第二章

贝叶斯统计决策理论

目 录

2.1 引言

- **确定性的模式：**如果试验对象和测量条件相同，所有测量值具有重复性，即在多次测量中，它们的结果不变，获得的模式称为确定性的模式。
- **随机模式：**测量值是随机的，这样的模式称为随机性的模式，简称随机模式。
- 下面介绍随机模式的分类方法。与概率论及数理统计知识密切相关。

2.1 引言

- 基本概念
  - 先验概率：先验概率是预先已知的或者可以估计的模式识别系统位于某种类型的概率。
  - 类（条件）概率密度：它是系统位于某种类型条件下，模式样本**x**出现的概率密度分布函数，常用 $\rho(\mathbf{x}\,|\,A), \rho(\mathbf{x}\,|\,B)$ 表示。
  - 后验概率：它是系统在某个具体的模式样本**x**条件下，位于某种类型的概率，常以 $p(A\,|\,\mathbf{x}), p(B\,|\,\mathbf{x})$ 表示。

2.1 引言

– Bayes法则

设有R类样本，分别为 $\omega_1, \omega_2, ..\omega_R$，若每类的先验概率为 $p(\omega_i), i = 1, 2, .., R$。对于一随机矢量$\mathbf{x}$，每类的条件概率为 $\rho(\mathbf{x} \mid \omega_i)$

根据Bayes公式，后验概率为：

$$p(\omega_i \mid \mathbf{x}) = \frac{\rho(\mathbf{x} \mid \omega_i) p(\omega_i)}{\sum_{k=1}^{R} p(\omega_k) p(\mathbf{x} \mid \omega_k)}$$

2.2 判决规则

- 判决规则定义

  目标：给定变量属于哪一类的规则(判决函数)

  这个规则会将输入空间划分成几个决策区域 R_k

  每个区域对应一个类别，如果变量落入了R_k我们就判定它属于 ω_k这个类别

  规则的定义依据后验概率的关系给出

$$L = \{p(\omega_1 | \mathrm{x}), p(\omega_2 | \mathrm{x}), p(\omega_3 | \mathrm{x}), \ldots, p(\omega_R | \mathrm{x})\}$$

2.2 判决规则

- 判决规则
  - 最小错误率判决规则
  - 最小风险判决规则
  - **Neyman-Pearsen**判决规则
  - 最大似然比判决规则
  - 最小最大判决规则

2.2.1 最小错误率判决规则

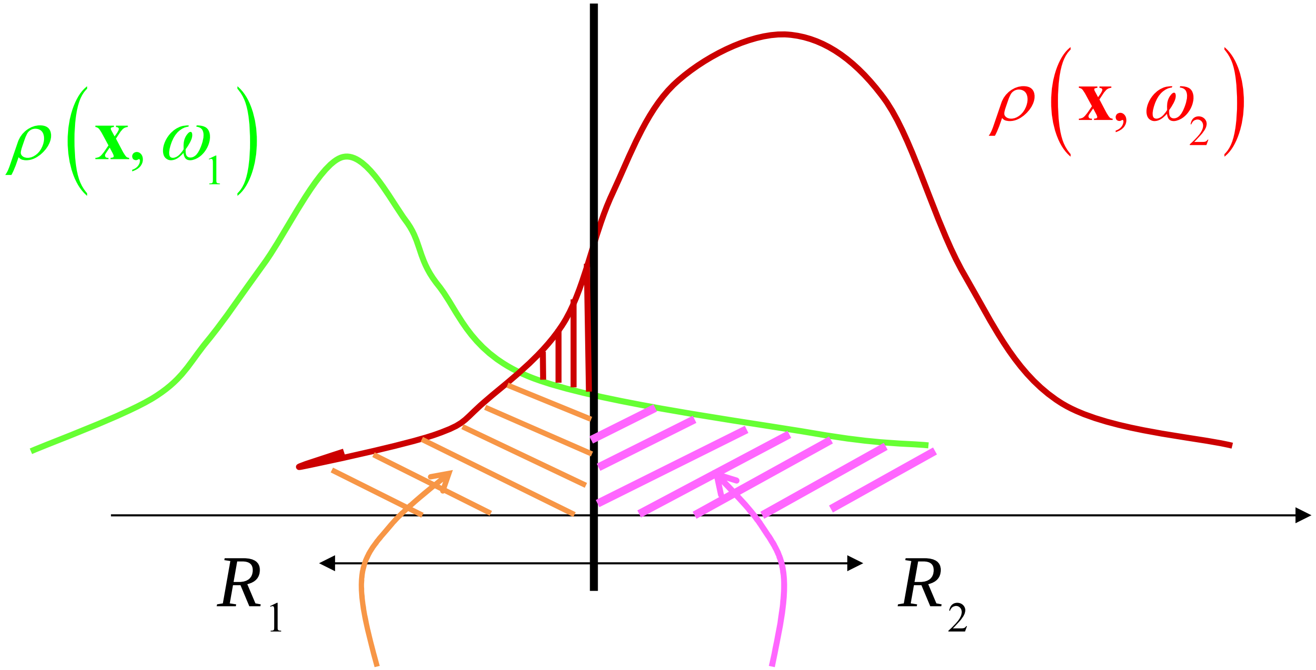
- 目标：尽可能减少错分类
- 以两类问题为例。
- 在两类问题中若出现错误，应该是本该是属于 ω_1类别的变量却被判定为ω_2，或者是本该属于ω_2 类别的变量被判定为 ω_1，这个错误的概率可以表示为：

$$p(\text{mistake}) = p(\mathbf{x} \in \mathrm{R}_1, \omega_2) + p(\mathbf{x} \in \mathrm{R}_2, \omega_1)$$

$$p(\text{mistake}) = \int_{R_1} p(\mathbf{x}, \omega_2) d\mathbf{x} + \int_{R_2} p(\mathbf{x}, \omega_1) d\mathbf{x}$$

最小错误率判决规则

- 目标：使 $p(\text{mistake})$最小



最小错误率判决规则

- 为了使 $p(\text{mistake})$ 最小，x应该被判定属于能够使上述积分达到较小值的类别
- 如果 $p(\mathbf{x},\omega_1) > p(\mathbf{x},\omega_2)$，此时我们将变量**x**归于类别$\omega_1$，这样一来在 R_1 区域$p(\mathbf{x},\omega_2)$的值较小，上述积分右边第一部分就达到最小了；第二部分的道理同上。

$$\begin{cases} p(\mathbf{x},\omega_1) > p(\mathbf{x},\omega_2), & \mathbf{x} \in \omega_1 \\ p(\mathbf{x},\omega_1) < p(\mathbf{x},\omega_2), & \mathbf{x} \in \omega_2 \\ p(\mathbf{x},\omega_1) = p(\mathbf{x},\omega_2), & otherwise \end{cases}$$

最小错误率

最小错误率判决规则

- 根据贝叶斯公式，可知$p(\mathbf{x},\omega_k)=p(\omega_k\mid\mathbf{x})p(\mathbf{x})$，

等价于

$$\boxed{\begin{cases} p(\omega_1\mid\mathbf{x})>p(\omega_2\mid\mathbf{x}), & then\ \mathbf{x}\in\omega_1 \\ p(\omega_1\mid\mathbf{x})<p(\omega_2\mid\mathbf{x}), & then\ \mathbf{x}\in\omega_2 \\ p(\omega_1\mid\mathbf{x})=p(\omega_2\mid\mathbf{x}), & otherwise \end{cases}}$$

最小错误率判决规则

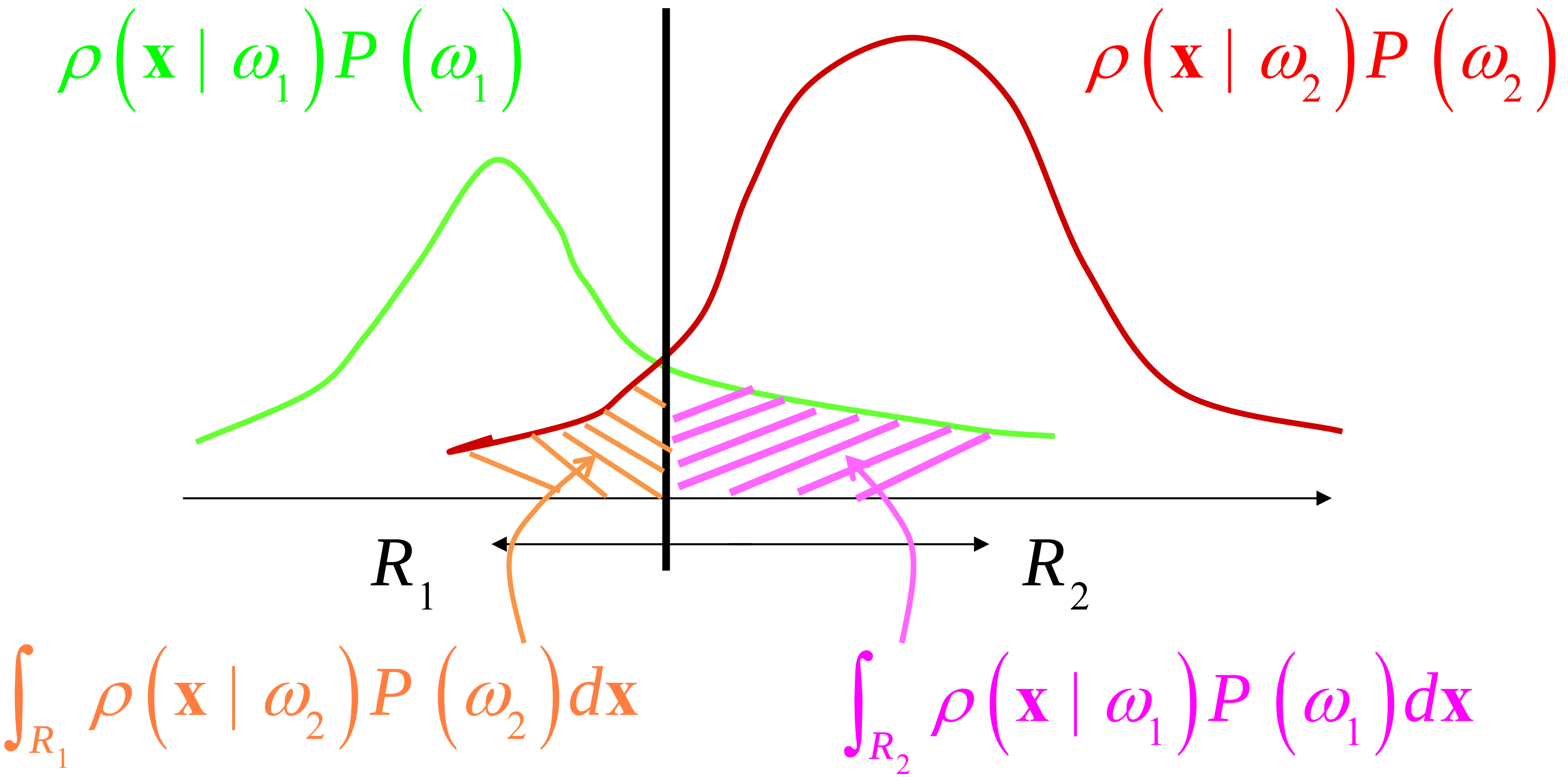
$$\begin{aligned} &p(\mathrm{x}\mid\omega_1)\cdot p(\omega_1)>p(\mathrm{x}\mid\omega_2)\cdot p(\omega_2), && then && \mathrm{x}\in\omega_1 \\ &p(\mathrm{x}\mid\omega_1)\cdot p(\omega_1)<p(\mathrm{x}\mid\omega_2)\cdot p(\omega_2), && then && \mathrm{x}\in\omega_2 \\ &p(\mathrm{x}\mid\omega_1)\cdot p(\omega_1)=p(\mathrm{x}\mid\omega_2)\cdot p(\omega_2), && otherwise && \end{aligned}$$

最小错误率判决规则

当 p(mistake)最小时，$\mathbf{x} \in \omega_1$ 但却落入R_2中；

$\mathbf{x} \in \omega_2$ 但却落入R_1中；

上图中的红色区域面积为零



最小错误率判决规则

故总错误率$p(\text{mistake})$

$$\begin{aligned} p(\text{mistake}) &= p(\mathbf{x} \in R_2, \omega_1) + p(\mathbf{x} \in R_1, \omega_2) \\ &= p(\mathbf{x} \in R_2 \mid \omega_1) p(\omega_1) + p(\mathbf{x} \in R_1 \mid \omega_2) p(\omega_2) \\ &= \int_{R_2} p(\mathbf{x} \mid \omega_1) p(\omega_1) d\mathbf{x} + \int_{R_1} p(\mathbf{x} \mid \omega_2) p(\omega_2) d\mathbf{x} \end{aligned}$$

其中 $\varepsilon_1 = \int_{R_2} p\left(\mathbf{x} \mid \omega_1\right) d\mathbf{x}$ 表示将类1的样本分到类2的错误率，$\varepsilon_2 = \int_{R_1} p\left(\mathbf{x} \mid \omega_2\right) d\mathbf{x}$ 表示将类2的样本分到类1的错误率。

举例1

- 为了对癌症进行诊断，对一批人进行一次普查，各每个人打试验针，观察反应，然后进行统计，规律如下：

（1）这一批人中，每1000个人中有5个癌症病人；

（2）这一批人中，每100个正常人中有一个试验呈阳性反应；

（3）这一批人中，每100个癌症病人中有95人试验呈阳性反应。

问：若某人（甲）呈阳性反应，甲是否正常？

解答

- 假定x表示实验反应为阳性，

（1）人分为两类：w_1—正常人，w_2—癌症患者

$p(w_1)+p(w_2)=1$

（2）由已知条件计算概率值：

先验概率：$p(w_1)=0.995$， $p(w_2)=0.005$

类条件概率密度：$p(x|w_1)=0.01$, $p(x|w_2)=0.95$

（3）决策过程：

解答

$$p(w2\mid x)=\frac{p(x\mid w_2)\cdot p(w_2)}{p(x\mid w_1)\cdot p(w_1)+p(x\mid w_2)\cdot p(w_2)}$$

$$=\frac{0.95\times 0.005}{0.01\times 0.995+0.95\times 0.005}$$

$$=0.323$$

$$p(x\mid w_1)\cdot p(w_1)=0.00995,\quad p(w_1\mid x)=1-p(w_2\mid x)=1-0.323=0.677$$

$$p(x\mid w_2)\cdot p(w_2)=0.00475$$

$$p(w_1\mid x)>p(w_2\mid x)\qquad p(x\mid w_1)\cdot p(w_1)>p(x\mid w_2)\cdot p(w_2)$$

- 由最小错误判决规则，可知：甲 $\in w_1$

最小错误率几种特例

最小错误率判决规则(判决函数)

$$
\begin{cases}
p(\omega_1 \mid \mathbf{x}) > p(\omega_2 \mid \mathbf{x}), & then\ \ \mathbf{x} \in \omega_1 \\
p(\omega_1 \mid \mathbf{x}) < p(\omega_2 \mid \mathbf{x}), & then\ \ \mathbf{x} \in \omega_2 \\
p(\omega_1 \mid \mathbf{x}) = p(\omega_2 \mid \mathbf{x}), & otherwise
\end{cases}
$$

- 对于一个两类问题，有如下结论：

1.若对于某一样本**x**，有 $\rho\left(\mathbf{x} \mid \omega_1\right) = \rho\left(\mathbf{x} \mid \omega_2\right)$，则说明**x**没有提供关于类别状态的任何信息，完全取决于先验概率。

2.若 $p(\omega_1) = p(\omega_2)$，则判决完全取决于条件概率。

最小错误率—最大似然比

若 $\mathbf{x} \in w_1$，最小错误率判决规则：

$$p(\mathbf{x} \mid w_1) \cdot p(w_1) > p(\mathbf{x} \mid w_2) \cdot p(w_2)$$

两边同时除以 $p(\mathbf{x} \mid w_2) \cdot p(w_1)$ 有：

$$\frac{p(\mathbf{x} \mid w_1)}{p(\mathbf{x} \mid w_2)} > \frac{p(w_2)}{p(w_1)}$$

定义类型w_1与w_2的似然比为：

$$l_{12}(\mathbf{x}) = \frac{p(\mathbf{x} \mid w_1)}{p(\mathbf{x} \mid w_2)}$$

最小错误率—最大似然比

对例1用似然比的方法求解

先验概率：$p(w_1)$=0.995， $p(w_2)$=0.005

类条件概率密度：$p(\mathbf{x}|w_1)$=0.01, $p(\mathbf{x}|w_2)$=0.95

决策过程：

$$l_{12}(\mathbf{x})=\frac{p(\mathbf{x}|w_1)}{p(\mathbf{x}|w_2)}=\frac{0.01}{0.95}=0.0105$$

$$\theta_{12}=\frac{p(w_2)}{p(w_1)}=\frac{0.005}{0.995}=0.005$$

$l_{12}(\mathbf{x})>\theta_{12}$，根据似然比的判定规则可知：甲$\in w_1$，与例1结果一致。

最小错误率—最大似然比

则判决门限为： $\theta_{12} = \dfrac{p(w_2)}{p(w_1)}$

一般先验概率已知，θ_{12} 也就已知了。

所以“最小错误率判决规则”就变为：

若：$l_{12}(x) > \theta_{12}$， 则 $\mathbf{x} \in w_1$

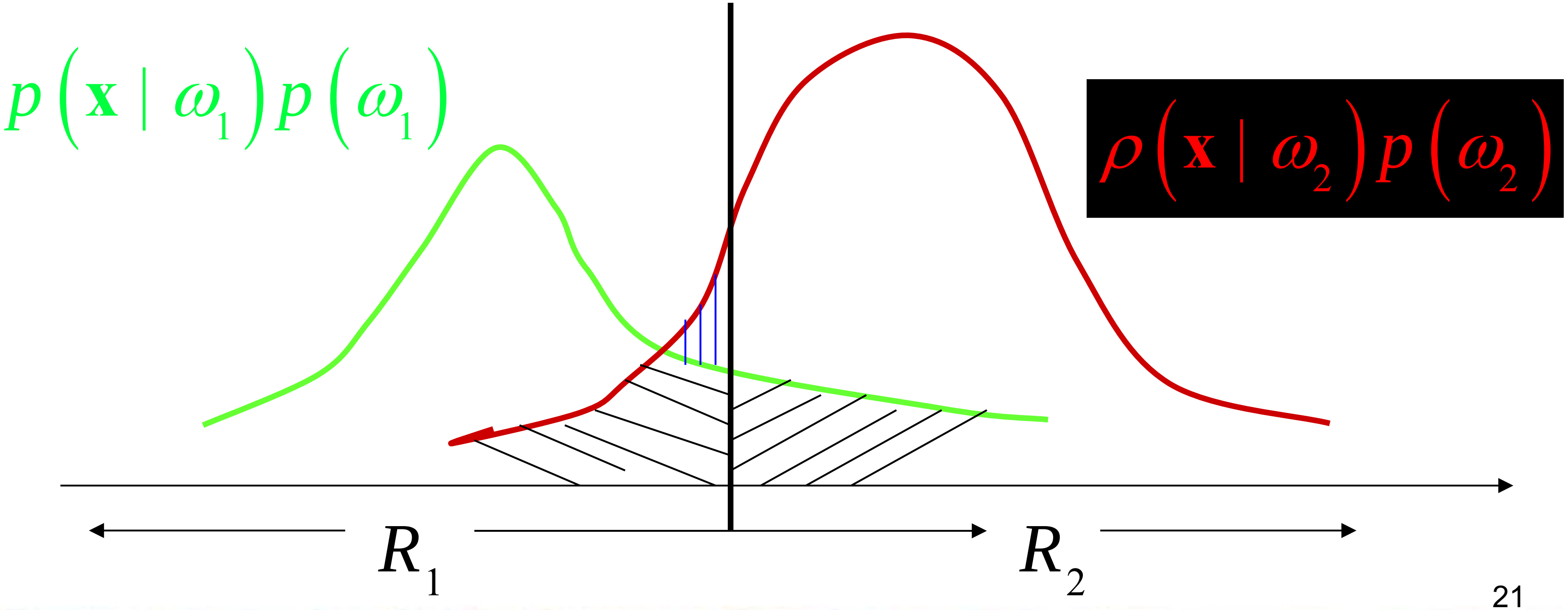
$l_{12}(x) < \theta_{12}$， 则 $\mathbf{x} \in w_2$

$l_{12}(x) = \theta_{12}$， $\mathbf{x} \in w_1$ 或 $\mathbf{x} \in w_2$

两类问题决策面方程

- 二类问题：判别函数

$$g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) \begin{cases} > 0, & \mathbf{x} \in \omega_1 \\ < 0, & \mathbf{x} \in \omega_2 \end{cases}$$



两类问题决策面方程

- 决策面方程 $g(\mathbf{x})=0$
- 也可以表示为：

$$p(\mathbf{x}\mid w_1)\cdot p(w_1)-p(\mathbf{x}\mid w_2)\cdot p(w_2)=0$$

两类问题决策面方程

- 对例1写出其判决函数和决策面方程。
- 解：1. 对于**例1**，用判决函数：

$$g(\mathbf{x}) = p(\mathbf{x} \mid w_1) \cdot p(w_1) - p(\mathbf{x} \mid w_2) \cdot p(w_2)$$

得到对应的判决函数为：

$$g(\mathbf{x}) = 0.995\, p(\mathbf{x} \mid w_1) - 0.005\, p(\mathbf{x} \mid w_2)$$

决策面方程为：

$$0.995\, p(\mathbf{x} \mid w_1) - 0.005\, p(\mathbf{x} \mid w_2) = 0$$

最小错误率判决规则—多类情况

- 把上述两类问题导出的最小错误率判决规则一般化，推广到R类问题中，表达为若

$$p(w_i \mid \mathbf{x}) = \max_{j=1,\cdots,R} \{p(w_j \mid \mathbf{x})\}$$

则 $\mathbf{x} \in w_i$

- 等价于：若 $p(\mathbf{x} \mid w_i) \cdot p(w_i) = \max_{j=1,\cdots,R} \{p(\mathbf{x} \mid w_j) \cdot p(w_j)\}$

  则 $\mathbf{x} \in w_i$

最小错误率判决规则—多类情况

假定

$$g_i(\mathbf{x}) = p(\omega_i \mid \mathbf{x})$$

$$g_i(\mathbf{x}) = p(\mathbf{x} \mid \omega_i)p(\omega_i)$$

$$g_i(\mathbf{x}) = \log p(\mathbf{x} \mid \omega_i) + \log p(\omega_i)$$

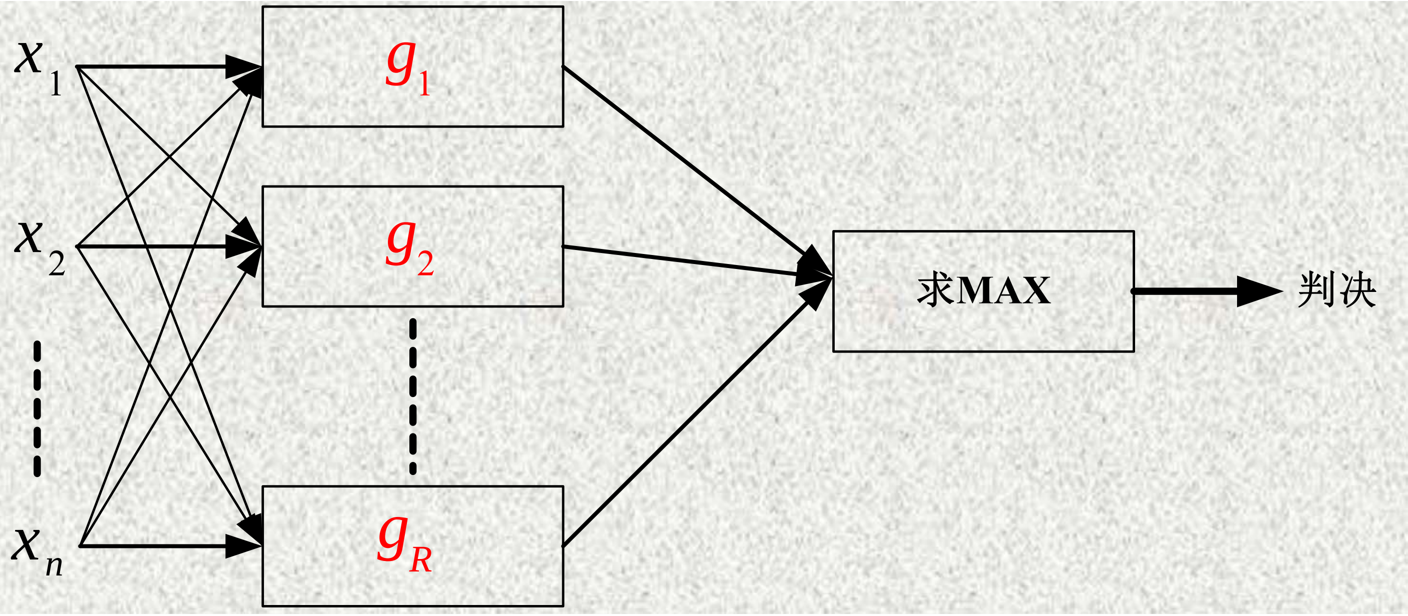
但是它们的效果是一样的，都是把特征空间分割成多个不同的决策区域。

$$g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i \Rightarrow \mathbf{x} \in \omega_i$$

最小错误率判决规则—多类情况

- 多类分类器结构

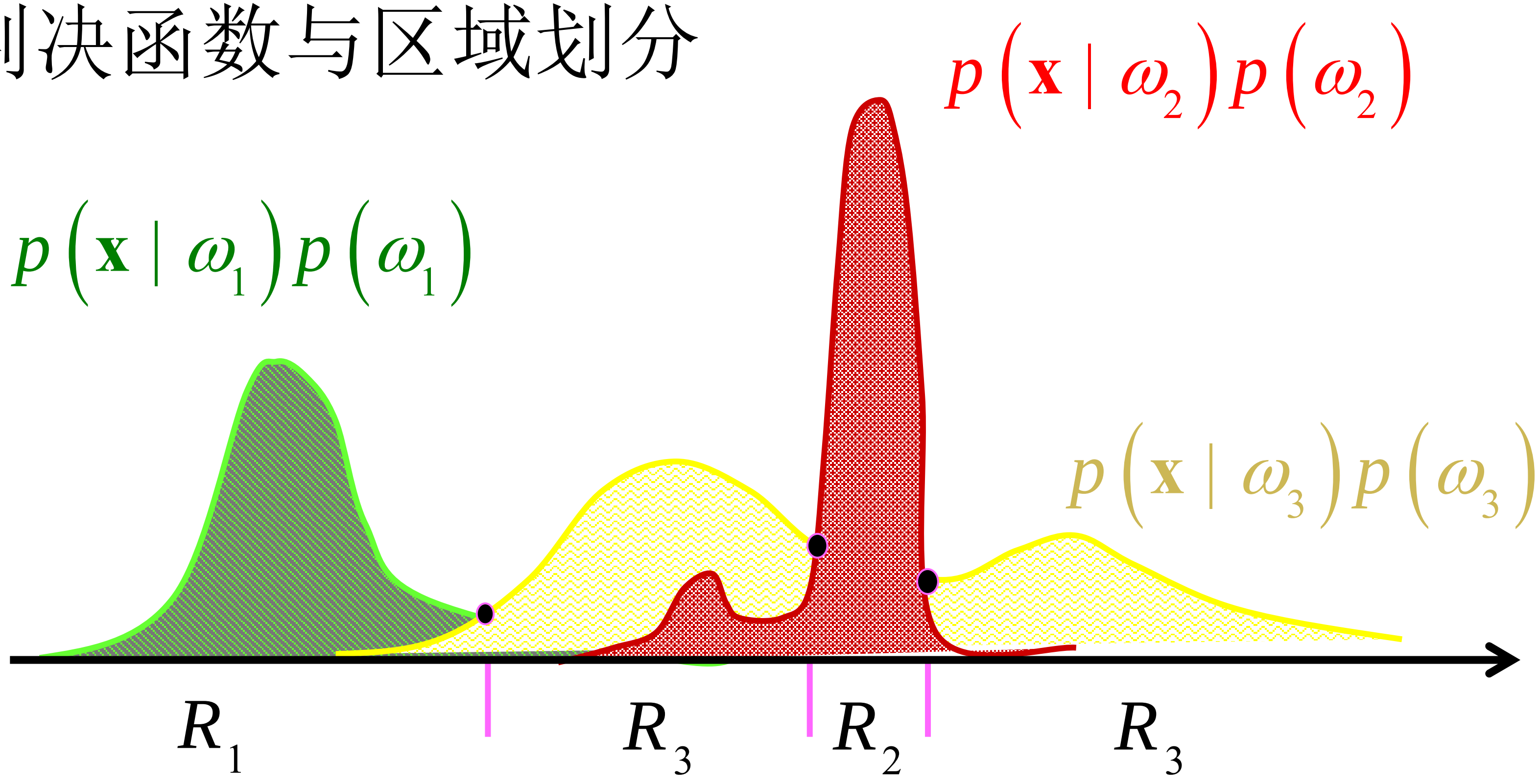
$$\mathbf{x} = (x_1, x_2, ..x_n)^T \qquad \omega_1, \omega_2, ..\omega_R$$



若有 $g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i \Rightarrow \mathbf{x} \in \omega_i$

多类问题决策面方程

- 判决函数与区域划分



决策面：$g_i(\mathbf{x}) = g_j(\mathbf{x})$ 且 R_i 与 R_j 相邻

最小错误率判决规则—多类情况

- C类情况下： $w_1, w_2, \cdots\cdots, w_R$

$$l_{ij}(\mathbf{x}) = \frac{\rho(\mathbf{x} \mid w_i)}{\rho(\mathbf{x} \mid w_j)}$$

- 由最小错误率判决规则导出结论：

若 $l_{ij}(\mathbf{x}) > \theta_{ij}$ ，则 $\mathbf{x} \in w_i$

其中 $\theta_{ij} = \dfrac{p(w_j)}{p(w_i)}$

2.2.2 最小风险判决规则

- 目标：判定过程中，考虑不同判决时产生代价不同
- 风险：对于某一样本 $\mathbf{x} \in \omega_j$ ，而采取判决 α_i ，则招致损失 $L(\alpha_i \mid \omega_j)$，简记为$L_{ji}$，称$R(\alpha_i \mid \mathbf{x})$为取行动 α_i 时的条件风险

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} L(\alpha_i \mid \omega_j) p(\omega_j \mid \mathbf{x})$$

- 总风险 $R = \int R(\alpha(\mathbf{x}) \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}$

其中$\alpha(\mathbf{x})$ 是对每一个$\mathbf{x}$ 所可能取得行动α_1 α_2 α_R中的一个。

最小风险判决规则

- 两类问题
  - 决策1： α_1 ----〉 决策为ω_1类
  - 决策2： α_2 ----〉 决策为ω_2类
  - L_{ij} 表示真类别为 ω_i，判决为ω_j所招致的损失

$$L_{ij} = L(\alpha_j \mid \omega_i)$$

则

$$R(\alpha_1 \mid \mathbf{x}) = L_{11}p(\omega_1 \mid \mathbf{x}) + L_{21}p(\omega_2 \mid \mathbf{x})$$

$$R(\alpha_2 \mid \mathbf{x}) = L_{12}p(\omega_1 \mid \mathbf{x}) + L_{22}p(\omega_2 \mid \mathbf{x})$$

最小风险判决规则

- 判断准则：

若 $R(\alpha_1 \mid \mathbf{x}) < R(\alpha_2 \mid \mathbf{x})$

则采取行动 α_1 的损失较小

最小风险判决规则

- 实施最小风险判决规则的步骤如下：

（1）在给定样本x条件下，计算各类后验概率 $p(w_j \mid \mathbf{x})$, $j=1,2,\cdots\cdots,c$ 。

（2）求各种判决的条件平均风险 $R(\alpha_i \mid \mathbf{x})$ ，

$j=1,2,\cdots\cdots,a$ ，为此，需要知道风险矩阵。

（3）比较各种判决的条件平均风险，把样本x归属于条件平均风险最小的那一种判决。

举例2

- 在例1的基础上，考虑决策风险

损失 状态 / 决策	ω_1	ω_2
α_1	**0.5**	**6**
α_2	**2**	**0.5**

试对该病人x进行分类。

解答

解：已知条件

$P(\omega_1) = 0.995$ $P(\omega_2) = 0.005$

$\rho(\mathbf{x} \mid \omega_1) = 0.01$ $\rho(\mathbf{x} \mid \omega_2) = 0.95$

$L_{11} = 0.5$ $L_{21} = 6$

$L_{12} = 2$ $L_{22} = 0.5$

后验概率 $P\left(\omega_1 \mid \mathbf{x}\right) = 0.677$ $P\left(\omega_2 \mid \mathbf{x}\right) = 0.323$

条件风险

$$R(\alpha_1 \mid \mathbf{x}) = L_{11}P(\omega_1 \mid \mathbf{x}) + L_{21}P(\omega_2 \mid \mathbf{x}) = 2.2765$$

$$R(\alpha_2 \mid \mathbf{x}) = L_{12}P(\omega_1 \mid \mathbf{x}) + L_{22}P(\omega_2 \mid \mathbf{x}) = 1.5155$$

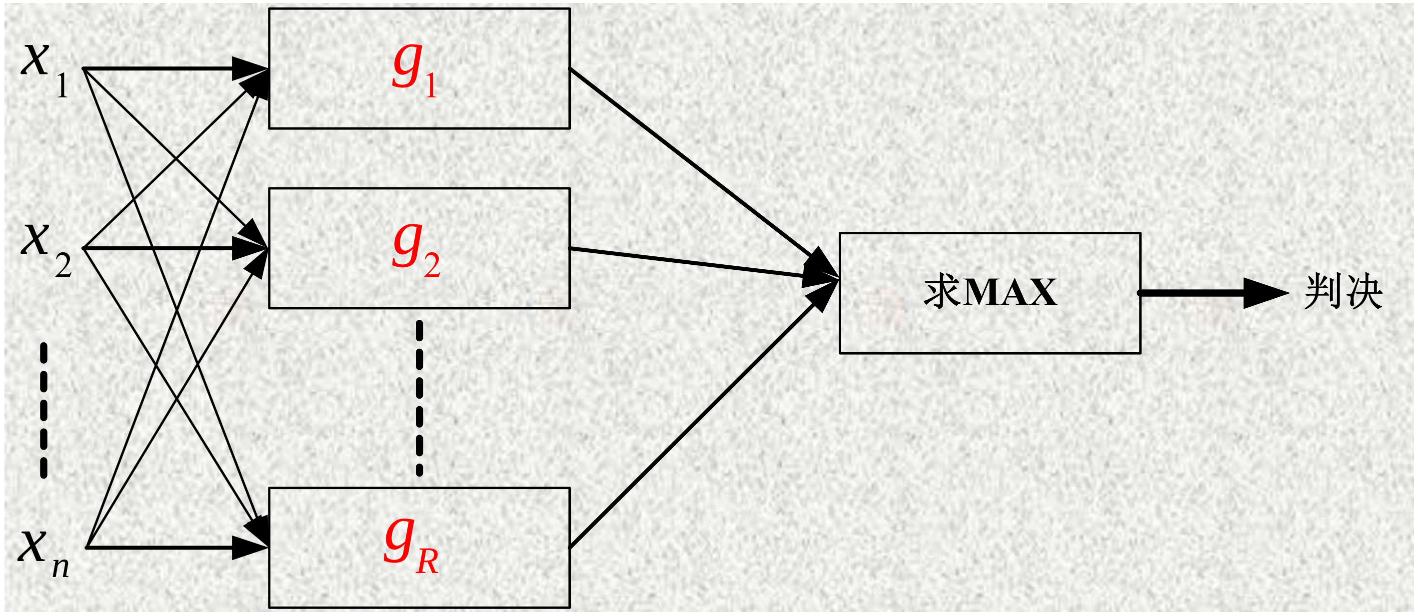
$$\boxed{R(\alpha_2 \mid \mathbf{x}) < R(\alpha_1 \mid \mathbf{x})}$$

结论：决策x为癌症病人，与例1 的结论相反

最小错误率判决规则—多类情况

- 多类分类器结构

$$\mathbf{x} = (x_1, x_2, ..x_n)^T \qquad \omega_1, \omega_2, ..\omega_R$$



若有 $g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i \Rightarrow \mathbf{x} \in \omega_i$

最小风险判决规则—决策面

- 对于例2，判决函数定义为： $g(\mathbf{x})=g_1(\mathbf{x})-g_2(\mathbf{x})$

其中：$g_1(\mathbf{x})=1-R(a_1=w_1\,|\,\mathbf{x})$，$g_2(\mathbf{x})=1-R(a_2=w_2\,|\,\mathbf{x})$，

带入上式：

$$g(\mathbf{x})=g_1(\mathbf{x})-g_2(\mathbf{x})$$
$$=R(a_2=w_2\,|\,\mathbf{x})-R(a_1=w_1\,|\,\mathbf{x})$$

$$R(\alpha_1=w_1\,|\,\mathbf{x})=\sum_{j=1}^{2}L(\alpha_1\,|\,w_j)\cdot p(w_j\,|\,\mathbf{x})=0.5*p(w_1\,|\,\mathbf{x})+6*p(w_2\,|\,\mathbf{x})$$
$$R(\alpha_2=w_2\,|\,\mathbf{x})=\sum_{j=1}^{2}L(\alpha_2\,|\,w_j)\cdot p(w_j\,|\,\mathbf{x})=2*p(w_1\,|\,\mathbf{x})+0.5*p(w_2\,|\,\mathbf{x})$$

$$g(\mathbf{x})=1.5*p(w_1\,|\,\mathbf{x})-5.5*p(w_2\,|\,\mathbf{x})$$

决策面方程为：$1.5*p(w_1\,|\,\mathbf{x})-5.5*p(w_2\,|\,\mathbf{x})=0$。

两种决策规则的关系

- 结论：在0－1损失函数情况下，最小风险判决规则退化为最小错误率判决规则。也就是说，最小错误率判决规则是最小风险判决规则的一个特例。
- 推导：
- 现假设正确判决损失为0，错误判决损失为1，且判决数目与类型数目相等。即有0－1损失函数：

$$L(\alpha_i \mid w_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

两种决策规则的关系

可令$L(\alpha_i \mid w_j) = 1 - \delta_{ij}$，其中 $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

$$R(\alpha_i \mid w_j) = \sum_{i=1}^{c} L(\alpha_i \mid w_j) \cdot P(w_j \mid \mathbf{x})$$

$$= \sum_{i=1}^{c} (1 - \delta_{ij}) \cdot P(w_j \mid \mathbf{x}) = \sum_{i=1}^{c} P(w_j \mid \mathbf{x}) - \sum_{j=1}^{c} \delta_{ij} \cdot P(w_j \mid \mathbf{x})$$

$$= 1 - P(w_i \mid \mathbf{x})$$

得到若$P(w_i \mid \mathbf{x}) = \max_{k=1,2,\cdots,c} \{P(w_k \mid \mathbf{x})\}$，则 $\mathbf{x} \in w_i$

这就是最小错误率判决规则。

2.2.3 最大似然比判决规则

- 判断准则：

若 $R(\alpha_1 \mid \mathbf{x}) < R(\alpha_2 \mid \mathbf{x})$

则采取行动 α_1 的损失较小

- 推导：

$$L_{11}p(\omega_1 \mid \mathbf{x}) + L_{21}p(\omega_2 \mid \mathbf{x}) < L_{12}p(\omega_1 \mid \mathbf{x}) + L_{22}p(\omega_2 \mid \mathbf{x})$$

即

$$(L_{21} - L_{22})p(\omega_2 \mid \mathbf{x}) < (L_{12} - L_{11})p(\omega_1 \mid \mathbf{x}) \Rightarrow \mathbf{x} \in \omega_1$$

最大似然比判决规则

- 用先验概率与条件概率代替后验概率

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

$$\Rightarrow \quad \mathbf{x}\in\omega_1$$

- 似然比

$$l(\mathbf{x})=\frac{p(\mathbf{x}\mid\omega_1)}{p(\mathbf{x}\mid\omega_2)}>\frac{(L_{21}-L_{22})p(\omega_2)}{(L_{12}-L_{11})p(\omega_1)}=\theta_{12}$$

与X无关
的常数

- 结论：若似然比大于某一与x无关的阈值，则决策为ω_1，被称为最小风险的Bayes决策。

最大似然比判决规则

结论：最小风险的Bayes判决准则

似然比 $l(\mathbf{x})\begin{cases} > \dfrac{(L_{21}-L_{22})p(\omega_2)}{(L_{12}-L_{11})p(\omega_1)} \\ < \dfrac{(L_{21}-L_{22})p(\omega_2)}{(L_{12}-L_{11})p(\omega_1)} \end{cases} \Rightarrow \mathbf{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases}$

其中：$\begin{cases} L_{11} \text{ 是 } \mathbf{x}\in\omega_1\text{，且被分入域}R_1\text{中的“损失”；} \\ L_{21} \text{ 是 } \mathbf{x}\in\omega_2\text{，且被分入域}R_1\text{中的“损失”；} \\ L_{12} \text{ 是 } \mathbf{x}\in\omega_1\text{，且被分入域}R_2\text{中的“损失”；} \\ L_{22} \text{ 是 } \mathbf{x}\in\omega_2\text{，且被分入域}R_2\text{中的“损失”；} \end{cases}$

最大似然比判决规则

例2用似然比的方法求解

先验概率： p(w1)=0.995， p(w2)=0.005

类条件概率密度：p (**x**|w1)=0.01, p(**x**|w2)=0.95

风险：L_{11}=0.5, L_{12}=6, L_{21}=2, L_{22}=0.5

决策过程:

$$l_{12}(\mathbf{x}) = \frac{\rho(\mathbf{x} \mid w_1)}{\rho(\mathbf{x} \mid w_2)} = \frac{0.01}{0.95} = 0.0105$$

$$\theta_{12} = \frac{L(\alpha_1 \mid w_2) - L(\alpha_2 \mid w_2)}{L(\alpha_2 \mid w_1) - L(\alpha_1 \mid w_1)} \cdot \frac{p(w_2)}{p(w_1)} = \frac{6-0.5}{2-0.5} \times \frac{0.005}{0.995} = 0.0184$$

$l_{12}(\mathbf{x}) < \theta_{12}$ 根据似然比的判定规则可知甲$\in w_1$与例2结果保持一致。

2.2.4 Neyman-Pearson判决规则

- 问题
  - 通常情况下，无法确定损失$L_{11}, L_{12}, L_{21}, L_{22}$ 。
  - 先验概率未知
- 基本思想
  - 限定或者约束某一个错误概率，与此同时通过一些计算，使得另一个错误概率最小
- 实现
  - 设计一个辅助函数

$$\gamma = \varepsilon_1 + \mu\varepsilon_2 \quad \varepsilon_1 = \int_{R_2} p(\mathbf{x} \mid \omega_1) d\mathbf{x} \quad \varepsilon_2 = \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x}$$

Neyman-Pearson判决规则

- 由于

$$\int_{R_2} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} + \int_{R_1} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} = 1$$

所以

$$\begin{aligned}
\gamma &= \int_{R_2} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} + \mu \int_{R_1} \rho(\mathbf{x} \mid \omega_2) d\mathbf{x} \\
&= \int_{R_2} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} + \int_{R_1} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} + \mu \int_{R_1} \rho(\mathbf{x} \mid \omega_2) d\mathbf{x} - \int_{R_1} \rho(\mathbf{x} \mid \omega_1) d\mathbf{x} \\
&= 1 + \int_{R_1} \left[\mu \rho(\mathbf{x} \mid \omega_2) - \rho(\mathbf{x} \mid \omega_1) \right] d\mathbf{x}
\end{aligned}$$

$\min(\gamma)$ $\Rightarrow$ $\mu \rho(\mathbf{x} \mid \omega_2) < \rho(\mathbf{x} \mid \omega_1)$ 在R_1中

结论

$$\frac{\rho(\mathbf{x} \mid \omega_1)}{\rho(\mathbf{x} \mid \omega_2)} > \mu \Rightarrow \mathbf{x} \in \omega_1$$

Neyman-Pearson判决规则

- 同样地，在R_2中

$$\gamma = 1 + \int_{R_2} \left[\rho(\mathbf{x} \mid \omega_1) - \mu \rho(\mathbf{x} \mid \omega_2) \right] d\mathbf{x}$$

$$\min(\gamma) \Rightarrow \mu \rho(\mathbf{x} \mid \omega_2) > \rho(\mathbf{x} \mid \omega_1)$$

$$\frac{\rho(\mathbf{x} \mid \omega_1)}{\rho(\mathbf{x} \mid \omega_2)} < \mu \Rightarrow \mathbf{x} \in \omega_2$$

Neyman-Pearson判决规则

- 结论

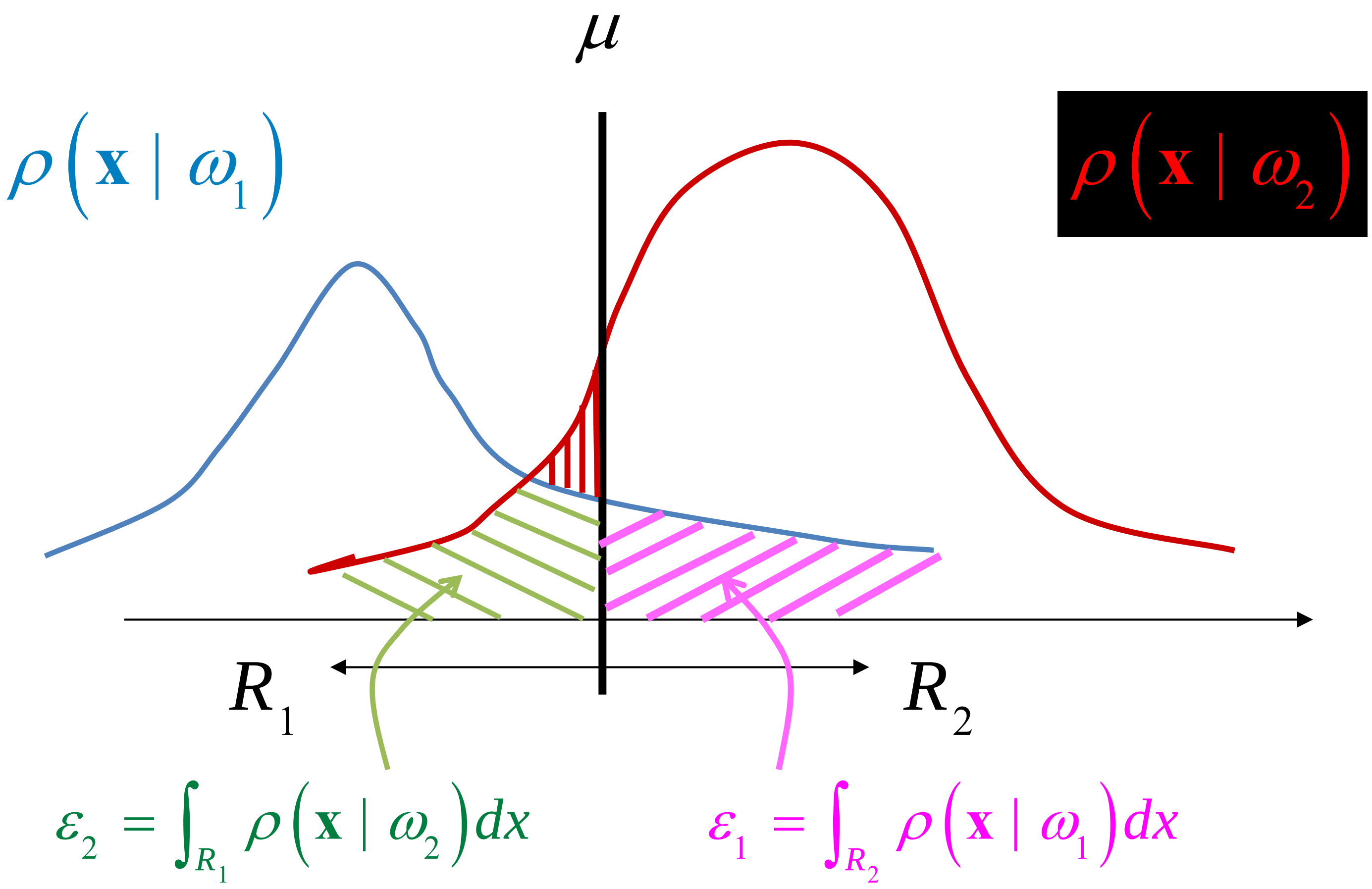
  Neyman-Pearson决策的任务就是寻找阈值常数μ，当 $\dfrac{\rho(\mathbf{x}\mid\omega_1)}{\rho(\mathbf{x}\mid\omega_2)}=\mu$ 时

  μ 就可以划分子空间R_1和R_2。而

$$\varepsilon_2=\int_{-\infty}^{\mu}\rho(\mathbf{x}\mid\omega_2)d\mathbf{x}$$

ε_2 已知 $\rightarrow$ μ $\rightarrow$ 子空间R_1和R_2

Neyman-Pearson判决规则



举例

- 两类的概率密度函数是正态的，两类的均值向量分别为$\mu_1=(-1,0)^T$和$\mu_2=(0,1)^T$，协方差矩阵相等且为单位矩阵。给定$\varepsilon_2=0.046$，试确定N-P判决门限t。

解答

- 根据给定的条件，很容易写出两类的类概率密度函数，即：$\rho(\mathbf{x}\,|\,w_1)\sim N(\mu_1,\Sigma)$ 和 $\rho(\mathbf{x}\,|\,w_2)\sim N(\mu_2,\Sigma)$。$\Sigma = E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T]$。

而 $\mu_1=(-1,0)^T$，$\mu_2=(1,0)^T$，$\Sigma = I$，得到：

$$\rho(\mathbf{x}\,|\,w_1)=\frac{1}{(2\pi)^{d/2}\,|\,\Sigma\,|^{1/2}}\exp[-\frac{1}{2}(\mathbf{x}-\mu_1)^T\cdot\Sigma^{-1}(\mathbf{x}-\mu_1)]$$

$$=\frac{1}{2\pi}\exp(-\frac{1}{2}[(x_1+1)^2+x_2{}^2])=\frac{1}{2\pi}\exp(-\frac{1}{2}(x_1{}^2+2x_1+1+x_2{}^2))$$

解答

$$\rho(\mathbf{x}\,|\,w2)=\frac{1}{2\pi}\exp(-\frac{1}{2}[(x_1-1)^2+x_2{}^2])=\frac{1}{2\pi}\exp(-\frac{1}{2}[(x_1{}^2-2x_1+x_2{}^2))$$

$$\frac{\rho(\mathbf{x}\,|\,w_1)}{\rho(\mathbf{x}\,|\,w_2)}=\exp(-2x_1)$$

- 故：$\mu(\mathbf{x})=\exp(-2x_1)$，$\mu$只是$x_1$的函数，与$x_2$无关。有$x_1=-\frac{1}{2}\ln\mu$。又$\rho(\mathbf{x}\,|\,w_2)$的边缘密度为$\rho(x_1\,|\,w_2)$：

解答

$$\rho(x_1 \mid w_2) = \int_{-\infty}^{\infty} \rho(\mathbf{x} \mid w_2) dx_2$$

$$= \int \frac{1}{2\pi} \exp[-\frac{1}{2}(x_1^2 - 2x_1 + 1 + x_2^2)] dx_2$$

$$= \frac{1}{2\pi} \exp[-\frac{1}{2}(x_1^2 - 2x_1 + 1)] \cdot \int_{-\infty}^{\infty} \exp(-\frac{1}{2} x_2^2) dx_2$$

$$= \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(x_1 - 1)^2]$$

对于给定的正数 ε_2，可由下式计算：

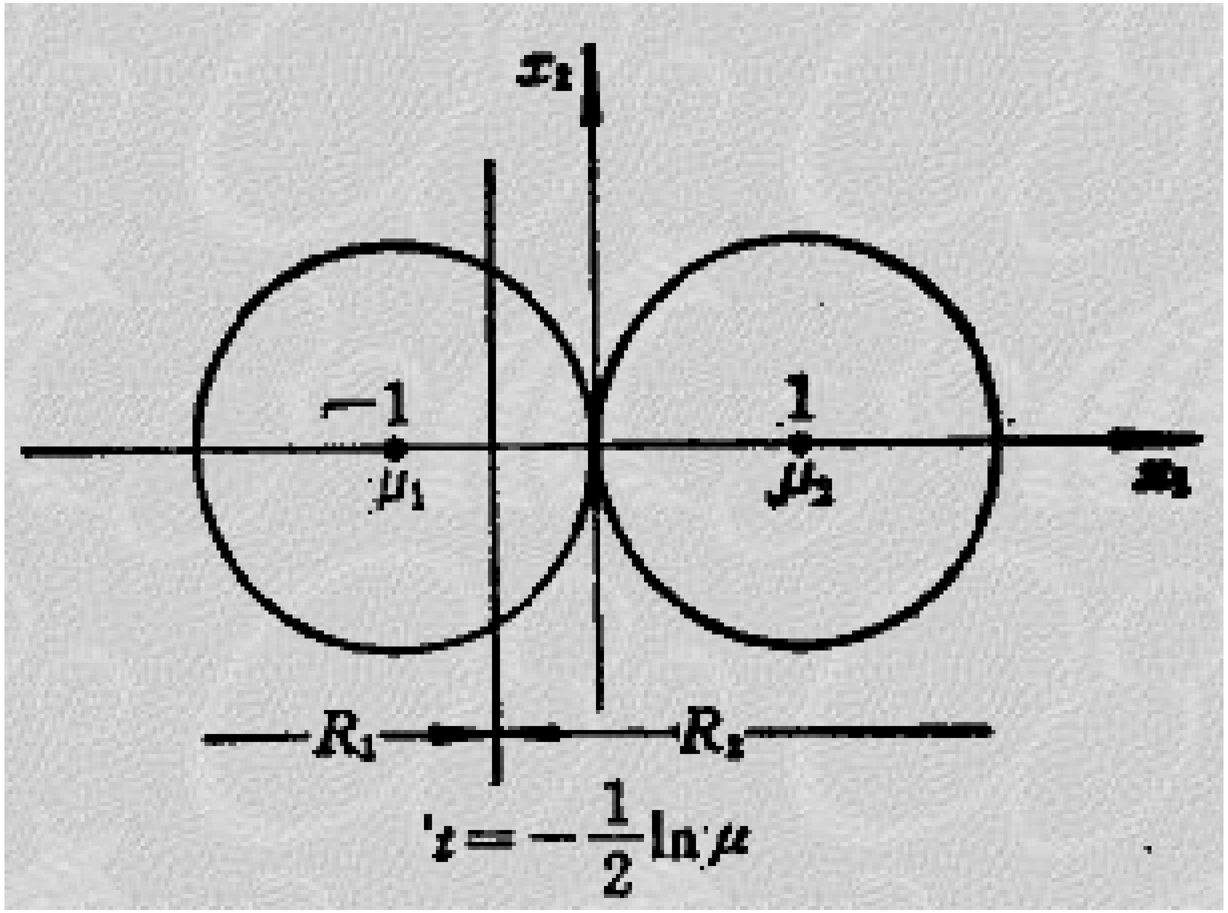
$$\varepsilon_2 = \int_{-\infty}^{-\frac{1}{2}\ln\mu} \frac{1}{\sqrt{2\pi}} \exp[-\frac{(x_1 - 1)^2}{2}] dx_1 = \int_{-\infty}^{-\frac{1}{2}\ln\mu - 1} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) dy$$

解答

- 显然，y是服从标准正态分布的随机变量，令 $y_1=-\frac{1}{2}\ln\mu-1$ ，则：$\varepsilon_2=\Phi(y_1)$
- ε_2与y_1具有一一对应的关系，有表可查。当

$\varepsilon_2=0.046$ 时，$y_1=-1.693$，$\mu=4$，$x_1=-0.693$， 因此判决门限 $t=x_1=-0.693$， 如下图所示。分区界线是 $x_1=-0.693$ 的一条直线，对于样本 $\mathbf{x}=(x_1,x_2)^T$ 的分类判决，只需考察特征x_1 ，判决规则为，若：$x_1<-0.693$， 则 $\mathbf{x}\in w_1$。否则，x属于w_2。

解答



选择门限图

2.2.5 最小最大判决规则

- 在先验概率已知的情况下
  - 利用似然比和各个先验概率比值确定的阈值，可以确定某一样本的正确分类
- 在先验概率变化的情况下
  - 能否找到一种决策，使得Bayes风险仍最小？
- 基本思想
  - 在训练过程中使平均风险达到最大值，恰好在分类判决中使最大可能的平均风险达到最小值。
- 两类问题

最小最大判决规则

- 总平均风险 $R = \int R(\alpha(\mathbf{x}) \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}$

由于， $p(\omega_2) = 1 - p(\omega_1)$

$$\int_{R_1} p(\mathbf{x} \mid \omega_i) d\mathbf{x} = 1 - \int_{R_2} p(\mathbf{x} \mid \omega_i) d\mathbf{x}$$

则

$$\begin{aligned} R = & \int_{R_1} L_{11} p(\mathbf{x} \mid \omega_1) p(\omega_1) d\mathbf{x} + \int_{R_1} L_{21} p(\mathbf{x} \mid \omega_2) p(\omega_2) d\mathbf{x} \\ & + \int_{R_2} L_{12} p(\mathbf{x} \mid \omega_1) p(\omega_1) d\mathbf{x} + \int_{R_2} L_{22} p(\mathbf{x} \mid \omega_2) p(\omega_2) d\mathbf{x} \end{aligned}$$

最小最大判决规则

$$R = L_{22} + (L_{21} - L_{22})\boxed{\int_{R_1} p(\mathbf{x} \mid \omega_2)d\mathbf{x}} + p(\omega_1)[(L_{11} - L_{22})$$

$$+(L_{12} - L_{11})\underbrace{\boxed{\int_{R_2} p(\mathbf{x} \mid \omega_1)d\mathbf{x}}}_{\varepsilon_1} - (L_{21} - L_{22})\underbrace{\boxed{\int_{R_1} p(\mathbf{x} \mid \omega_2)d\mathbf{x}]}}_{\varepsilon_2}$$

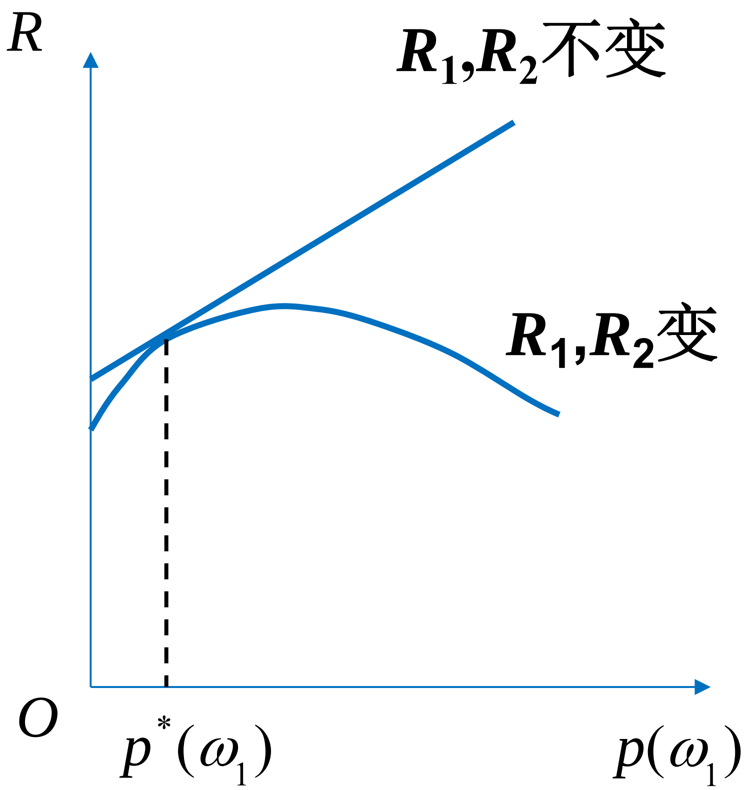
$$R = \underbrace{L_{22} + (L_{21} - L_{22})\varepsilon_2} + p(\omega_1)\left[\underbrace{(L_{11} - L_{22}) + (L_{12} - L_{11})\varepsilon_1 - (L_{21} - L_{22})\varepsilon_2}\right]$$

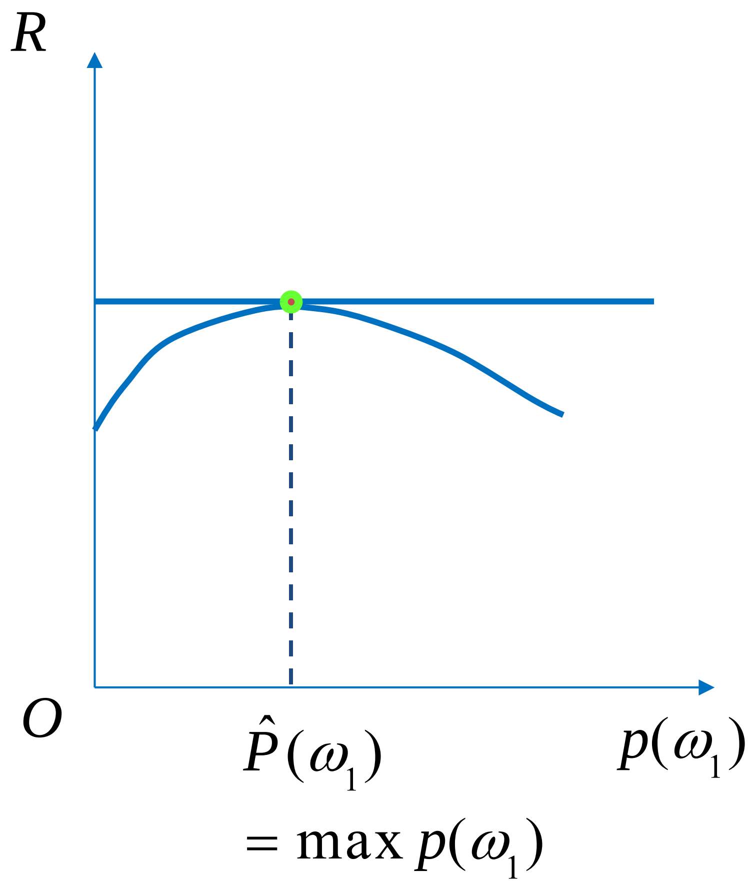
$$= a + bp(\omega_1)$$

- 结论：一旦域R_1,R_2确定了，总风险R就是$p(\omega_1)$的一个线性函数。

最小最大判决规则

- Bayes风险R与$p(\omega_1)$的关系曲线





最小最大判决规则

- 求总风险的极大值 $\dfrac{dR}{dp(\omega_1)} = 0$

$$\varepsilon_1 = \int_{R_2} p\left(X \mid \omega_1\right) dx$$

$$(L_{21} - L_{22}) + (L_{12} - L_{11})\int_{R_2} p(X \mid \omega_1)dx - (L_{21} - L_{22})\int_{R_1} p(X \mid \omega_2)dx = 0$$

或 $(L_{11} - L_{22}) + (L_{12} - L_{11})\varepsilon_1 - (L_{21} - L_{22})\varepsilon_2 = 0$

$$\varepsilon_2 = \int_{R_1} p\left(X \mid \omega_2\right) dx$$

如果对于特定的损失 $L_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$

则风险极大的条件是 $\varepsilon_1 = \varepsilon_2$ 即所选定的边界来自两个类别的错误概率相等。

2.3 正态分布时的统计决策

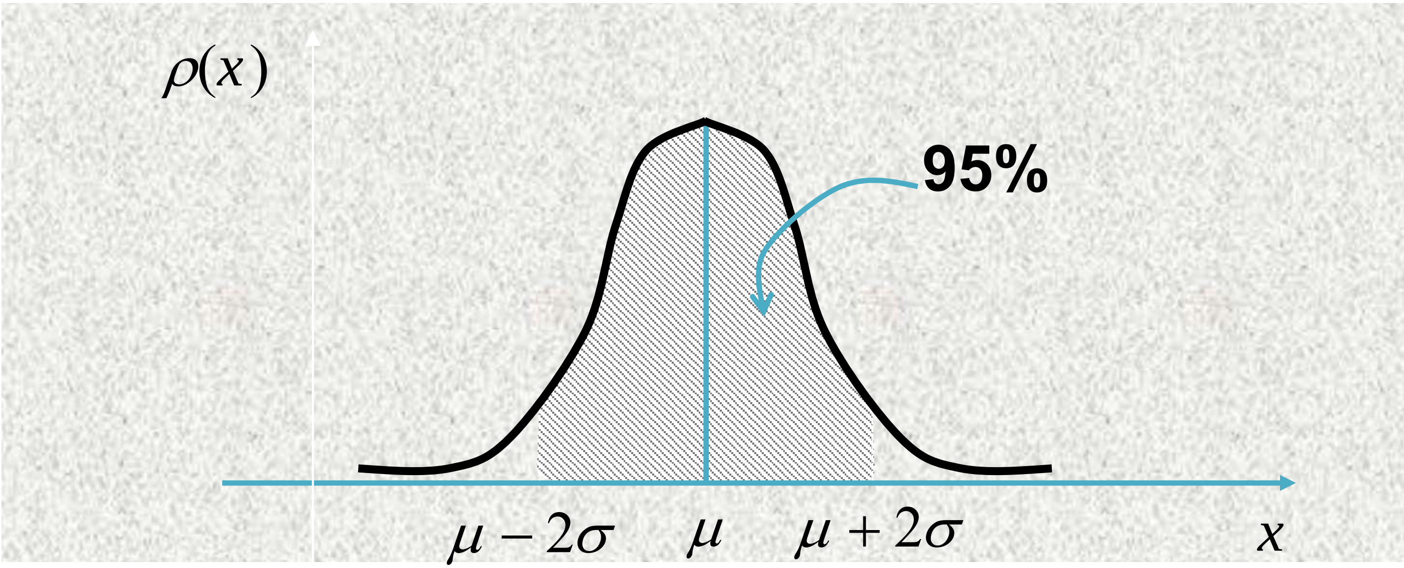
- 正态分布概率密度函数（**PDF**）的定义及性质
- 多元正态分布的性质
- 正态分布中的**Bayes**分类方法

2.3.1 正态分布PDF的定义及性质

- 正态分布(单变量)

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

$$\rho(x)\sim N(\mu,\sigma^2)$$



单变量正态分布

- $\rho(x)$ 具有以下性质：

$$\begin{cases} \rho(x) \geq 0, (-\infty < x < \infty) \\ \int_{-\infty}^{\infty} \rho(x) dx = 1 \end{cases}$$

- 从$\rho(x)$的图形上可以看出，只要有两个参数就可以完全确定其曲线。为了简单，常记$\rho(x)$为$N(\mu,\sigma^2)$。

正态密度函数

- 一维：

$$\rho(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\gamma^2}{2}\right]\qquad \gamma=\left(\frac{x-\mu}{\sigma}\right)$$

散布程度归一化距离

- d维：

$$\rho(\mathbf{x})=\frac{1}{(2\pi)^{\frac{d}{2}}\,|\,\Sigma\,|^{\frac{1}{2}}}\exp\left[-\frac{\gamma^2}{2}\right]$$

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

马氏距离

$$\rho\left(X\right)\sim N(\mu,\Sigma)$$

2.3.2 多元正态分布的性质

（1）参数μ和Σ对分布的决定性:

$\rho(x)$可由μ、Σ完全确定

（2）等密度点的轨迹为一超椭球面:

由$\rho(x)$的定义公式可知，当右边指数项为常数时，密度$\rho(x)$的值不变，所以等密度点满足:

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = constant$$

多元正态分布的性质

（3）不相关性等价于独立性：

对服从正态分布的两个分量互不相关，则它们之间一定独立。

（4）边缘分布与条件分布的等价性：

不难证明正态随机向量的边缘分布与条件分布仍服从正态分布。

（5）线性变换的正态性：

对于多元随机向量的线性变换，仍为多元正态分布的随机向量。

多元正态分布的性质

（6）线性组合的正态性

若**x**为多元正态随机向量，则线性组合 $y = a^T x$ 是一维的正态随机变量：

$$\rho(y) \sim N(a^T \mu, a^T \sum a)$$

其中，**a**与**x**同维。

2.3.3正态分布中的Bayes分类方法

- 符合Bayes判决

$$g_i(\mathbf{x}) = \rho(\mathbf{x} \mid \omega_i)p(\omega_i)$$

- 若条件概率密度为

$$\rho(\mathbf{x} \mid \omega_i) \sim N(\mu_i, \Sigma_i)$$

则

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

正态分布中的Bayes分类方法

情况分类

各特征分量统计独立，且具有相同的方差

协方差矩阵相同

协方差矩阵互不相同

情况一

- 各特征分量统计独立，且具有相同的方差

$$\Sigma_i = \sigma^2 I = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

则 $|\Sigma_i| = \sigma^2 \qquad \Sigma_i^{-1} = \dfrac{I}{\sigma^2}$

特点: 抽样落在同样大小的超球内。第i类抽样，以 μ_i 为中心。

情况一

- 判决函数

$$g_i(\mathbf{x}) = -\frac{(\mathbf{x}-\mu_i)^T(\mathbf{x}-\mu_i)}{2\sigma^2} + \log p(\omega_i)$$

若 $P(\omega_i)$ 相同，则

$$g_i(\mathbf{x}) = -\frac{\boxed{(\mathbf{x}-\mu_i)^T(\mathbf{x}-\mu_i)}}{2\sigma^2}$$

最小距离分类器

欧氏距离

展开

情况一

- 判决函数

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

$$g_i(\mathbf{x}) = W_i^T X + W_{i0}$$

其中，$W_i = \frac{1}{\sigma^2}\mu_i \qquad W_{i0} = -\frac{1}{2\sigma^2}\mu_i^T\mu_i + \log p(\omega_i)$

说明：$\mathbf{x}^T\mathbf{x}$ 对于所有的i均相同。

情况一

- 决策面

若类 ω_i 与类 ω_j 相邻，则决策面由

$$g_i\left(\mathbf{x}\right) = g_j\left(\mathbf{x}\right)$$

确定。

$$g_i\left(\mathbf{x}\right) = -\frac{1}{2\sigma^2}(-2\mu_i^T\mathbf{x} + \mu_i^T\mu_i) + \log p(\omega_i)$$

$$g_j\left(\mathbf{x}\right) = -\frac{1}{2\sigma^2}(-2\mu_j^T\mathbf{x} + \mu_j^T\mu_j) + \log p(\omega_j)$$

下页

情况一

- 决策面 $W^T\mathbf{x} - W^T\mathbf{x}_0 = 0$

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

$$= \frac{1}{\sigma^2}(\mu_i^T - \mu_j^T)\mathbf{x} - \frac{1}{\sigma^2}\left[\frac{1}{2}(\mu_i^T\mu_i - \mu_j^T\mu_j) - \sigma^2\log\frac{p(\omega_i)}{p(\omega_j)}\right]$$

$$= (\mu_i^T - \mu_j^T)\mathbf{x} \qquad (\mu_i^T - \mu_j^T) = (\mu_i - \mu_j)^T \qquad (\mu_i^T\mu_i - \mu_j^T\mu_j) = (\mu_i - \mu_j)^T(\mu_i + \mu_j)$$

$$-(\mu_i - \mu_j)^T\left[\frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{(\mu_i - \mu_j)^T(\mu_i - \mu_j)}(\mu_i - \mu_j)\log\frac{p(\omega_i)}{p(\omega_j)}\right]$$

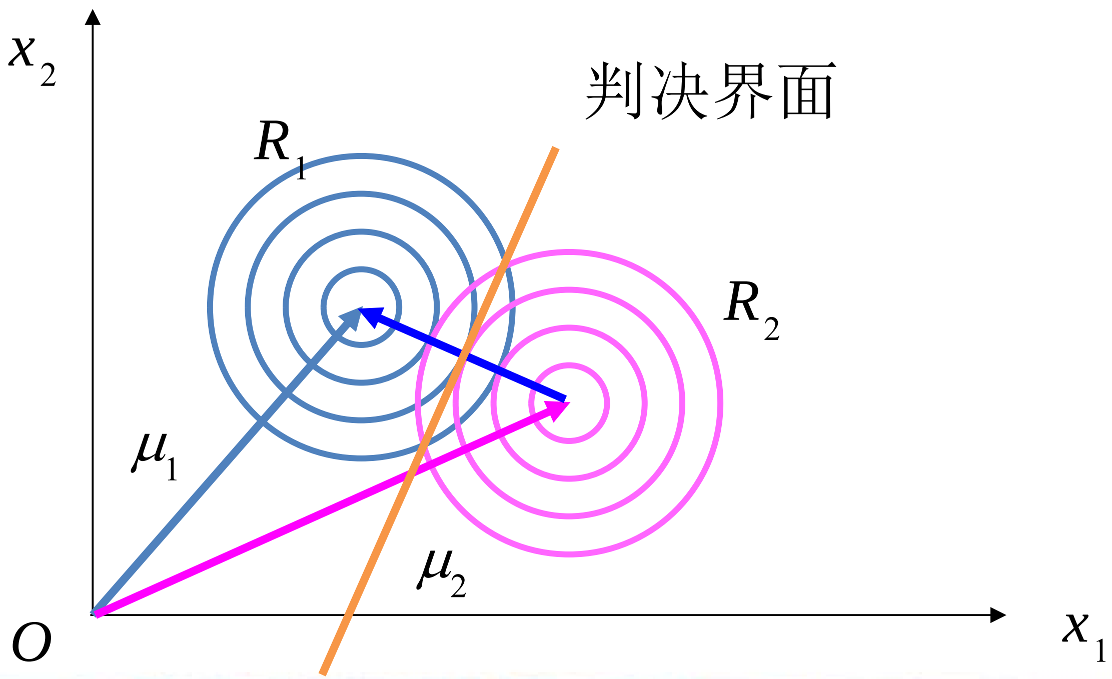
$$= W^T\mathbf{x} - W^T\mathbf{x}_0$$

$$W = (\mu_i - \mu_j) \qquad \underset{\mathbf{x}_0}{||}$$

情况一

- 结论 $W^T\mathbf{x} - W^T\mathbf{x}_0 = 0$

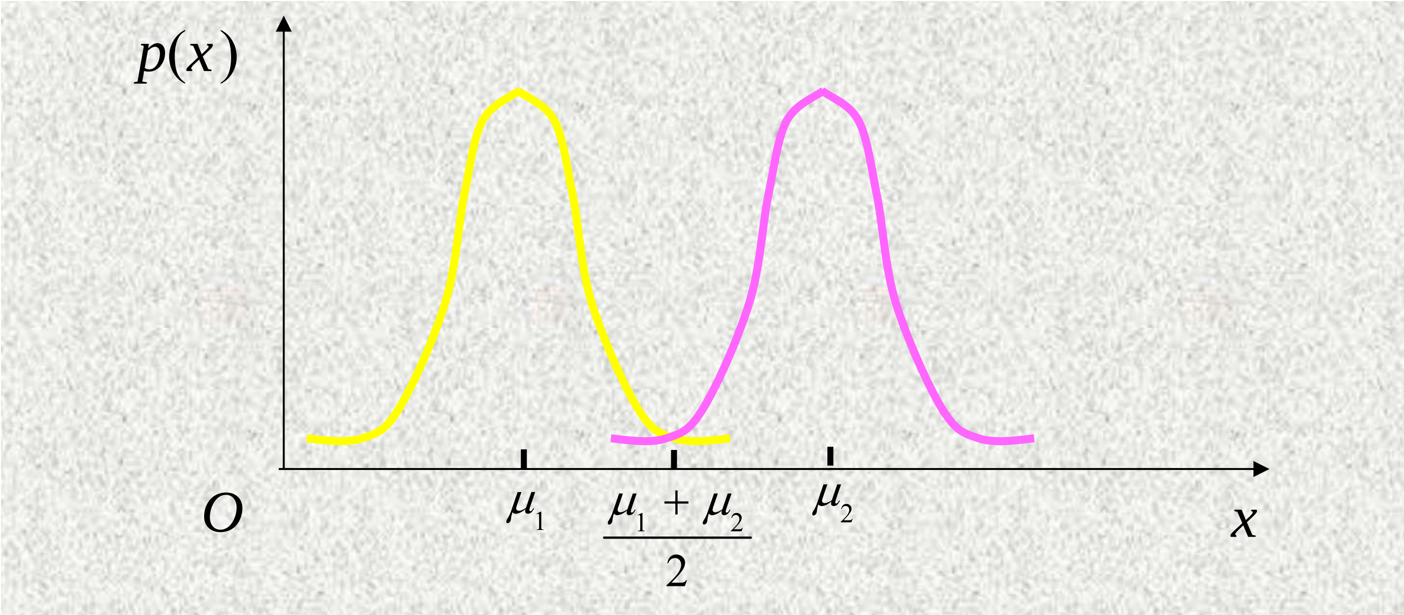
  (1) 满足此方程的超平面通过$\mathbf{x}_0$平行于W^T。由于$W = (\mu_i - \mu_j)$，故超平面垂直于均值之间的连线。



情况一

- 结论

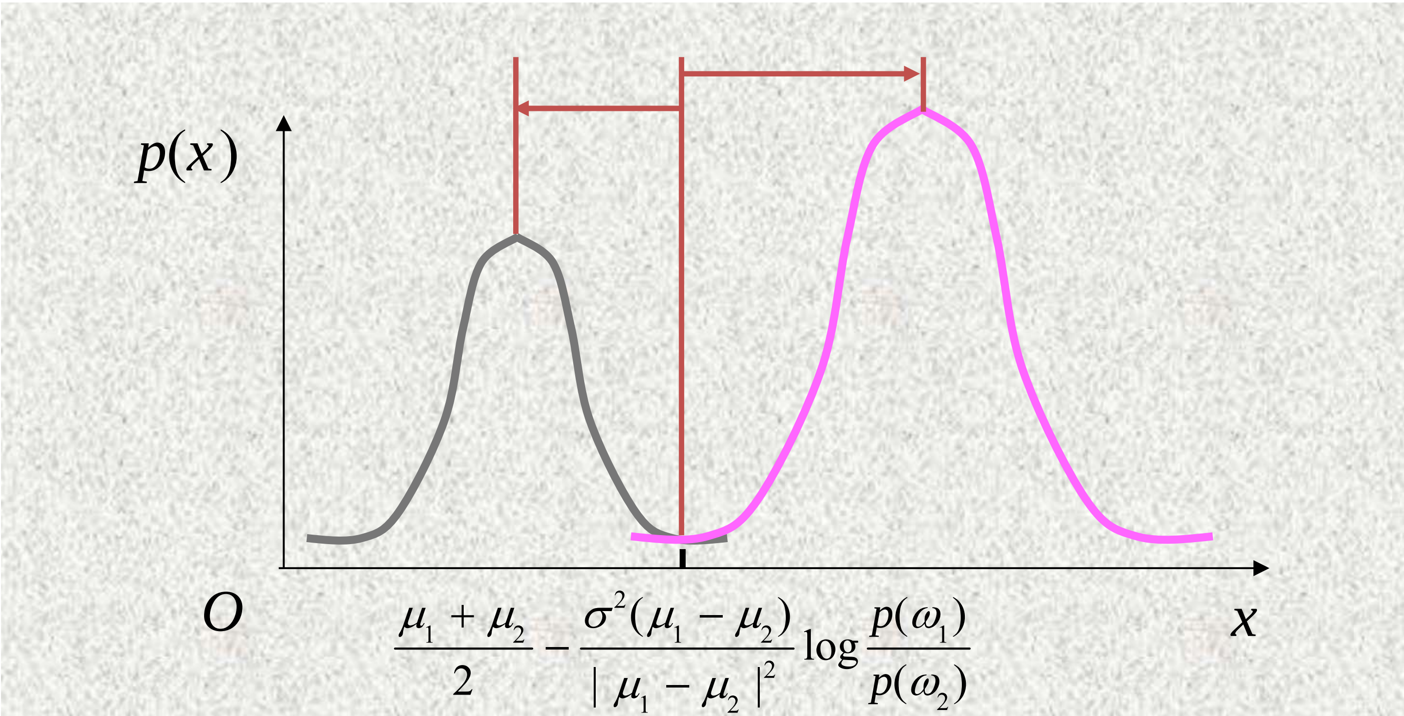
(2) 如果 $p(\omega_i) = p(\omega_j)$ ，则$\mathbf{x}_0$将通过均值连线的中点。



情况一

- 结论

(3) 如果 $p(\omega_i) \neq p(\omega_j)$ ，则$\mathbf{x}_0$离开先验概率较大的均值。



情况一

- 结论

(4) 如果 $\sigma^2 \ll \|\mu_i - \mu_j\|$，则决策面的位置受先验概率的影响较小。

即 $W^T\mathbf{x} - W^T\mathbf{x}_0 = 0$

$$\mathbf{x}_0 = \frac{\mu_i + \mu_j}{2} - \underbrace{\boxed{\frac{\sigma^2(\mu_i - \mu_j)}{\|\mu_i - \mu_j\|^2}\log\frac{p(\omega_1)}{p(\omega_2)}}}_{\approx\, 0} \approx \frac{\mu_i + \mu_j}{2}$$

情况二

- 协方差矩阵相同 $\Sigma_i = \Sigma \quad \rho \neq 0 \quad \sigma_1^2 \neq \sigma_2^2$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

与类别i无关

$$g_i(\mathbf{x}) = \boxed{-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_i|}$$

$$-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i) + \log p(\omega_i)$$

$$= -\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma^{-1}(\mathbf{x}-\mu_i) + \log p(\omega_i)$$

展开

特点：抽样落在超椭球簇内。第i类的簇以 μ_i为中心。

情况二

- 判决函数

$$g_i(\mathbf{x}) = W_i^T \mathbf{x} + W_{i0}$$

$$W_i = \Sigma^{-1}\mu_i \qquad W_{i0} = -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \log p(\omega_i)$$

- 判决面

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

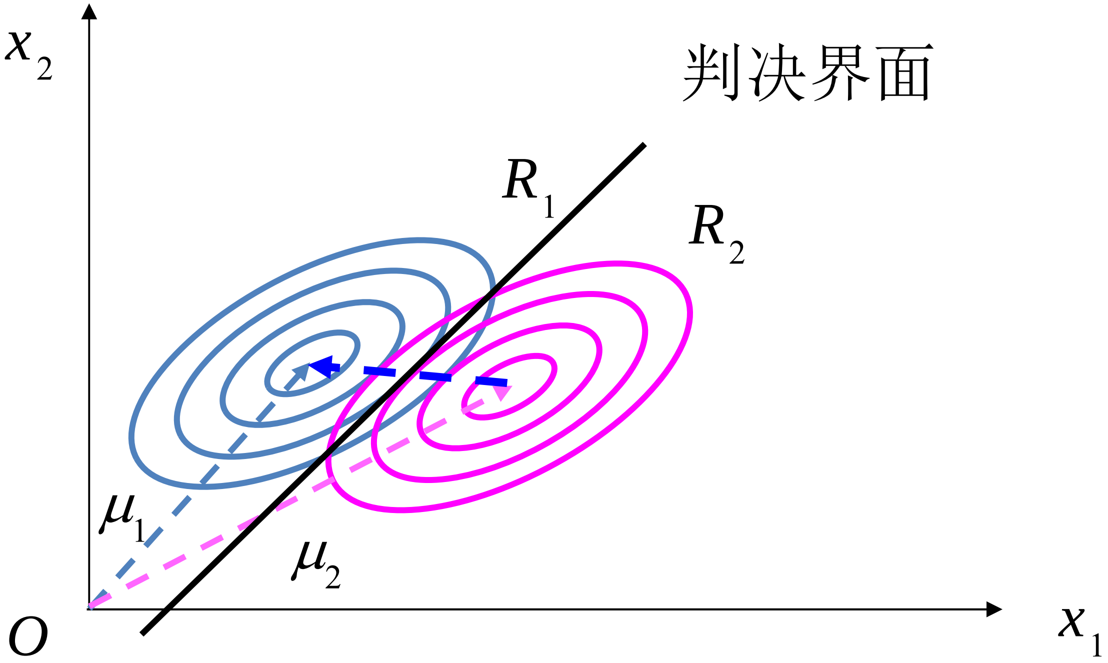
$$W = \Sigma^{-1}(\mu_i - \mu_j)$$

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

情况二

- 结论

  (1) 由于$\Sigma^{-1}(\mu_i-\mu_j)$通常不与$(\mu_i-\mu_j)$方向一致，故分割R_1与R_2的超平面一般不垂直于均值的连线。



情况二

- 结论

(2)若先验概率相等，则此超平面与均值连线相交在均值连线的中点，即

$$\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j)$$

(3)若先验概率不相等，则判断界面就是离开先验概率较大的那个均值。

情况三

- 协方差矩阵互不相同
  - 判别函数不再是线性的，而是二次型的。

$$g_i(\mathbf{x}) = \mathbf{x}^T W_i \mathbf{x} + W_{i1}^T \mathbf{x} + W_{i0}$$

$$W_i = -\frac{1}{2}\Sigma_i^{-1}$$

$$W_{i1} = \Sigma_i^{-1}\mu_i$$

$$W_{i0} = -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i - \frac{1}{2}\log\left|\Sigma_i\right| + \log p(\omega_i)$$

情况三

- 决策面
  - 超二次曲面对：超平面对、超球对、超椭球对、超抛物面对、超双曲面对
  - 二维情况下，假设分量x,y是类条件独立的($\rho=0$)，故协方差矩阵是对角形的，从而不同的决策面与各自的方差有关。

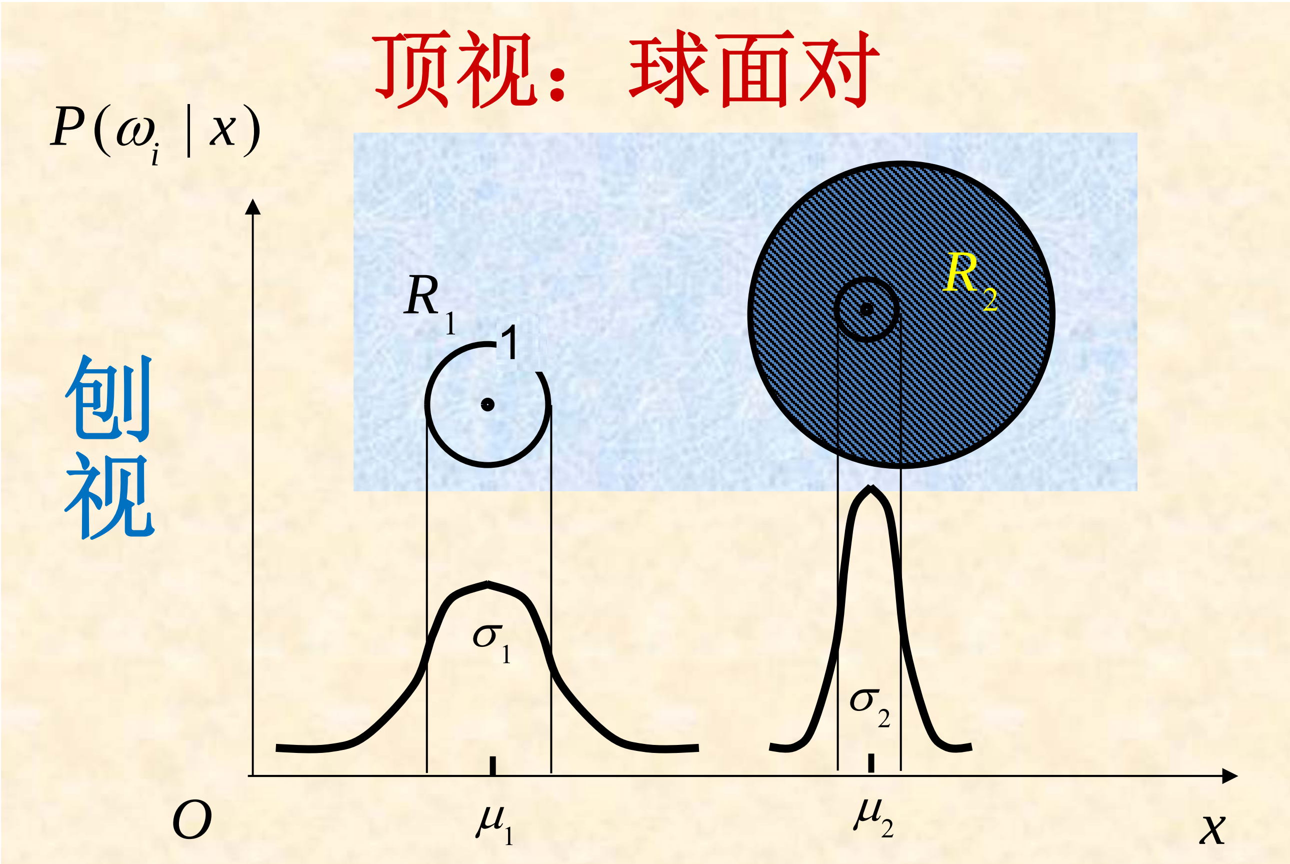
情况三

- 球面对

$\sigma_{1x} = \sigma_{1y}$

$\sigma_{2x} = \sigma_{2y}$

$\sigma_2 < \sigma_1$

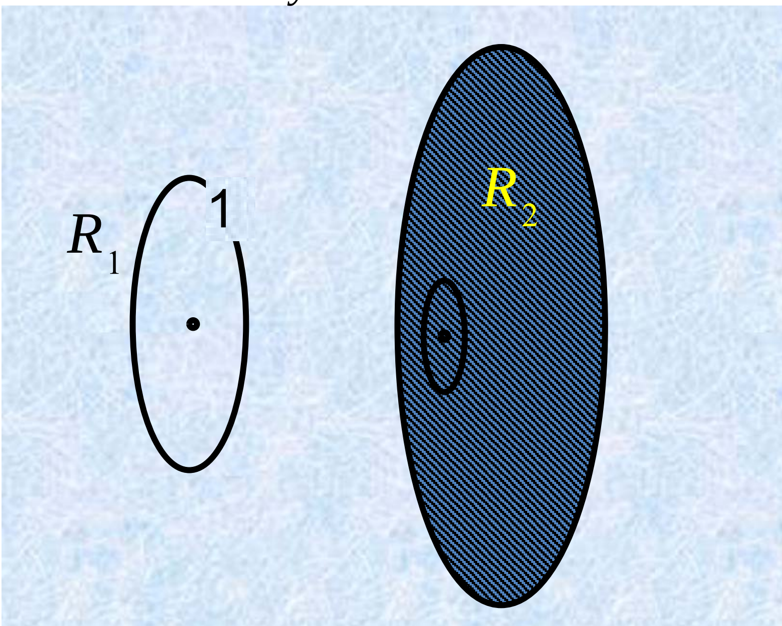


情况三

- 椭圆对

$\sigma_{1x} < \sigma_{1y}$ 且 $\sigma_2 < \sigma_1$

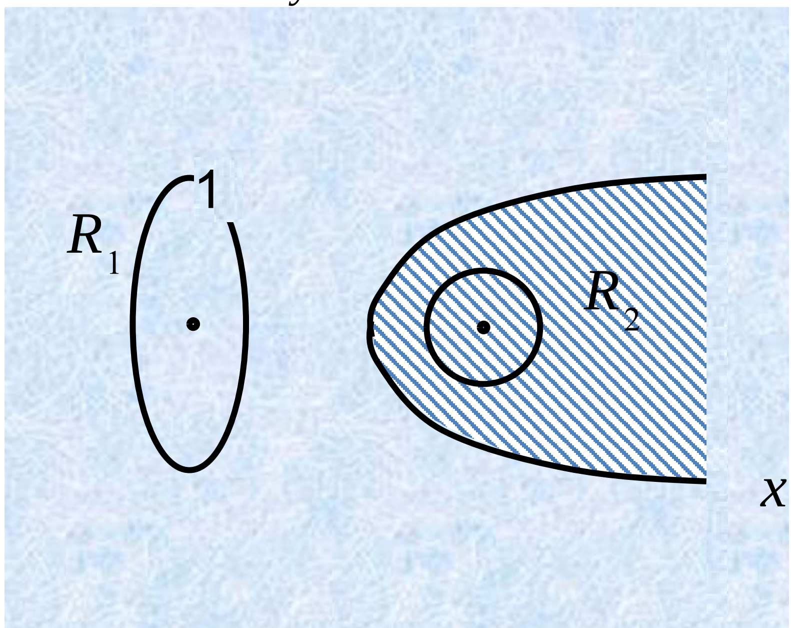
$\sigma_{2x} < \sigma_{2y}$



- 抛物线

$\sigma_{1x} < \sigma_{1y}$ 且 $\sigma_{1x} = \sigma_{2x}$

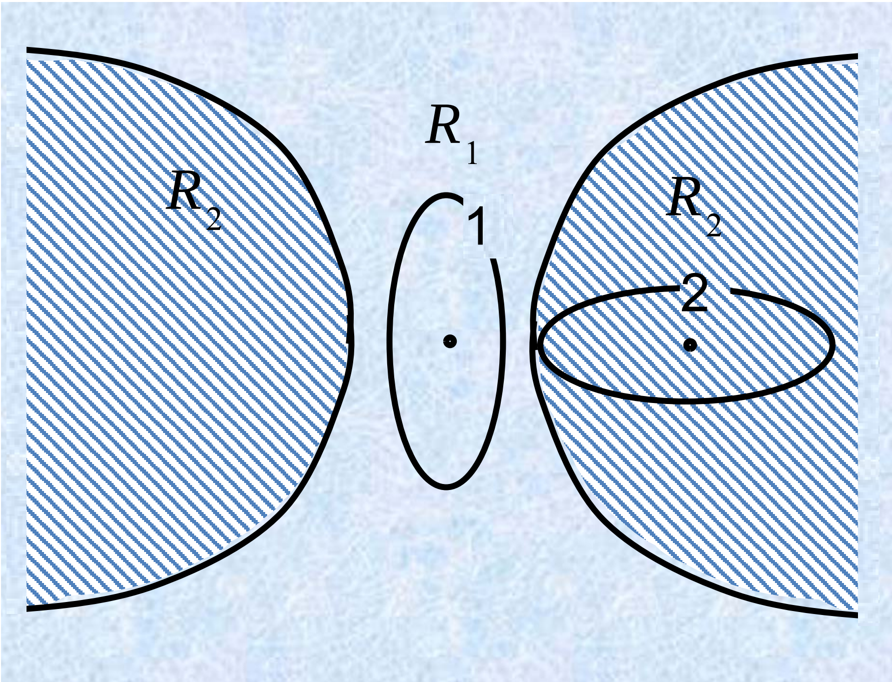
$\sigma_{2x} = \sigma_{2y}$



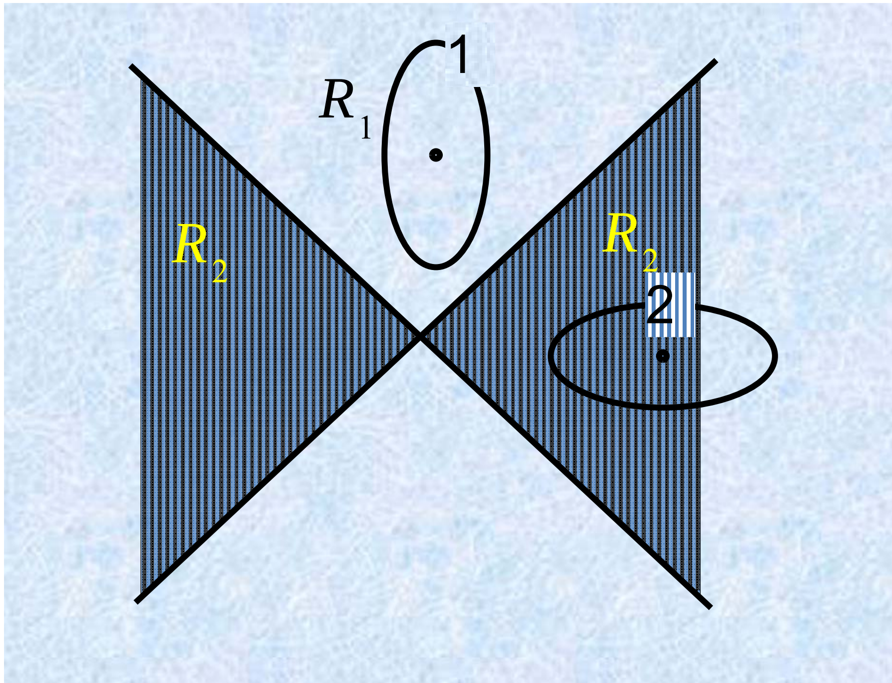
情况三

- 双曲线

$\sigma_{1x} < \sigma_{1y} \qquad \sigma_{2x} > \sigma_{2y}$

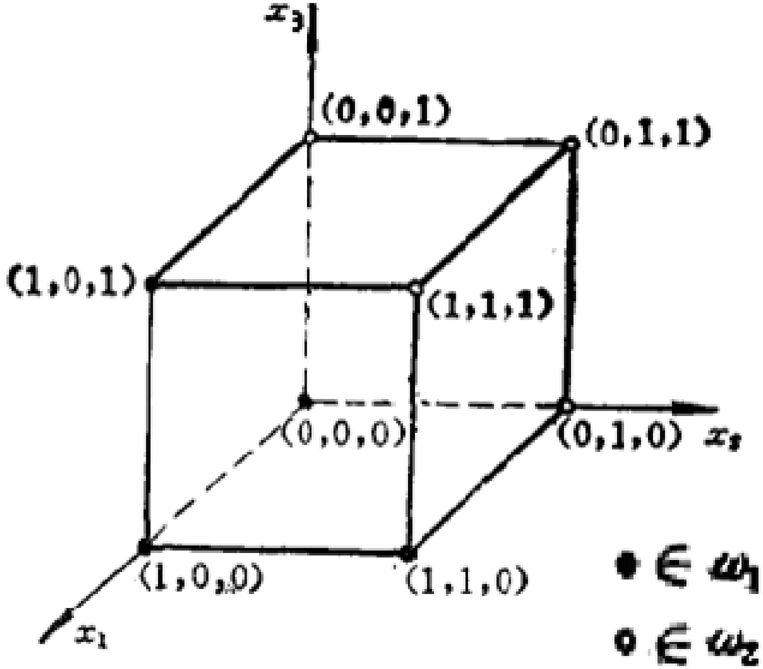


- 直线
  - (双曲线的退化形式)



举例

- 设在三维特征空间里，两类的类概率密度是正态分布的，分别在两个类型中获得4个样本，位于一个单位立方体的顶点上，如下图。两类的先验概率相等，试确定两类之间的决策面及相应的类型区域R_1和R_2。



解答

- w_1和w_2表示两个类型，由图可知，两个类型的样本：

w_1：$(0,0,0)^T,(1,0,0)^T,(1,1,0)^T,(1,0,1)^T$

w_2：$(0,1,0)^T,(0,0,1)^T,(0,1,1)^T,(1,1,1)^T$

用各类样本的算术平均值近似代替各类均值向量，也就是：

$$\mu_i \approx \frac{1}{N_i}\sum_{k=1}\mathbf{x}_{ik}$$

N_i为 w_i中的样本数，$\mathbf{x}_{ik}$表示 w_i 的第k个样本。

解答

- 协方差矩阵由其定义求得：

$$\Sigma_i = R_i - \mu_i \mu_j^T = \frac{1}{N_i}\sum_{k=1}^{N_i} \mathbf{x}_{ik} \cdot \mathbf{x}_{ik}^T - \mu_i \mu_i^T$$

式中R_i为类w_i的自相关函数。

由题中所给条件：$i = 1,2$，$N_1 = N_2 = 4$

有：$\mu_1 = \frac{1}{4}(3,1,1)^T$，$\mu_2 = \frac{1}{4}(1,3,3)^T$

$$\mu_1 \mu_1^T = (\frac{3}{4},\frac{1}{4},\frac{1}{4})^T \cdot (\frac{3}{4},\frac{1}{4},\frac{1}{4}) = \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} \cdot (\frac{3}{4},\frac{1}{4},\frac{1}{4}) = \frac{1}{16}\begin{bmatrix} 9 & 3 & 3 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

解答

$$\mu_2\mu_2^T=\begin{pmatrix}\frac{1}{4}\\ \frac{3}{4}\\ \frac{3}{4}\end{pmatrix}\cdot(\frac{1}{4},\frac{3}{4},\frac{3}{4})=\frac{1}{16}\begin{bmatrix}1&3&3\\3&9&9\\3&9&9\end{bmatrix}$$

$$R_1=\frac{1}{4}\left[\begin{pmatrix}0\\0\\0\end{pmatrix}\cdot(0,0,0)+\begin{pmatrix}1\\0\\0\end{pmatrix}\cdot(1,0,0)+\begin{pmatrix}1\\1\\0\end{pmatrix}\cdot(1,1,0)+\begin{pmatrix}1\\0\\1\end{pmatrix}\cdot(1,0,1)\right]=\frac{1}{4}\begin{bmatrix}3&1&1\\1&1&0\\1&0&1\end{bmatrix}\text{最大}$$

同理：

$$R_2=\frac{1}{4}\begin{bmatrix}1&1&1\\1&3&2\\1&2&2\end{bmatrix}$$

解答

$$\Sigma_1 = R_1 - \mu_1\mu_1^T = \frac{1}{4}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} - \frac{1}{16}\begin{bmatrix} 9 & 3 & 3 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix} = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

$$\Sigma_2 = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

因此，$\Sigma_1 = \Sigma_2 = \Sigma$ 符合情况二。用情况二的公式确定决策面。

$$\Sigma^{-1} = 4\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

解答

决策面为$g_1(\mathbf{x})-g_2(\mathbf{x})=0 \Rightarrow w^T(\mathbf{x}-\mathbf{x}_0)=0$，$w=\Sigma^{-1}(\mu_1-\mu_2)$，

$\mathbf{x}_0=\frac{1}{2}(\mu_1+\mu_2)$，先验概率相等 $p(w_1)=p(w_2)$

$$w=\Sigma^{-1}(\mu_1-\mu_2)=4\begin{bmatrix}2 & -1 & -1\\ -1 & 2 & 1\\ -1 & 1 & 2\end{bmatrix}\cdot\frac{1}{4}\begin{bmatrix}2\\ -2\\ -2\end{bmatrix}=\begin{bmatrix}8\\ -8\\ -8\end{bmatrix}$$

$$\mathbf{x}_0=\frac{1}{2}(\mu_1+\mu_2)=\frac{1}{2}(1,1,1)^T$$

决策方程： $w^T(\mathbf{x}-\mathbf{x}_0)=0$，

$$(8,-8,-8)\begin{bmatrix}x_1-\frac{1}{2}\\ x_2-\frac{1}{2}\\ x_3-\frac{1}{2}\end{bmatrix}=0$$

解答

也就是：$8(x_1-\frac{1}{2})-8(x_2-\frac{1}{2})-8(x_3-\frac{1}{2})=0$

$$8x_1-8x_2-8x_3+4=0$$

$$2x_1-2x_2-2x_3+1=0$$

如下图所示。

w指向的一侧为正，是w_1的区域R_1，负向的一侧为w_2。