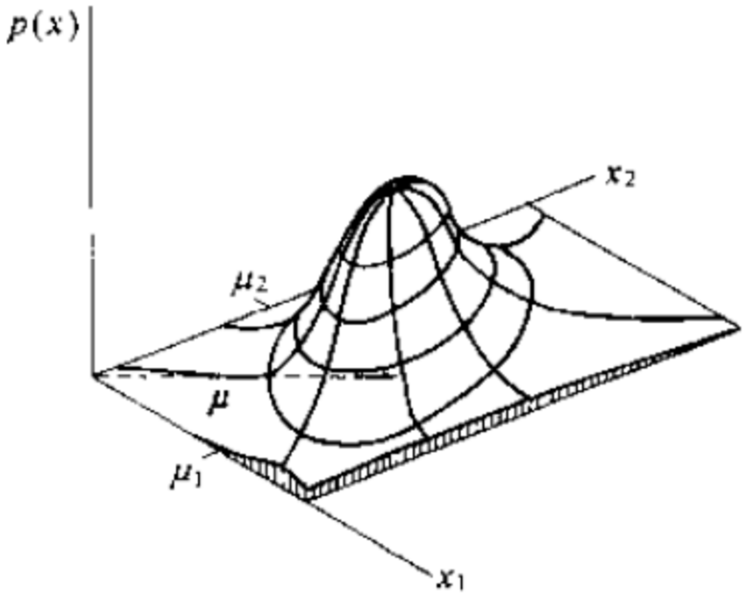
$$\begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

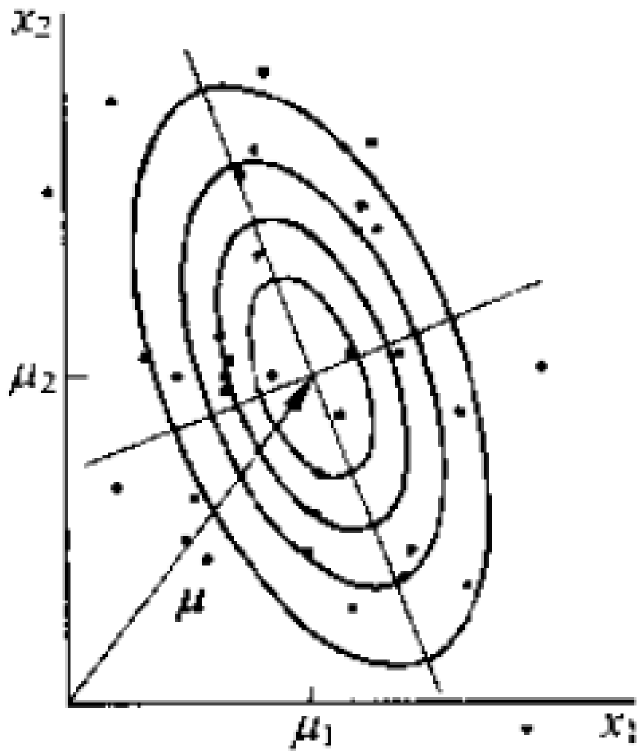
请问哪个是左图，
哪个是右图？





正态分布的等密度点的轨迹为超椭球面





正态分布 的性质：

（1） $p(x) \sim N(\mu, \Sigma)$。 $d + d(d+1)/2$ 个参数决定

（2）等密度点形成超椭球面 $(x-\mu)^T \Sigma^{-1}(x-\mu) = C \stackrel{\Delta}{=} \gamma^2$

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。

（3）对于正态分布的随机变量，不相关等价于独立

不相关：$E\left[x_i^T x_j\right] = E\left[x_i^T\right] E\left[x_j\right]$

独立：$p(x_i, x_j) = p(x_i) p(x_j)$， $\forall x_i, x_j$。 $\Rightarrow$ 不相关

推论：Σ 是对角阵 $\Leftrightarrow$ x 的各分量相互独立

（4）边缘分布仍是正态分布 $p(x_i) \sim N(\mu_i, \sigma_{ii}^2)$

条件分布也是正态分布。

正态分布 的性质：

（5）线性变换仍是正态分布

$$p(x) \sim N(\mu, \Sigma), \quad y = Ax$$

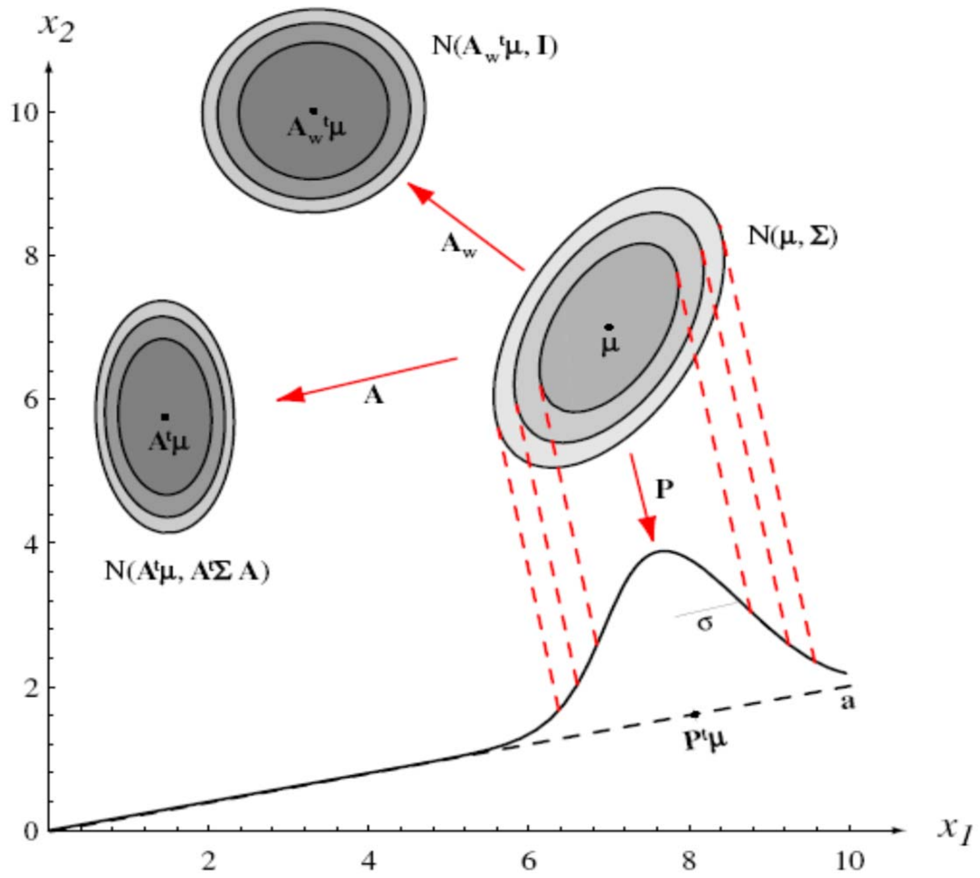
$$p(y) \sim N(A\mu, A\Sigma A^T)$$

（6）线性组合仍是正态分布（线性变换的特例）

$$p(x) \sim N(\mu, \Sigma), \quad y = a^T x$$

$$p(y) \sim N(a^T \mu, a^T \Sigma a)$$

特征空间中的一个线性变换将一个任意正态分布变换成另一个正态分布



Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

正态分布下的贝叶斯决策

$$p(x\mid\omega_i)=\frac{1}{(2\pi)^{d/2}\mid\sum_i\mid^{1/2}}\exp\left[-\frac{1}{2}(x-\mu_i)^T\sum_i^{-1}(x-\mu_i)\right]$$

考虑判别函数

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

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

决策面方程 $g_i(x)=g_j(x)$

$$-\frac{1}{2}\left[(x-\mu_i)^T\sum_i^{-1}(x-\mu_i)-(x-\mu_j)^T\sum_j^{-1}(x-\mu_j)\right]-\frac{1}{2}\ln\frac{\mid\sum_i\mid}{\mid\sum_j\mid}+\ln\frac{P(\omega_i)}{P(\omega_j)}=0$$

正态分布下的贝叶斯决策

一、$\Sigma_i = \sigma^2 I \quad i = 1,\cdots,c$

●如果 $P(\omega_i) \quad i = 1,\cdots,c$ 相等，可略去判别函数中与类别无关的项

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

$$g_i(X) = -(X-\mu_i)^T(X-\mu_i) = -\|X-\mu_i\|^2$$

最小距离分类器

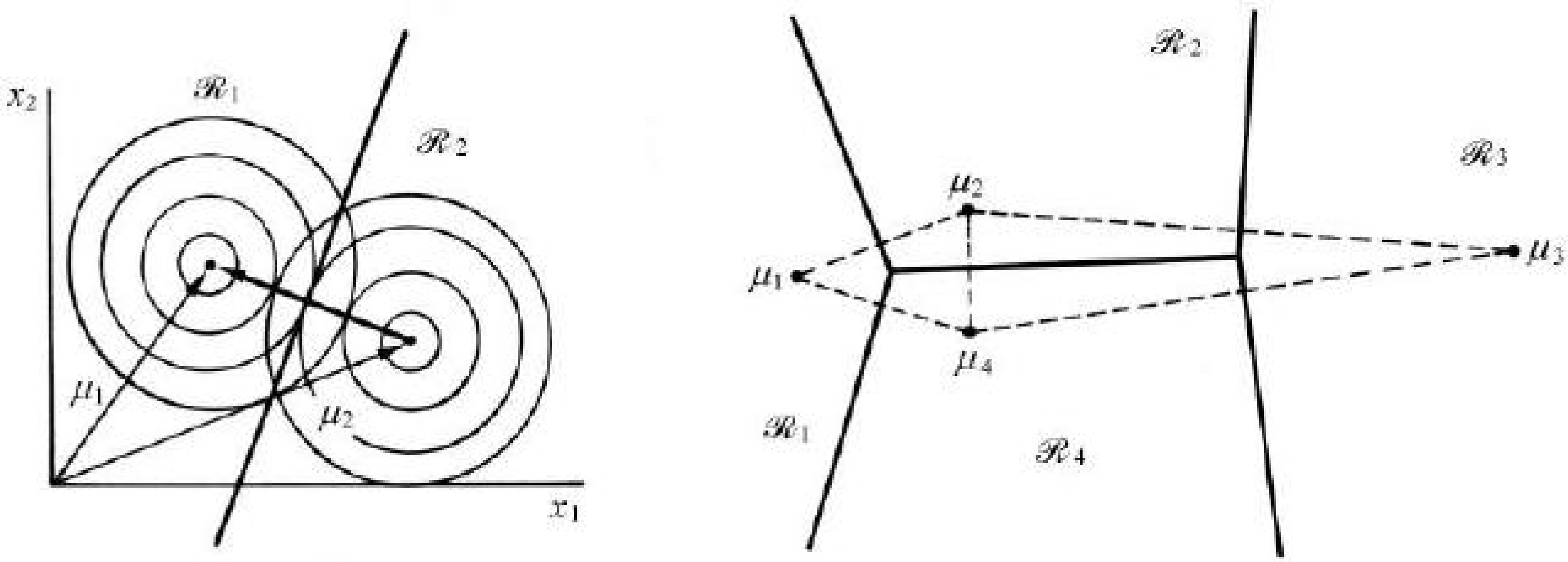
每个样本以它到每类样本均值的欧式距离的最小值确定其分类

正态分布下的贝叶斯决策

一、$\Sigma_i = \sigma^2 I \quad i = 1, \cdots, c$

- 如果 $P(\omega_i) \quad i = 1, \cdots, c$ 相等，可略去判别函数中与类别无关的项

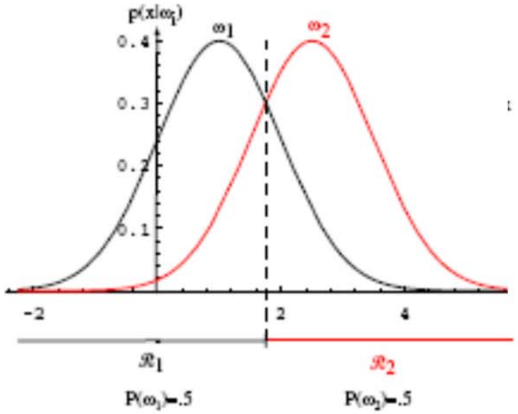
$$g_i(X) = -(X - \mu_i)^T (X - \mu_i) = -\|X - \mu_i\|^2$$

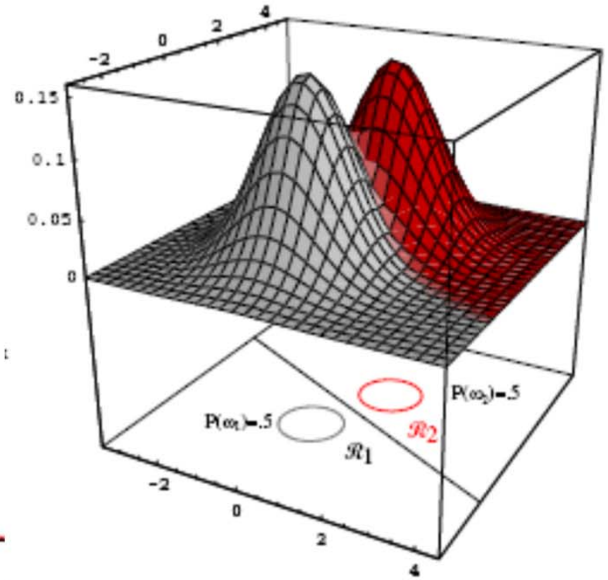


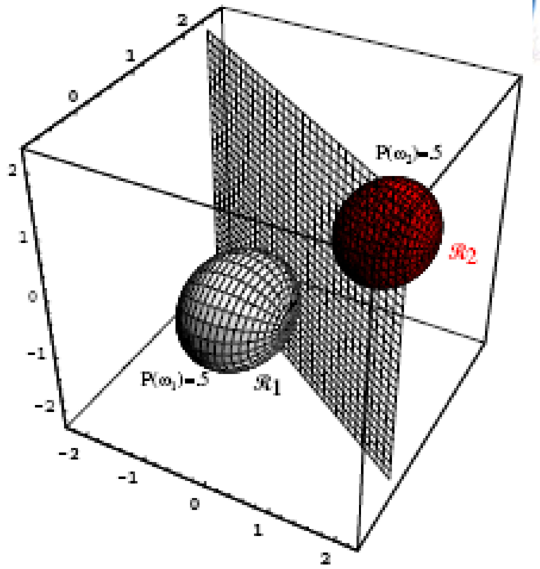
(a) 两类情况　　(b) 多类情况

在$\sum = \sigma^2 I, P(\omega_1) = P(\omega_2)$情况下，

显示$p(X|\omega_i)$和判别函数的







如果两种分布的协方差矩阵相等并且与单位阵成比例，那么它们呈**d**维球状分布，其判决边界是一个归一化超平面，垂直于两个中心的连线。

第二章 统计决策方法

正态分布下的贝叶斯决策

一、$\Sigma_i = \sigma^2 I \quad i = 1, \cdots, c$

● 如果 $P(\omega_i)$，$i = 1, \cdots, c$ 不相等，得

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

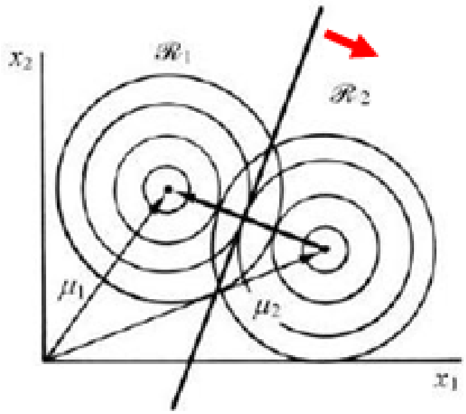
再略去与 i 无关的项 $x^T x$，整理可得

$$g_i(x) = w_i^T x + b_i$$ —— 线性判别函数

其中
$$w_i = \frac{1}{\sigma^2}\mu_i$$

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

决策面向先验概率小的方向偏移



正态分布下的贝叶斯决策

二、 $\Sigma_i = \Sigma \quad i = 1, \cdots, c$

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

令 $\gamma^2 = (x-\mu_i)^T \Sigma^{-1}(x-\mu_i)$------ x 到 μ_i 的 Mahalanobis 距离的平方

若 $P(\omega_i)$ 相等，则分类取决于样本到类中心的 Mahalanobis 距离。

一般，可得

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

略去 $x^T \Sigma^{-1} x$ 项，得

$$g_i(x) = w_i^T x + b_i$$ ——线性判别函数

其中， $w_i = \Sigma^{-1} \mu_i$， $b_i = -\frac{1}{2}\mu_i^T \Sigma^{-1} \mu_i + \ln P(\omega_i)$

正态分布下的贝叶斯决策

二、 $\Sigma_i = \Sigma \quad i = 1, \cdots, c$

决策面方程： $g_i(x) = g_j(x)$

可写为 $w^T(x - x_0) = 0$

其中 $w = \Sigma^{-1}(\mu_i - \mu_j)$

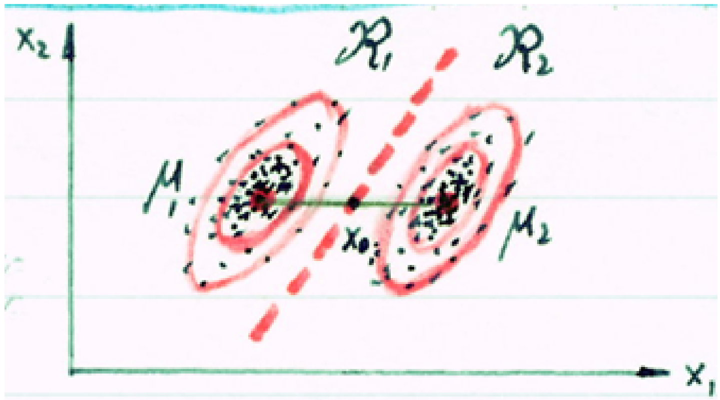
$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln(P(\omega_i)/P(\omega_j))}{(\mu_i - \mu_j)^T \Sigma^{-1} (\mu_i - \mu_j)}(\mu_i - \mu_j)$$

μ_i 到 μ_j 的马氏距离平方。

当 $P(\omega_i) = P(\omega_j)$ 时， $x_0 = \frac{1}{2}(\mu_i + \mu_j)$

Mahalanobis 距离考虑了方差因素。

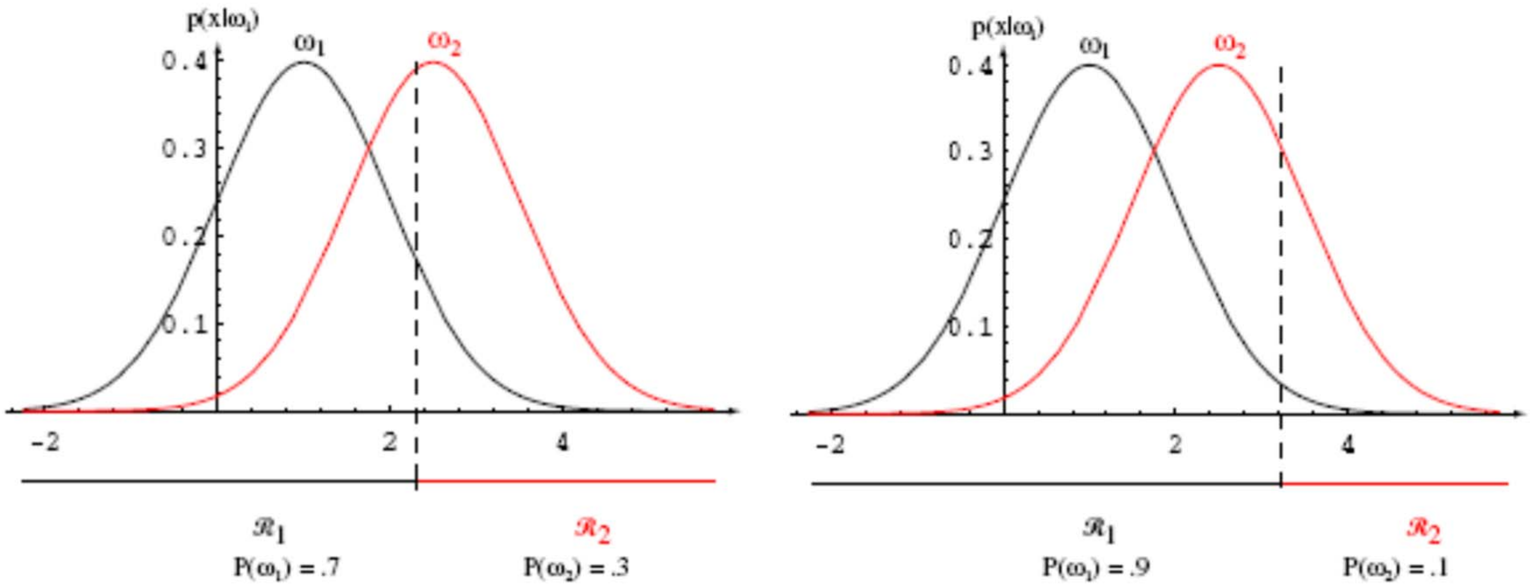
当 $\Sigma = I$ 时就是欧氏距离。



正态分布下的贝叶斯决策

二、 $\Sigma_i = \Sigma \quad i = 1, \cdots, c$

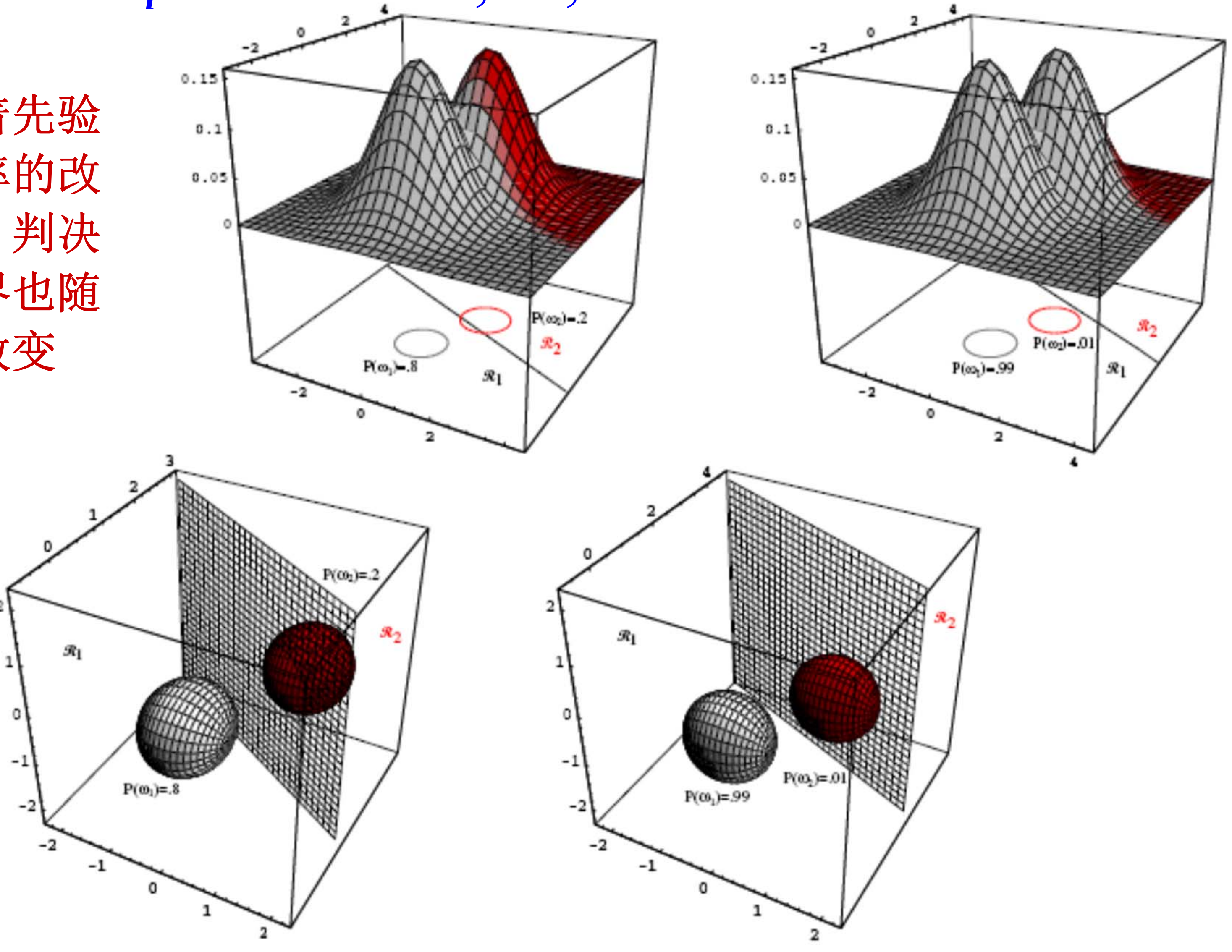
随着先验概率的改变，判决边界也随之改变



正态分布下的贝叶斯决策

二、 $\Sigma_i = \Sigma \quad i = 1, \cdots, c$

随着先验概率的改变，判决边界也随之改变



正态分布下的贝叶斯决策

三、一般情况：各类协方差 矩阵不等

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

$$= x^T W_i x + w_i^T x + w_{i0}$$

其中 $W_i = -\frac{1}{2}\Sigma_i^{-1}$

$$w_i = \Sigma_i^{-1}\mu_i$$

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

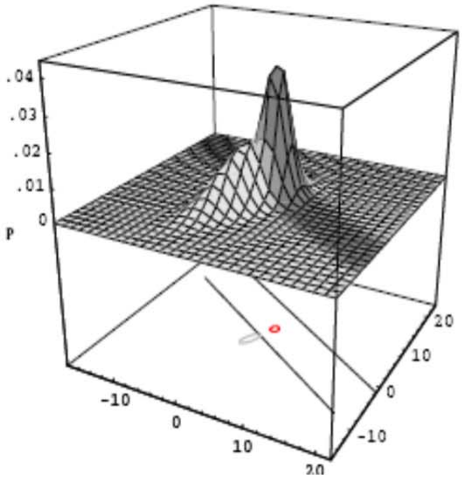
决策面 $g_i(x) = g_j(x)$

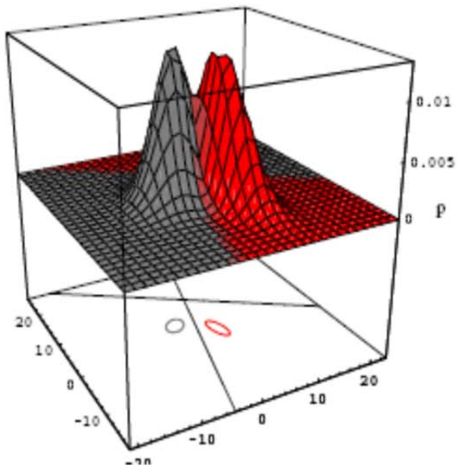
$$x^T(W_i - W_j)x + (w_i - w_j)^T x + w_{i0} - w_{j0} = 0$$ 为超二次曲面

正态分布下的贝叶斯决策

三、一般情况：各类协方差 矩阵不等

任意二维正态分布导致一般超二次曲面的**Bayes**判决边界

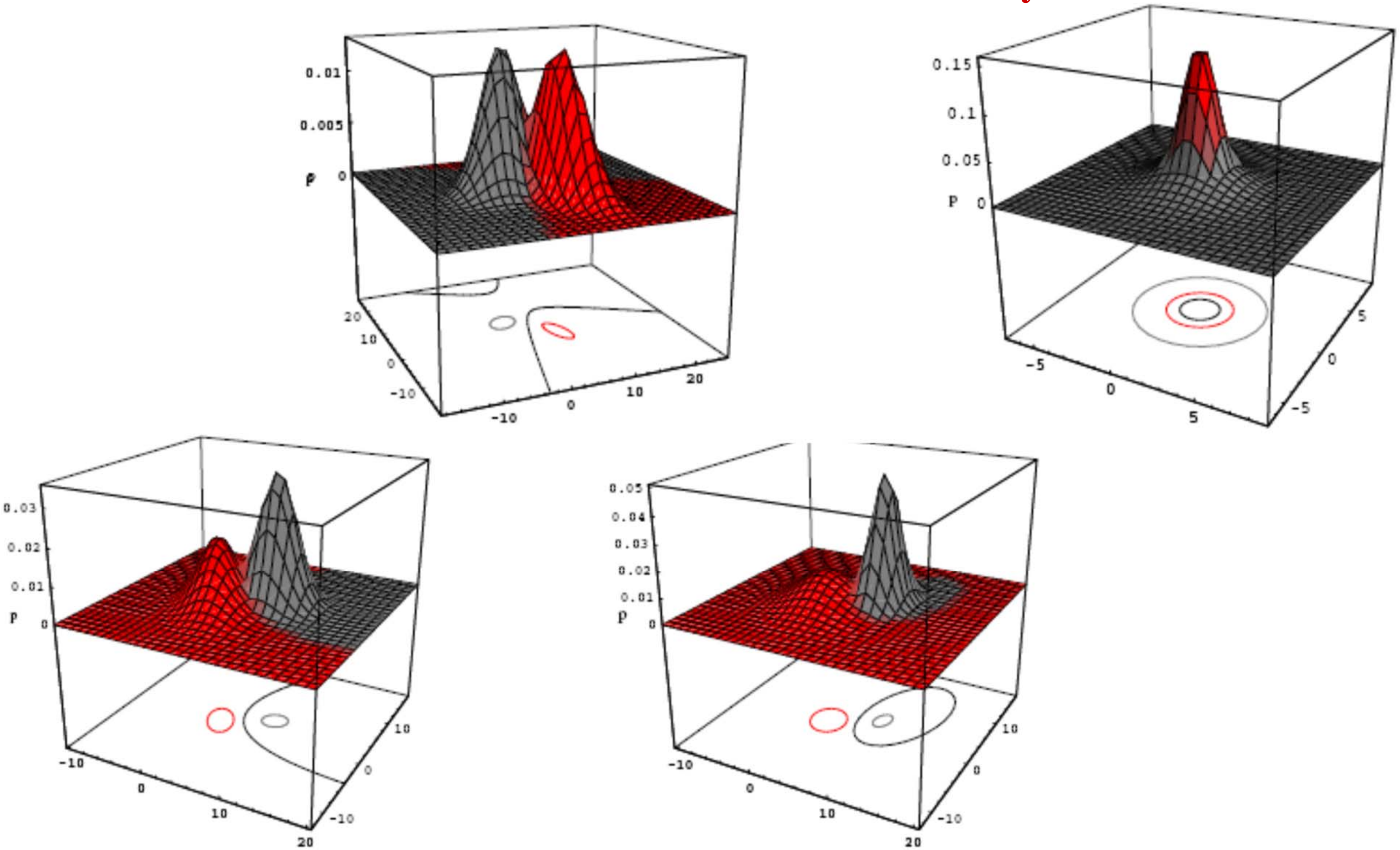




正态分布下的贝叶斯决策

三、一般情况：各类协方差 矩阵不等

任意二维正态分布导致一般超二次曲面的**Bayes**判决边界



Outline:

- 引言
- Bayes决策
  - ✓ 基于最小错误率的Bayes决策
  - ✓ 基于最小风险的Bayes决策
- Neyman-Pearson决策与ROC曲线
- 判别函数、决策面与分类器设计
- 正态分布的统计决策
  - ✓ 正态分布及性质
  - ✓ 正态分布概率模型下的最小错误率Bayes决策
- 错误率问题

错误率问题

- 研究错误的意义
  - ✓对于同一问题，不同的错误率反映了方法的优劣

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

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

错误率问题

- 三类处理方法：
  - ✓按理论公式计算
  - ✓计算错误率上限（估计）
  - ✓实验估计：
    - 交叉验证（cross-validation）：leave-one-out, 10-fold, …

错误率问题

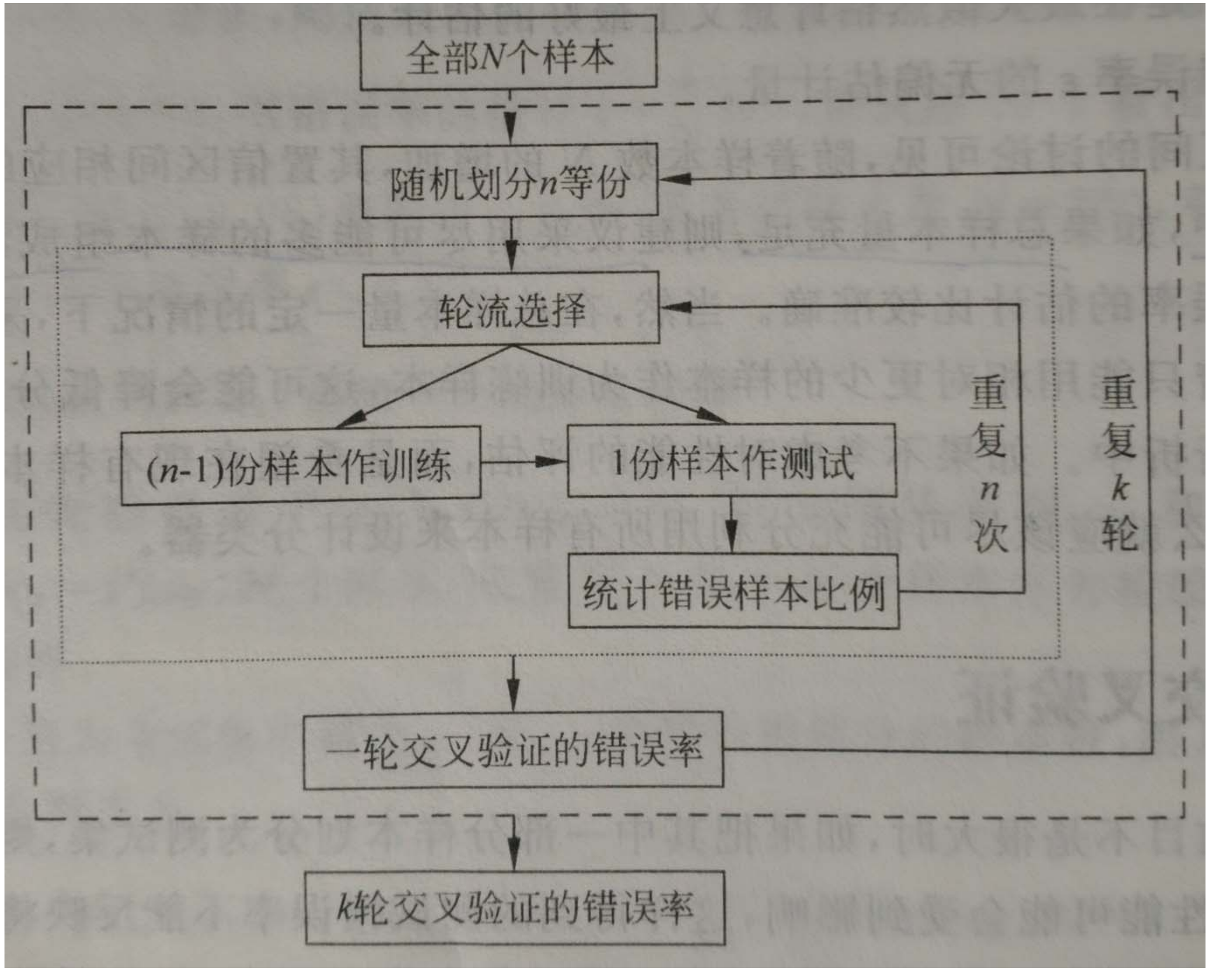
- 交叉验证法：
  - ✓ 基本思想：是在现有总样本不变的情况下，随机选用一部分样本作为临时的训练集，用剩余样本作为临时的测试集，得到一个错误率估计；然后随机选用另外一部分样本作为临时训练集，其余样本作为临时测试集，再得到一个错误率估计……如此反复多次，最后对所有错误率求平均。

错误率问题

- 交叉验证法：
  - ✓ n倍交叉验证法(n-fold cross-validation)：把全部样本随机地划分为n个等份，在一轮实验中轮流抽出其中的1份样本作为测试样本，用其余(n-1)份作为训练样本，得到n个错误率后进行平均，作为一轮交叉验证的错误率；由于对样本的一次划分是随意的，人们往往进行多轮这样的划分(比如k轮)，得到多个交叉验证错误率估计，最后将多个估计再求平均。

第二章 统计决策方法

错误率问题



n倍交叉验证法(n-fold cross-validation)示意图

错误率问题

- 交叉验证法：
  - ✓ 留一法交叉验证(leave-one-out cross-validation)：不把样本进行分组，而是每轮实验拿出一个样本来作为测试样本，用其余的N-1个样本作为训练样本集，训练分类器，测试对留出的那个样本的分类是否正确；在下一轮试验中，把之前测试的样本放回，拿出另外一个样本作为测试样本，用剩余的N-1个样本做训练，再对留出的样本做测试；以此类推，直到每个样本都被作为测试样本一次。全部N轮实验完成后，统计总共出现的测试错误数占总样本数的比例就是留一法交叉验证错误率。