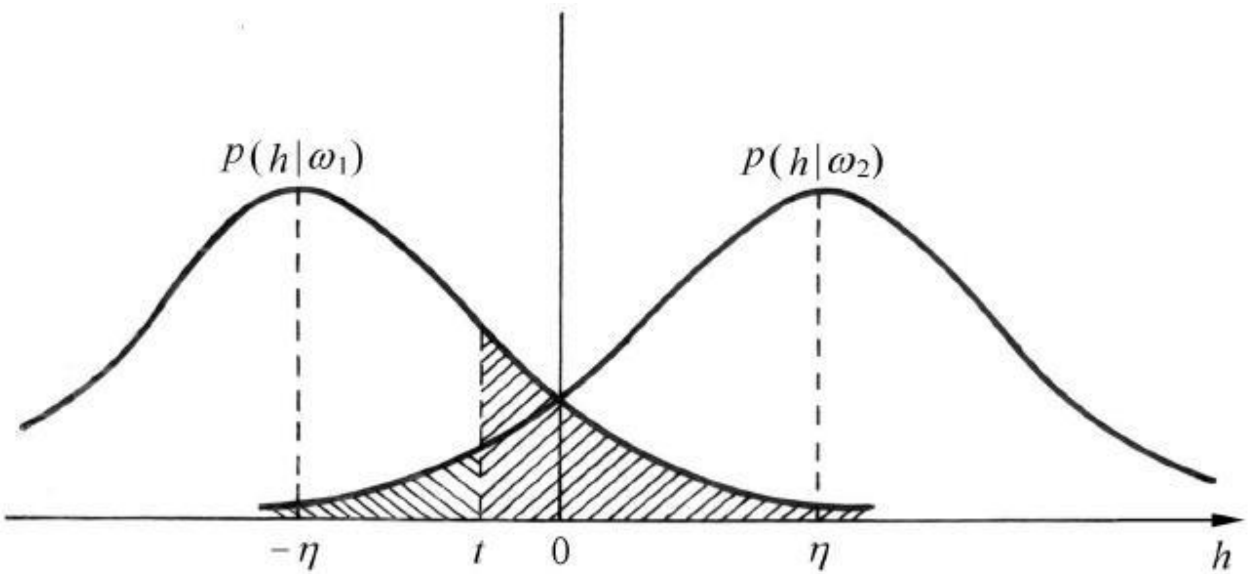
第二章
Bayes 决策理论

Bayes 决策的假设

- 概率方法
- 各类别总体概率分布已知；
- 分类数已知；
- 问题：根据高度信息对玉米与杂草分类



玉米和杂草图片来自百度

常用的决策规则 1.最小错误率

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

ω_1 = 玉米

ω_2 = 杂草

$$P(B|A)P(A) = P(B,A)$$

常用的决策规则　1. 最小错误率

如果$P(\omega_1 | \vec{x}) > P(\omega_2 | \vec{x})$,

则把x归类于玉米ω_1,

反之$P(\omega_1 | \vec{x}) < P(\omega_2 | \vec{x})$

则把x归类于杂草ω_2。

(1)如果$P(\omega_i | \vec{x}) = \max\limits_{j=1,2} P(\omega_j | \vec{x})$,则$\vec{x} \in \omega_i$

(2)如果$p(\vec{x} | \omega_i)P(\omega_i) = \max\limits_{j=1,2} p(\vec{x} | \omega_j)P(\omega_j)$,

则$\vec{x} \in \omega_i$

常用的决策规则 1. 最小错误率

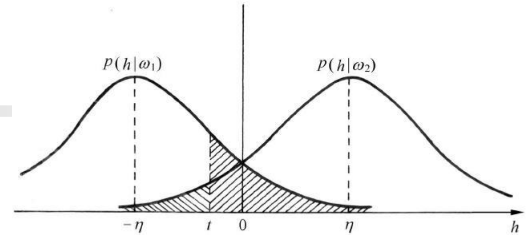
(3) 若 $l(\vec{x}) = \frac{P(\vec{x}|\omega_1)}{P(\vec{x}|\omega_2)} \begin{matrix} > \\ < \end{matrix} \frac{P(\omega_2)}{P(\omega_1)}$，则$\vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

(4) 若 $h(\vec{x}) = -\ln[l(\vec{x})] =$

$$-\ln p(\vec{x}|\omega_1) + \ln p(\vec{x}|\omega_2) \begin{matrix} < \\ > \end{matrix} \ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right),$$

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

常用的决策规则 1. 最小错误率
错误率分析



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

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

$$\begin{aligned} P(e) &= \int_{-\infty}^{t} P(\omega_2|\vec{x})P(\vec{x})d\vec{x} + \int_{t}^{+\infty} P(\omega_1|\vec{x})P(\vec{x})d\vec{x} \\ &= \int_{-\infty}^{t} P(\vec{x}|\omega_2)P(\omega_2)d\vec{x} + \int_{t}^{+\infty} P(\vec{x}|\omega_1)P(\omega_1)d\vec{x} \end{aligned}$$

常用的决策规则 1. 最小错误率错误率分析

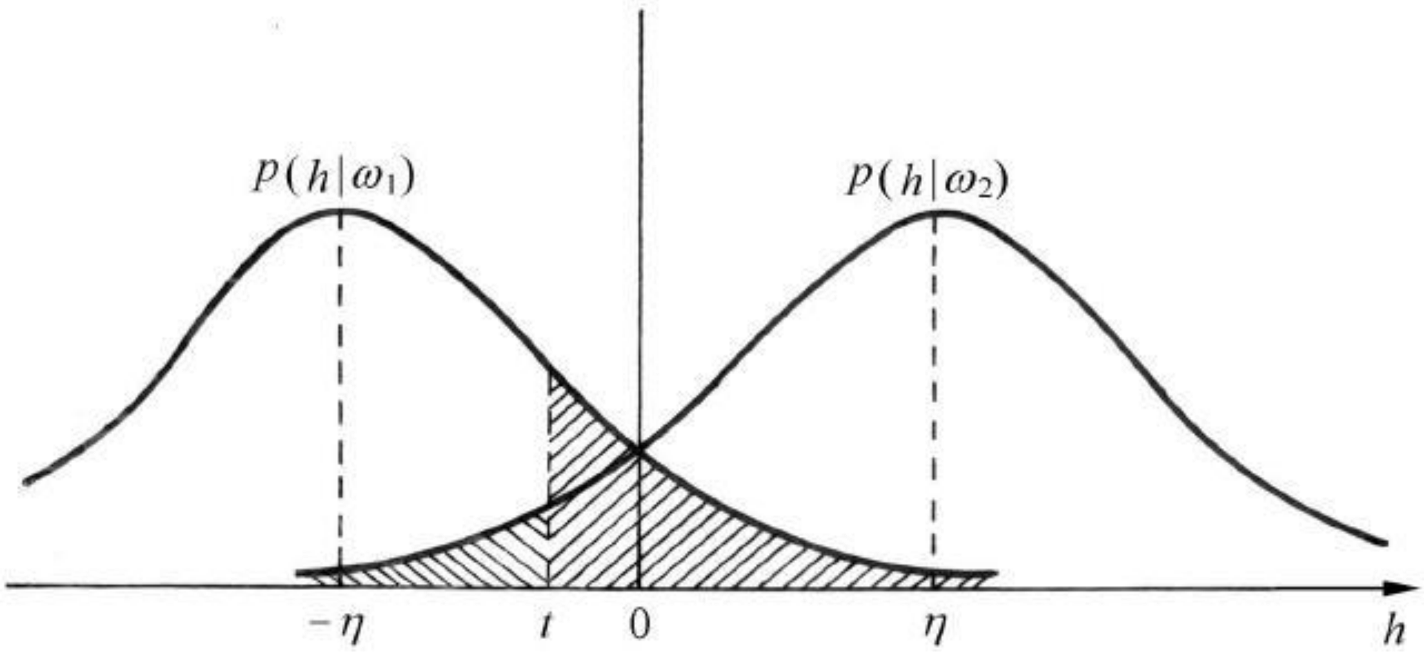
$$
\begin{aligned}
P(e) &= P(\vec{x} \in \Re_1, \omega_2) + P(\vec{x} \in \Re_2, \omega_1) \\
&= P(\vec{x} \in \Re_1 | \omega_2) P(\omega_2) + P(\vec{x} \in \Re_2 | \omega_1) P(\omega_1) \\
&= P(\omega_2) \int_{\Re`_1} P(\vec{x}|\omega_2) d\vec{x} + P(\omega_1) \int_{\Re`_2} P(\vec{x}|\omega_1) d\vec{x} \\
&= P(\omega_2) P_2(e) + P(\omega_1) P_1(e)
\end{aligned}
$$

常用的决策规则 1.最小错误率 错误率分析

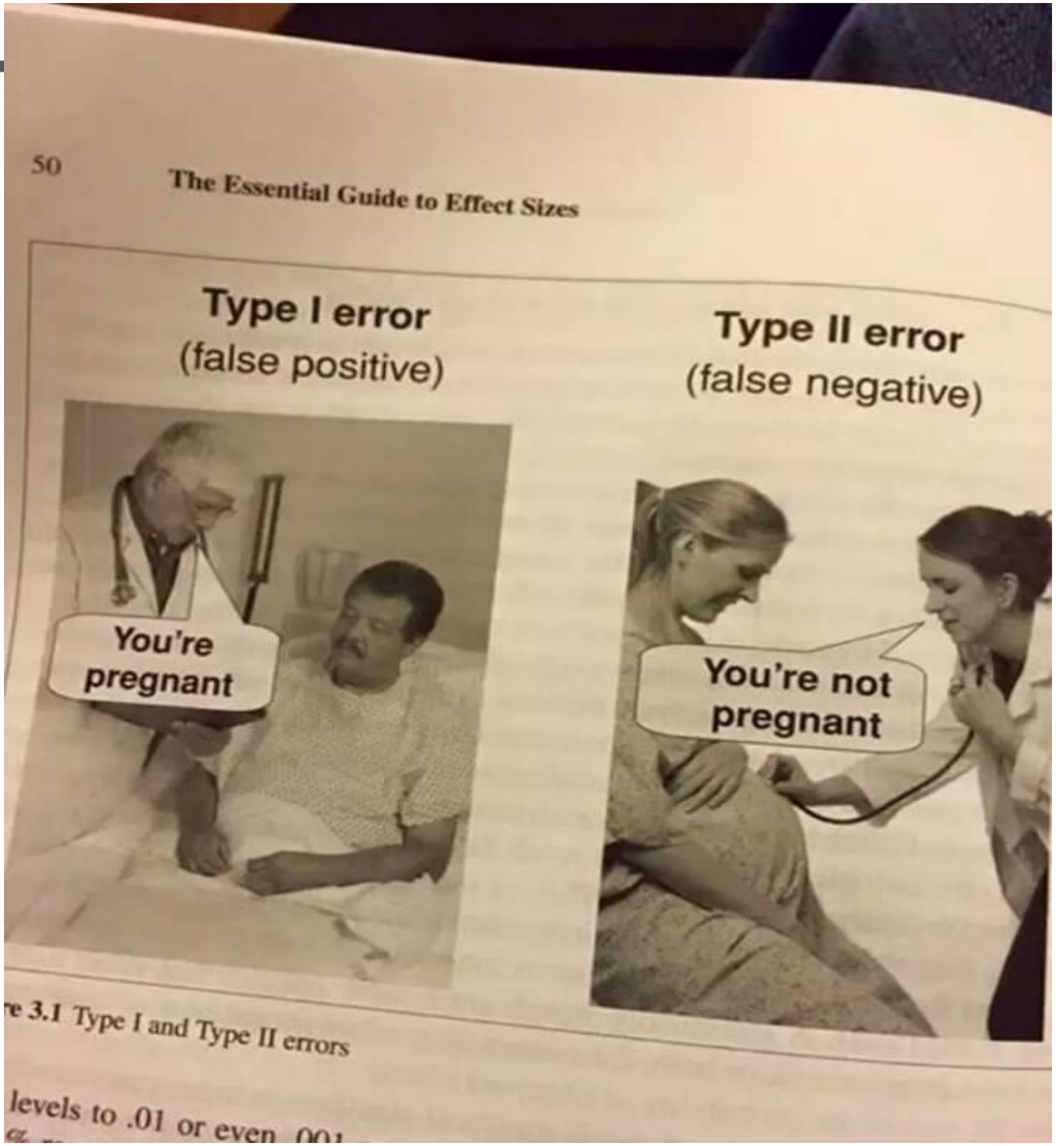
- 证明是最小错误率
- 两类(接收，拒绝)错误：虚警，漏警

$P_2(e)$

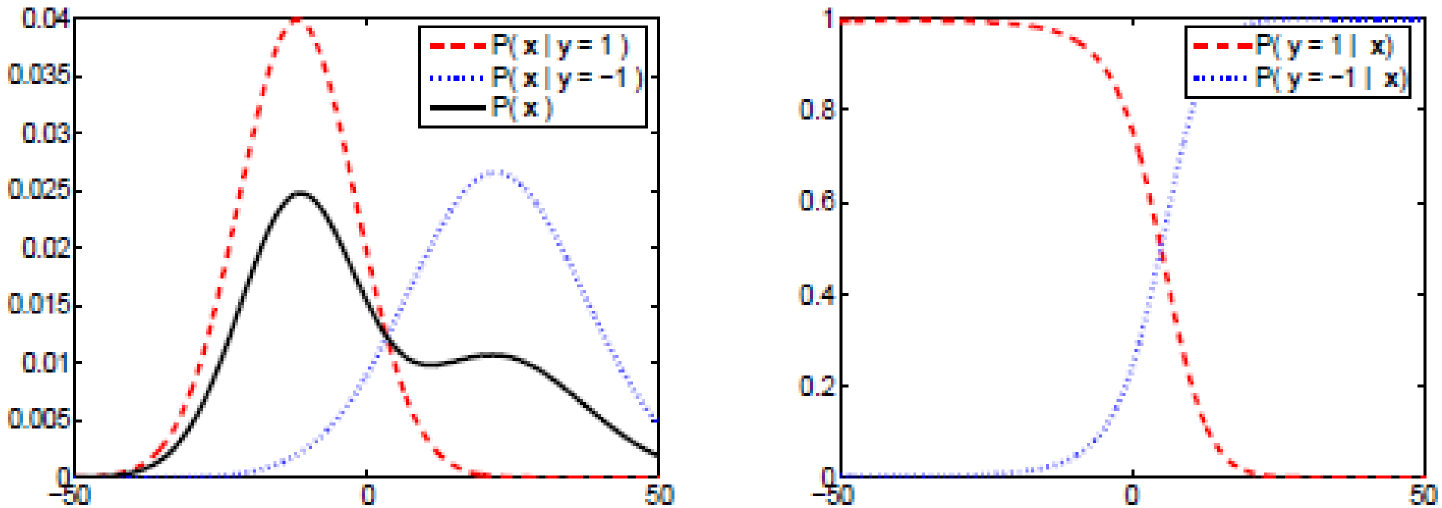
$P_1(e)$



两类错误


50
The Essential Guide to Effect Sizes
Type I error
(false positive)
You're pregnant
Type II error
(false negative)
You're not pregnant
re 3.1 Type I and Type II errors
levels to .01 or even .001

似然与后验

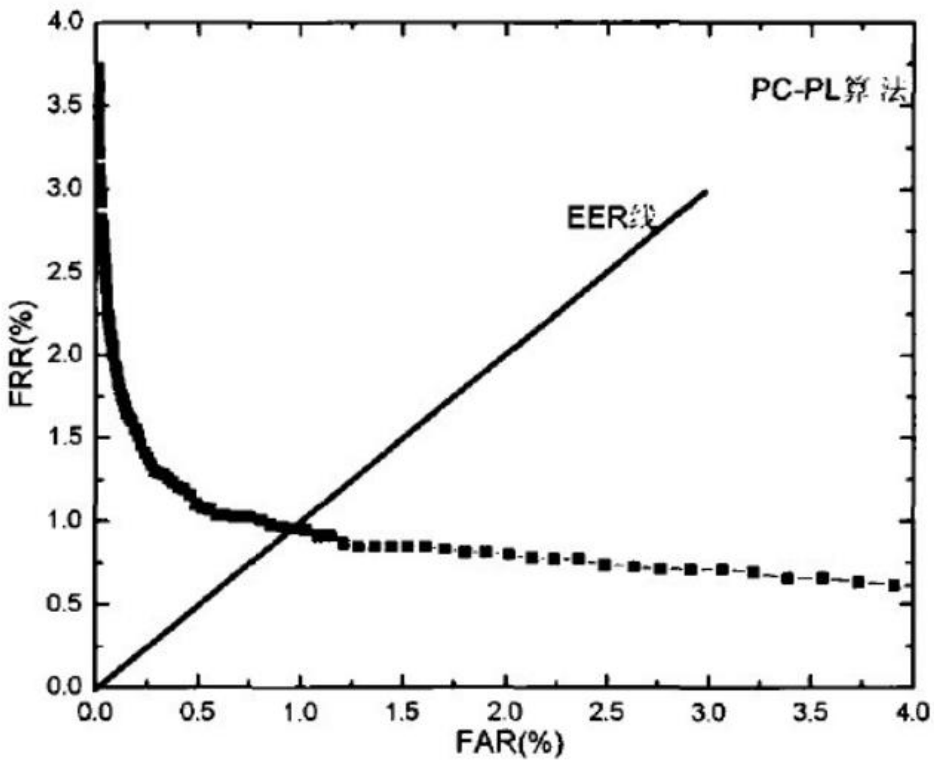


(a) 两类高斯分布以及其混合分布的概率密度　(b) 两类高斯分布的后验概率密度函数

图 2.1　贝叶斯公式的两类一维分布示例：第一类先验概率$P(y = 1) = 0.6$，第二类先验概率$P(y = 1) = 0.4$。

ROC曲线

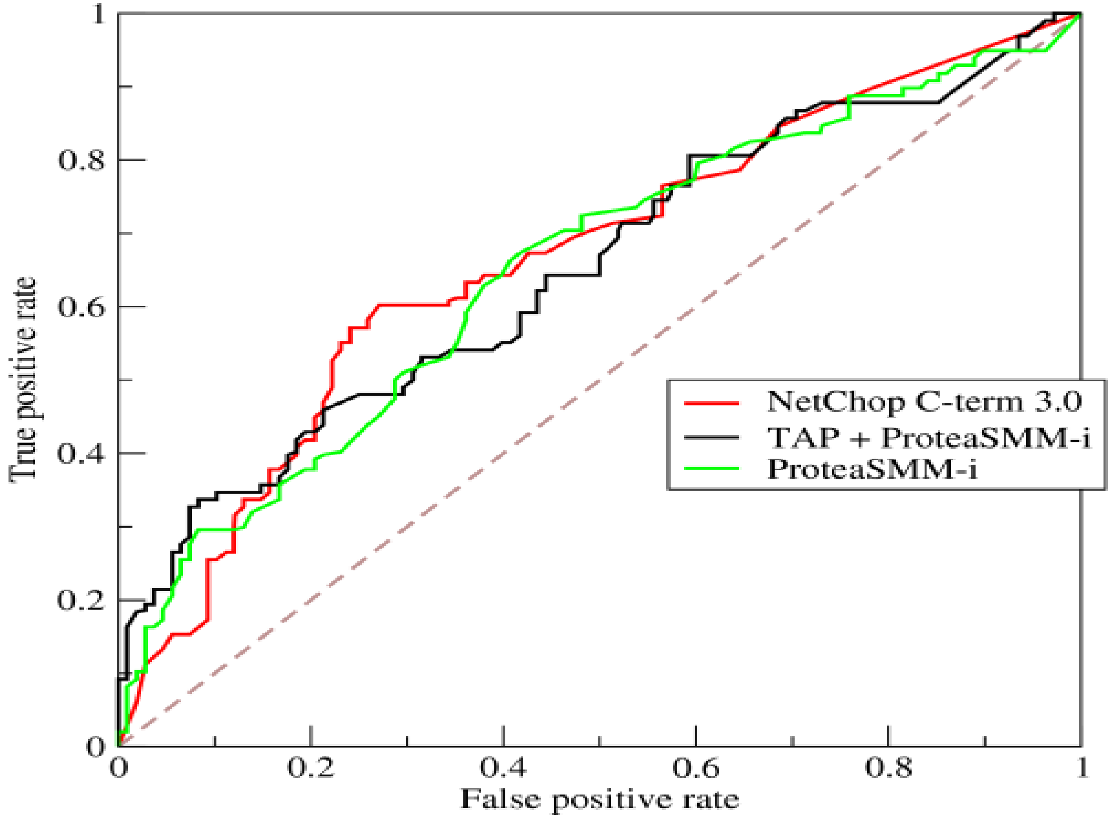
- ROC: Receiver Operating Characteristic



(*from http://blog.csdn.net/u014696921/article/details/74435229*)

ROC曲线

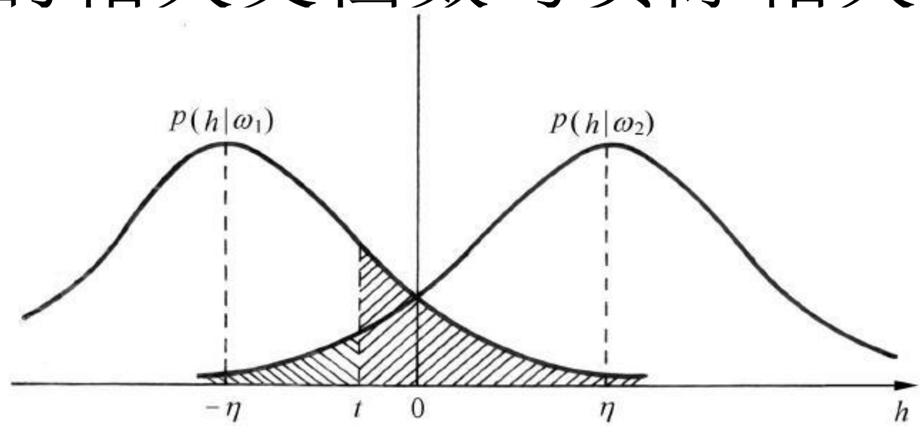
- ROC: Receiver Operating Characteristic



(ROC curve of three predictors of peptide cleaving in the proteasome *from https://en.wikipedia.org/wiki/Receiver_operating_characteristic*)

检索问题

- 检索问题：文本检索，图像检索
- 给出前k个搜索结果
- 准确率：检索出的相关文档数与检出文档数的比值
- 召回率：检索出的相关文档数与实际相关文档数的比值



常用的决策规则 1. 最小错误率
多类分类时的错误率分析

$$\left.\begin{aligned}P(e) \;=\; & [P(\vec{x}\in \Re_2|\omega_1)+P(\vec{x}\in \Re_3|\omega_1)+\ \ldots + P(\vec{x}\in \Re_c|\omega_1)]\,P(\omega_1)\\ & +[P(\vec{x}\in \Re_1|\omega_2)+P(\vec{x}\in \Re_3|\omega_2)+\ \ldots + P(\vec{x}\in \Re_c|\omega_2)]\,P(\omega_2)\\ & +[\ldots]\\ & +[P(\vec{x}\in \Re_1|\omega_c)+P(\vec{x}\in \Re_2|\omega_c)+\ \ldots + P(\vec{x}\in \Re_c|\omega_c)]\,P(\omega_c)\end{aligned}\right\} C\text{ 行}$$

$$\underbrace{\hspace{20em}}_{\text{每行}C\text{-1项}}$$

$$= \sum_{i=1}^{c}\sum_{j=1,j\neq i}^{c}[P\left(\vec{x}\in \Re_j|\omega_i\right)]\,P(\omega_i)$$

$$P(c) = \sum_{j=1}^{c} P\left(\vec{x} \in \Re_j | \omega_j\right) P\left(\omega_j\right) = \sum_{j=1}^{c} \int_{\Re_i} P\left(\vec{x} | \omega_j\right) P\left(\omega_j\right) d\vec{x}$$

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

2. 最小风险贝叶斯决策

- 不同的决策所带来的损失可能不同；
- 举例：
- 变量：x
- C个类别
- 决策空间　$\{\alpha_1, \alpha_2, \dots, \alpha_a\}$

$\lambda\left(\alpha_i, \omega_j\right)$：损失函数

$$R(\alpha_i|\vec{x}) = E[\lambda\left(\alpha_i, \omega_j\right)]$$

$$= \sum_{j=1}^{c} \lambda\left(\alpha_i, \omega_j\right) P\left(\omega_j|\vec{x}\right), i = 1,2, \ldots, a$$

$$R = \int R\left(\alpha \,|(\vec{x})|\vec{x}\right) P(\vec{x}) \boldsymbol{d}\, \vec{x}$$

$$R(\alpha_k|\vec{x}) = \min_{i=1,\dots,a} R(\alpha_i|\vec{x})$$

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

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

$$= \sum_{\substack{j=1 \\ j\neq i}}^{c} P\left(\omega_j|\vec{x}\right)$$

$$P\left(\omega_i|\vec{x}\right)$$

3. 限定一类错误率条件下使另一类错误率最小的两类决策

第二类的错误率不能大于一个值

$$\gamma = P_1(e) + \lambda(P_2(e) - \varepsilon_0)$$

$$P_1(e) = \int_{\Re_2} p(\vec{x}|\omega_1)d\vec{x}$$

$$P_2(e) = \int_{\Re_1} p(\vec{x}|\omega_2)d\vec{x}$$

$$\frac{\partial\gamma}{\partial\lambda} = \int_{\Re_1} p(\vec{x}|\omega_2)d\vec{x} - \varepsilon_0$$

$$\gamma = \int_{\Re_2} p(\vec{x}|\omega_1)d\vec{x} + \lambda[\int_{\Re_1} p(\vec{x}|\omega_2)d\vec{x} - \varepsilon_0]$$

$$= 1 - \int_{\Re_1} p(\vec{x}|\omega_1)d\vec{x} + \lambda[\int_{\Re_1} p(\vec{x}|\omega_2)d\vec{x} - \varepsilon_0]$$

$$= (1 - \lambda\varepsilon_0) + \int_{\Re_1} [\lambda p(\vec{x}|\omega_2) - p(\vec{x}|\omega_1)]d\vec{x}$$

- 令 $\frac{\partial \gamma}{\partial t} = 0$，及 $\frac{\partial \gamma}{\partial \lambda} = 0$

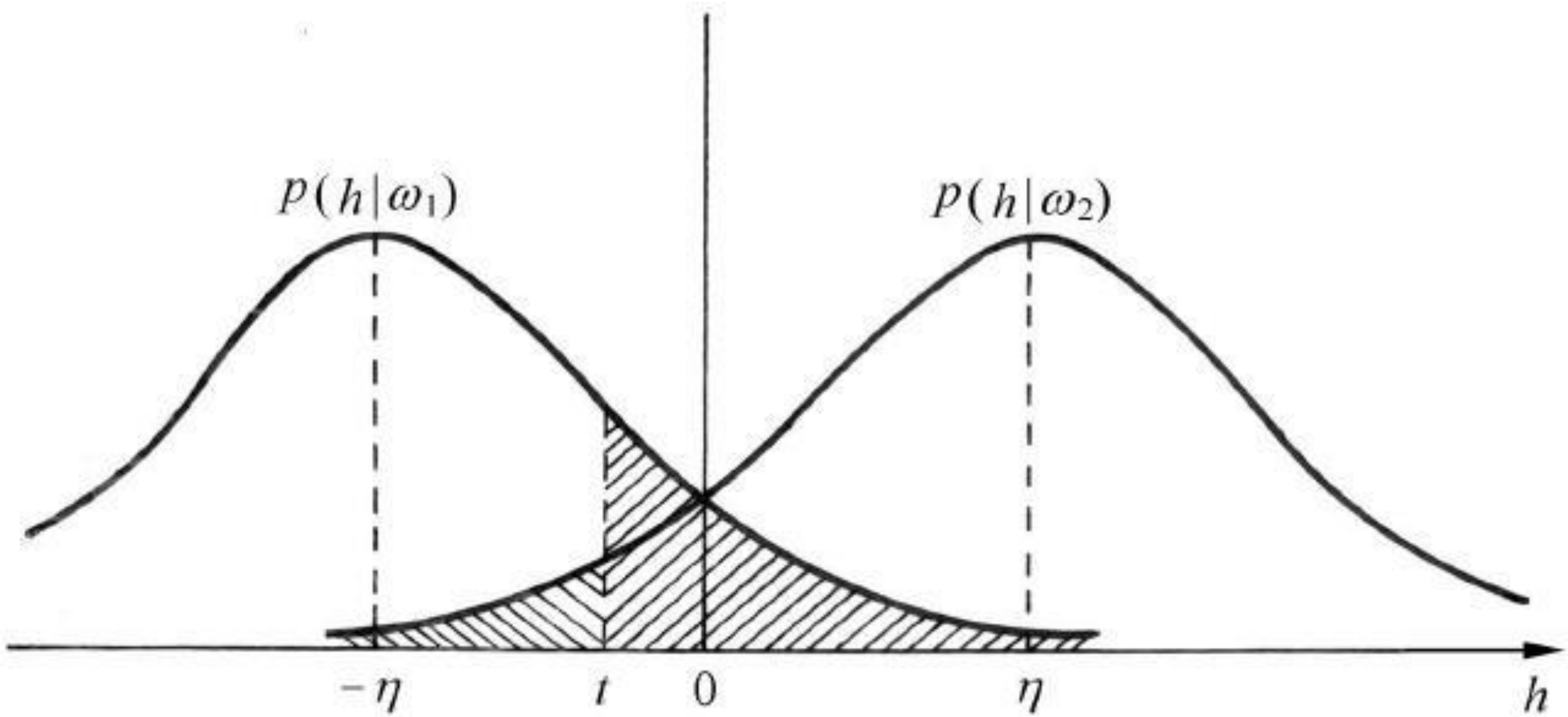
- 就得

$$\int_{\Re_1} p(x|\omega_2)dx = \varepsilon_0$$

$$\lambda = \frac{p(t|\omega_1)}{p(t|\omega_2)}$$

$$\lambda p(x|\omega_2) < p(x|\omega_1),$$

则$x \in \omega_1$,否则$x \in \omega_2$



$f(x)$: is continuous function.

$$\mathrm{Y}(y) = \{x : f(x) \ \geq y\}$$

$$m\big(Y(y)\big) = \int_{Y(y)} dx = \lambda_0$$

$$\int_{Y(y)} g(x) dx = \lambda_0$$

$$f(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)}, g(x) = p(x|\omega_2)$$

4．最小最大决策

先验概率未知

$$\lambda_{11} \qquad \lambda_{21} \qquad \lambda_{22} \qquad \lambda_{12}$$

$$\lambda_{21} > \lambda_{11} \qquad \lambda_{12} > \lambda_{22}$$

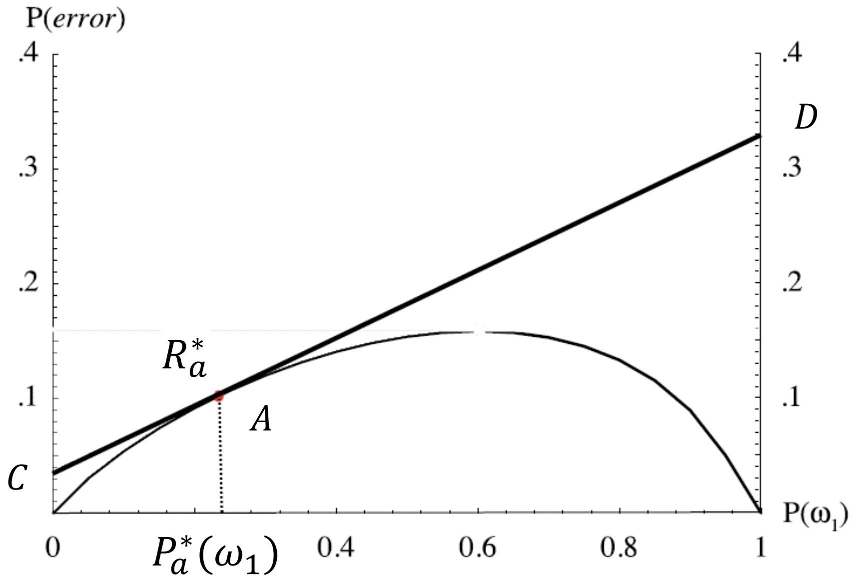
$$
\begin{aligned}
R &= \int R(\alpha(x)|x)p(x)dx \\
&= \int_{\mathfrak{R}_1} R(\alpha_1|x)p(x)dx + \int_{\mathfrak{R}_2} R(\alpha_2|x)p(x)dx \\
&= \int_{\mathfrak{R}_1} [\lambda_{11}p(\omega_1)p(x|\omega_1) + \lambda_{12}p(\omega_2)p(x|\omega_2)]dx \\
&+ \int_{\mathfrak{R}_2} [\lambda_{21}p(\omega_1)p(x|\omega_1) + \lambda_{22}p(\omega_2)p(x|\omega_2)]dx
\end{aligned}
$$

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

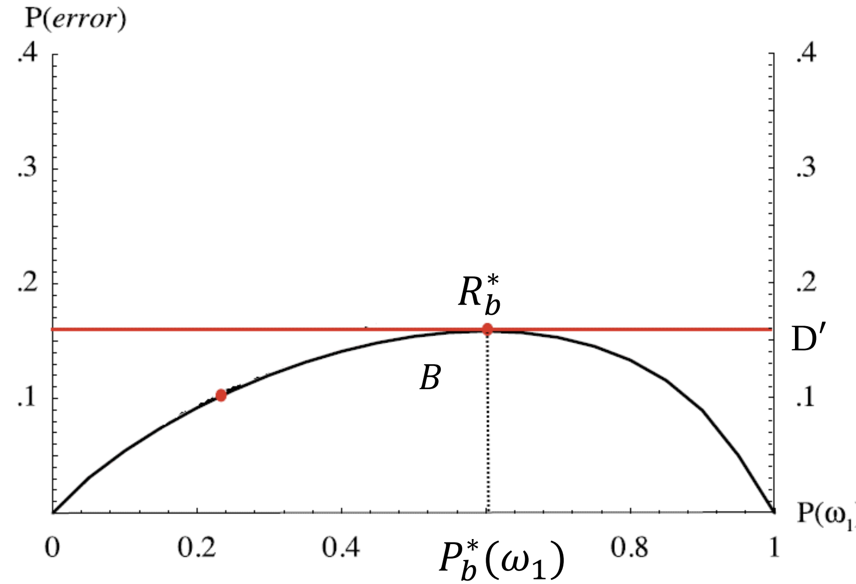
$$R = \ \lambda_{22} + (\lambda_{12} \ - \lambda_{22}) \int_{\Re_1} p(x|\omega_2) dx + P(\omega_1)[(\lambda_{11} \ - \lambda_{22})$$

$$+(\lambda_{21} \ - \lambda_{11}) \int_{\Re_2} p(x|\omega_1) dx - (\lambda_{12} \ - \lambda_{22}) \int_{\Re_1} p(x|\omega_2) dx]$$

$$= \ a + P(\omega_1) b$$



P(*error*)
.4
.3
.2
.1
D
R_a^*
A
C
P(ω1)
0
$P_a^*(\omega_1)$
0.4
0.6
0.8
1



P(*error*)
.4
.3
.2
.1
R_b^*
D′
B
P(ω1)
0
0.2
0.4
$P_b^*(\omega_1)$
0.8
1

分类器的设计-多类分类

- (1) $P(\omega_i|x) > P(\omega_j|x)$

  $j = 1,2,\ldots,c$ 且 $j \neq i \;\; \rightarrow \;\; x \in \omega_i$

- (2) $p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j)$

  $j = 1,2,\ldots,c$ 且 $j \neq i \;\; \rightarrow \;\; x \in \omega_i$

- (3) $$l(x) = \frac{p(x|\omega_i)}{p(x|\omega_j)} = \frac{P(\omega_j)}{P(\omega_i)}$$

  $$j = 1,2,\ldots,c \text{ 且 } j \neq i \quad \rightarrow \quad x \in \omega_i$$

- (4) $$\ln\ p(x|\omega_i) + \ln P(\omega_i) > \ln\ p\left(x|\omega_j\right) + \ln P(\omega_j)$$

  $$j = 1,2,\ldots,c \text{ 且 } j \neq i \quad \rightarrow \quad x \in \omega_i$$

- （1）判别函数

Discriminant functions

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

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

$$P(\omega_i|x) > P(\omega_j|x)$$

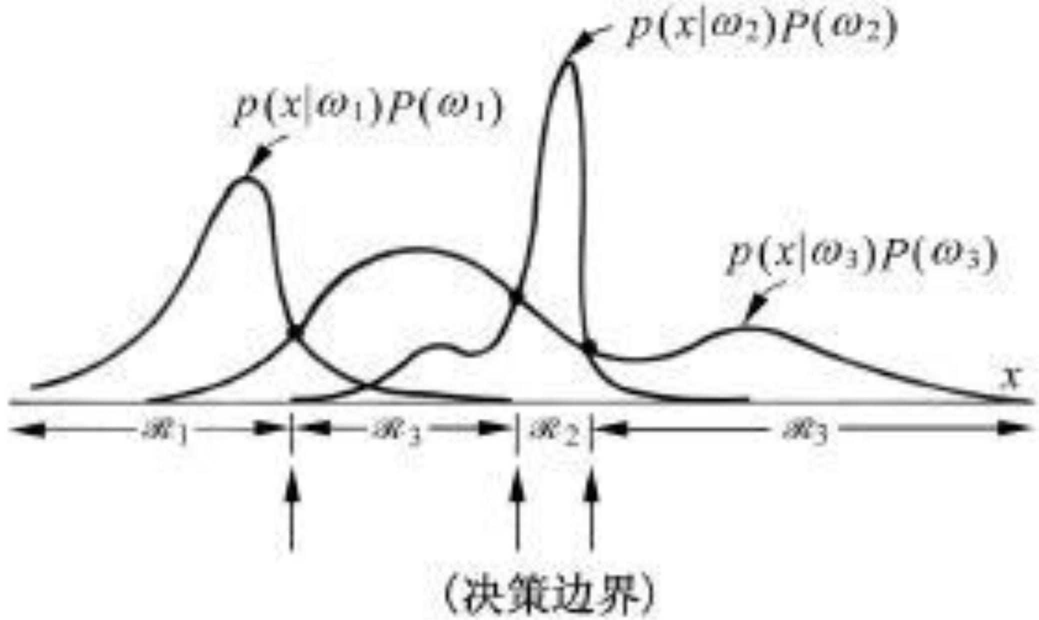
$$j = 1,2,\dots,c \text{ 且 } j \neq i \rightarrow x \in \omega_i$$

- (2) 决策面方程
- Decision boundary, decision surfaces

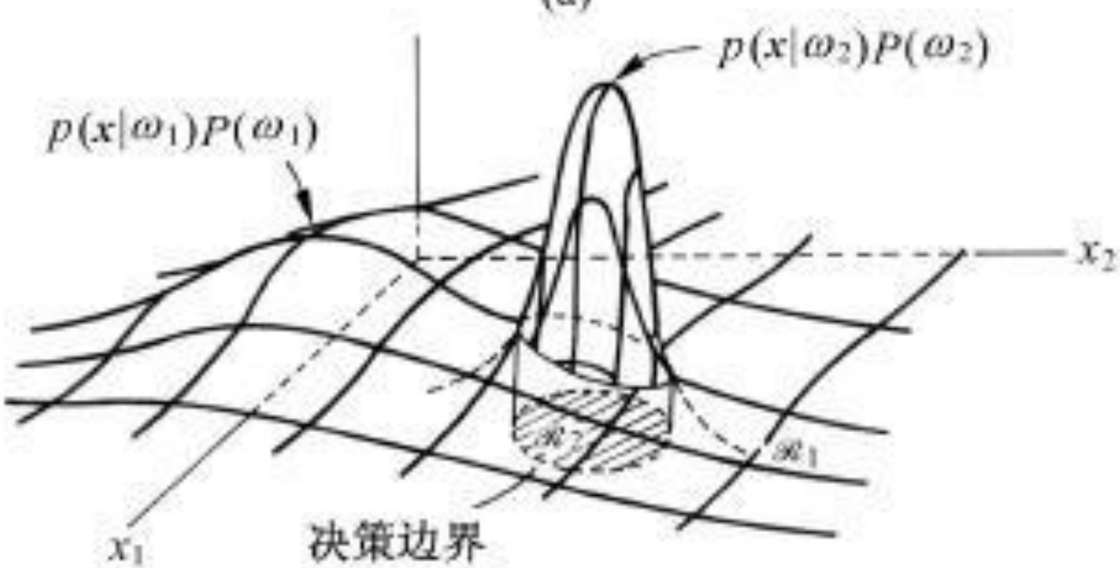
$$g_i(x) = g_j(x)$$

- (3)分类器设计(classifier design)
- 两类情况(the two-category case)

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


$p(x|\omega_2)P(\omega_2)$
$p(x|\omega_1)P(\omega_1)$
$p(x|\omega_3)P(\omega_3)$
x
$\mathscr{R}_1$
$\mathscr{R}_3$
$\mathscr{R}_2$
$\mathscr{R}_3$
（决策边界）

(a)

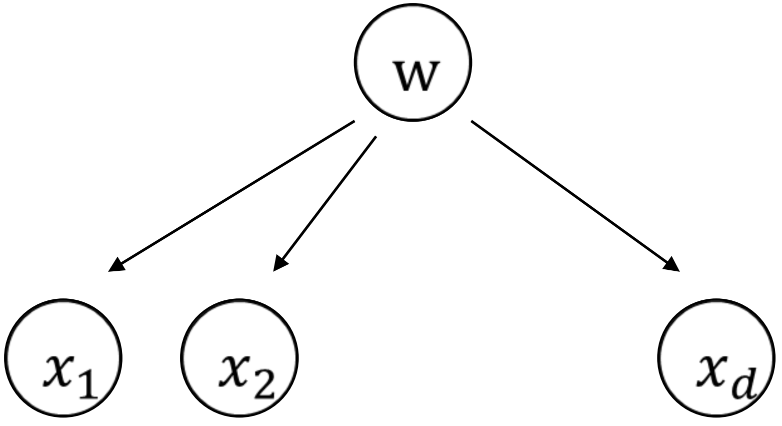

$p(x|\omega_2)P(\omega_2)$
$p(x|\omega_1)P(\omega_1)$
x_2
$\mathscr{R}_2$
$\mathscr{R}_1$
x_1
决策边界

(b)

朴素贝叶斯分类器（Naïve Bayes）

$$p(x_1, x_2, \ldots, x_d|w) = p(x_1|w)p(x_2|w)\ldots p(x_d|w)$$

- 贝叶斯网络
- 因果网络
- 贝叶斯信念网络



2.3　正态分布时的统计决策

- 正态分布及性质

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

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

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

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

$$\mu = [\mu_1, \mu_2, \dots \mu_d]^T$$

$\sum$ 是$d \times d$ 维协方差矩阵

正态分布的性质

- (1) μ, Σ 唯一决定分布$N(\mu, \Sigma)$
- (2)等密度点的轨迹为一个超椭球面

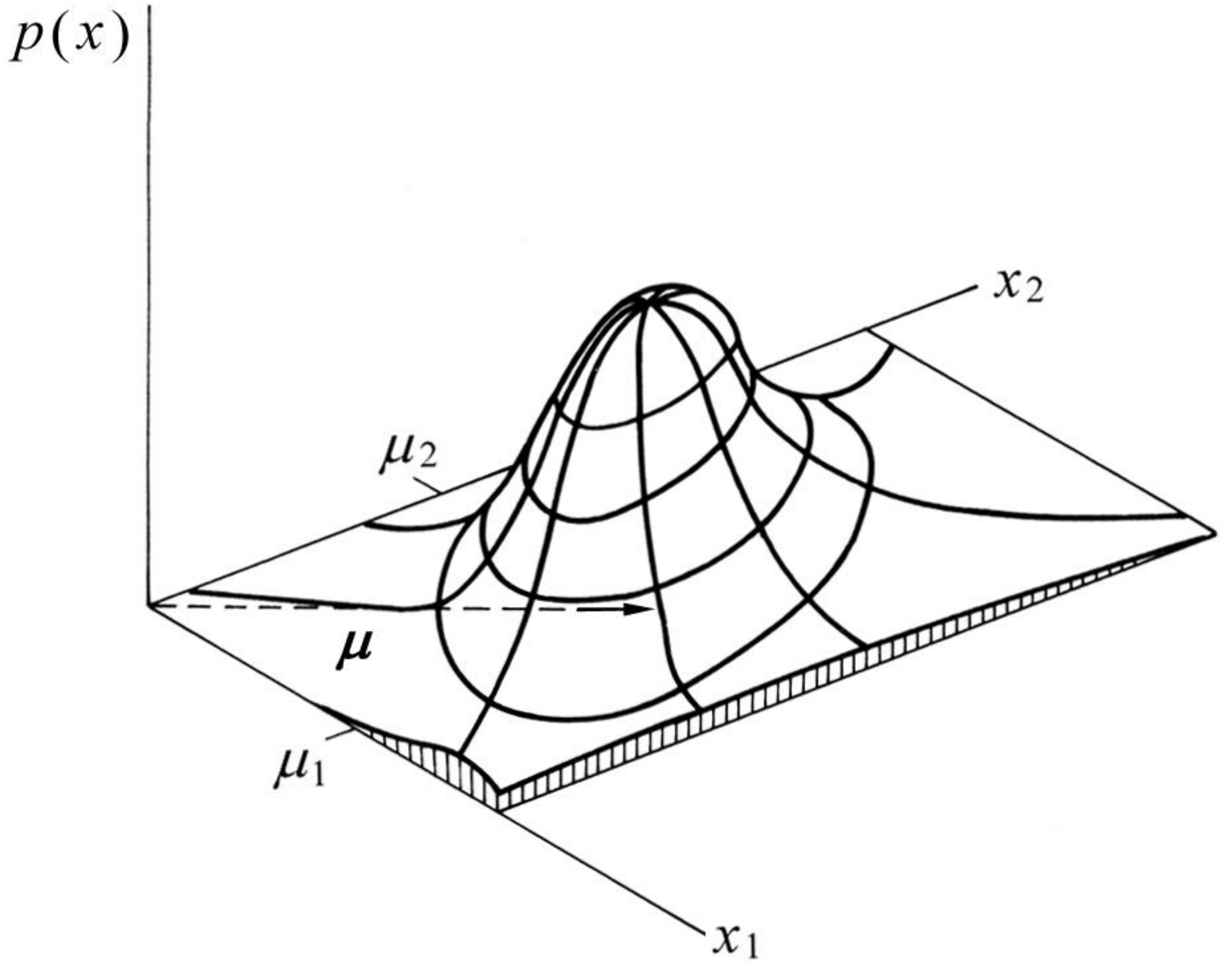
$$\gamma^2 = -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) = \text{常数}$$

x 到 μ 的M$ahalanobis$距离平方：马氏距离

- (3)不相关性=独立性

$$E\{x_i, x_j\} = E\{x_i\} \cdot E\{x_j\}$$

$$p(x_i, x_j) = p(x_i)p(x_j)$$


$p(x)$
x_2
μ_2
$\boldsymbol{\mu}$
μ_1
x_1

- (4)边缘分布与条件分布是正态的
- (5)线性变换的正态性

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

$$p(\mathrm{y}) \to N(A\mu, A\sum A^T)$$

- (6)线性组合的正态性

$$y = A^T x \quad p(\mathrm{y}) \to N(a^T\mu, a^T \sum a\ )$$

正态情况下贝叶斯判别

- 为什么研究正态分布下的贝叶斯判别？

- 1.合理
- 2.简便

正态情况下贝叶斯判别

$$p(x/\omega_i) \to N(\mu_i, \Sigma_i)$$

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

$$g(x) = \ln P(x/\omega_i) + \ln P(\omega_i)$$

$$= -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

$$-\frac{1}{2}(x-\mu_i)^T \sum_i^{-1} (x-\mu_i)$$

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

即： $$-\frac{1}{2}\left[(x-\mu_i)^T \sum_i^{-1} (x-\mu_i) - (x-\mu_j)^T \sum_j^{-1} (x-\mu_j)\right]$$

$$+\ln\frac{P(\omega_i)}{P(\omega_j)} - \frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|} = 0$$

• 第一种情况：

$$\sum_i = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}_{d\times d} = \sigma^2 I$$

$$\sum_i^{-1} = \frac{1}{\sigma^2} I$$

$$\left|\sum_{-1}\right| = \sigma^{2d}$$

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

$$x^T x = \sum_{i=1}^{c} x_i^2$$

- (1) $$P(\omega_i) = P(\omega_j)$$

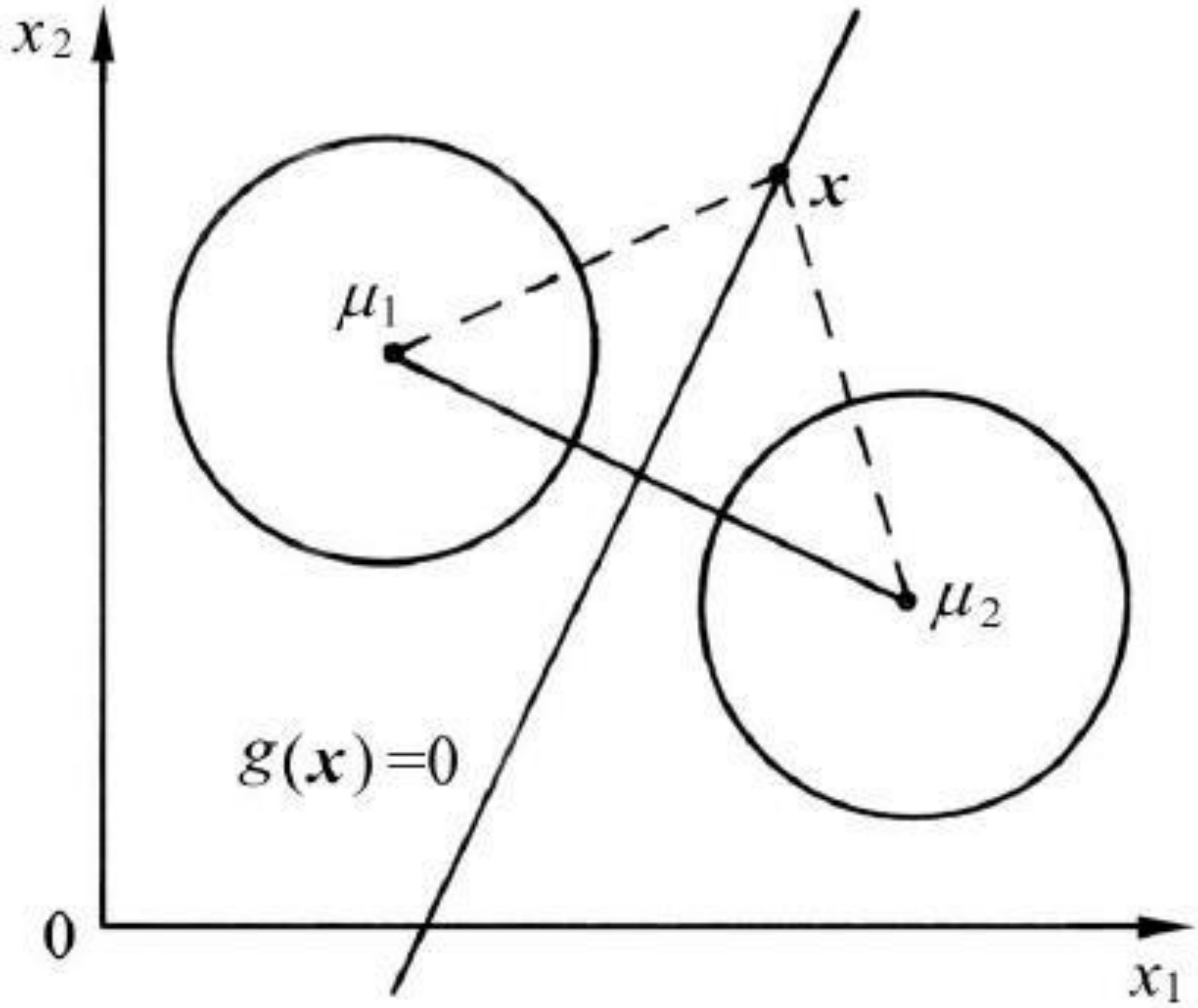
$$g_i(x) = -\frac{(x-\mu_i)^T(x-\mu_i)}{2\sigma^2}$$

$$g_i(x) = -(x-\mu_i)^T(x-\mu_i)$$

$$g_i(x) > g_j(x) \rightarrow x \in \omega_i$$

$$g_i(x) < g_j(x) \rightarrow x \in \omega_j$$

直观解释


x_2
x
μ_1
μ_2
$g(x)=0$
0
x_1

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

$$= \omega_i x + \omega_{i0}$$

$$\omega_i = \frac{1}{\sigma^2}\mu_i$$

$$\omega_{i0} = -\frac{1}{2\sigma^2}\mu_i^T \mu_i + \ln P(\omega_i)$$

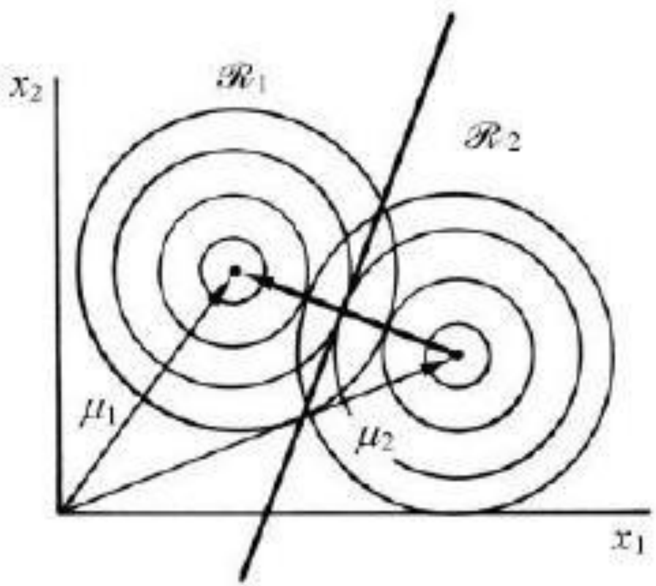
$$g_i(x) = \max_k g_k(x)$$

$$g_i(x) - g_j(x) = 0$$

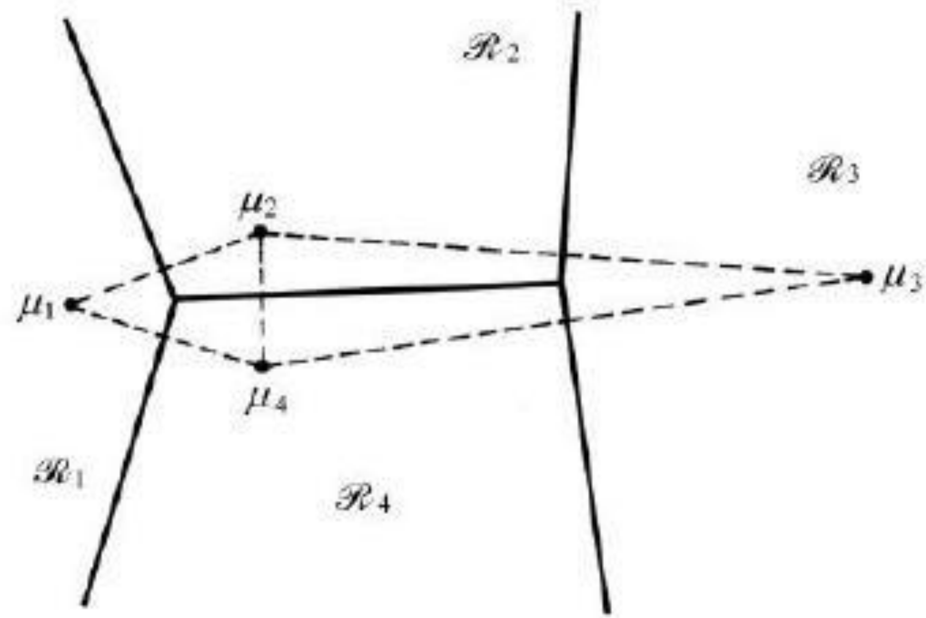
$$\omega^T(x - x_0) = 0, \omega = \mu_i - \mu_j$$

$$x_0 = \frac{1}{2}\left(\mu_i + \mu_j\right) - \frac{\sigma^2}{\left\|\mu_i - \mu_j\right\|^2} \times \ln \frac{P(\omega_i)}{P\left(\omega_j\right)}(\mu_i - \mu_j)$$

直观解释


x_2
$\mathscr{R}_1$
$\mathscr{R}_2$
μ_1
μ_2
x_1

(a) 两类情况


$\mathscr{R}_2$
$\mathscr{R}_3$
μ_2
μ_1
μ_3
μ_4
$\mathscr{R}_1$
$\mathscr{R}_4$

(b) 多类情况

- 第二种情况： $\sum_i = \sum$

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

如果先验概率相等，就是马氏距离

$$g_i(x) = (x-\mu_i)^T \sum^{-1}(x-\mu_i)$$

$$g_i(x) = \omega_i^T x + \omega_{i0}$$

$$\omega_i = \sum^{-1}\mu_i$$

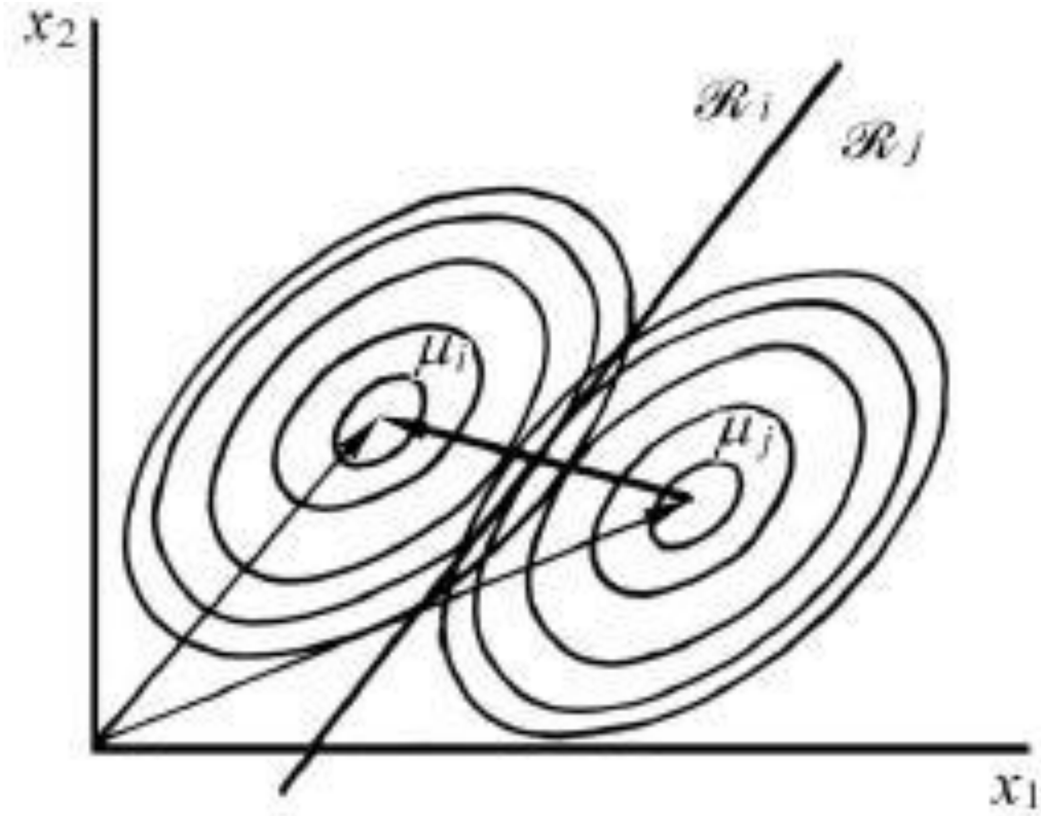
$$\omega_{i0} = -\frac{1}{2}{\mu_i}^2 \sum^{-1}\mu_i + \ln P(\omega_i)$$

$$g_i(x) - g_j(x) = 0$$

$$\omega^T(x - x_0) = 0$$

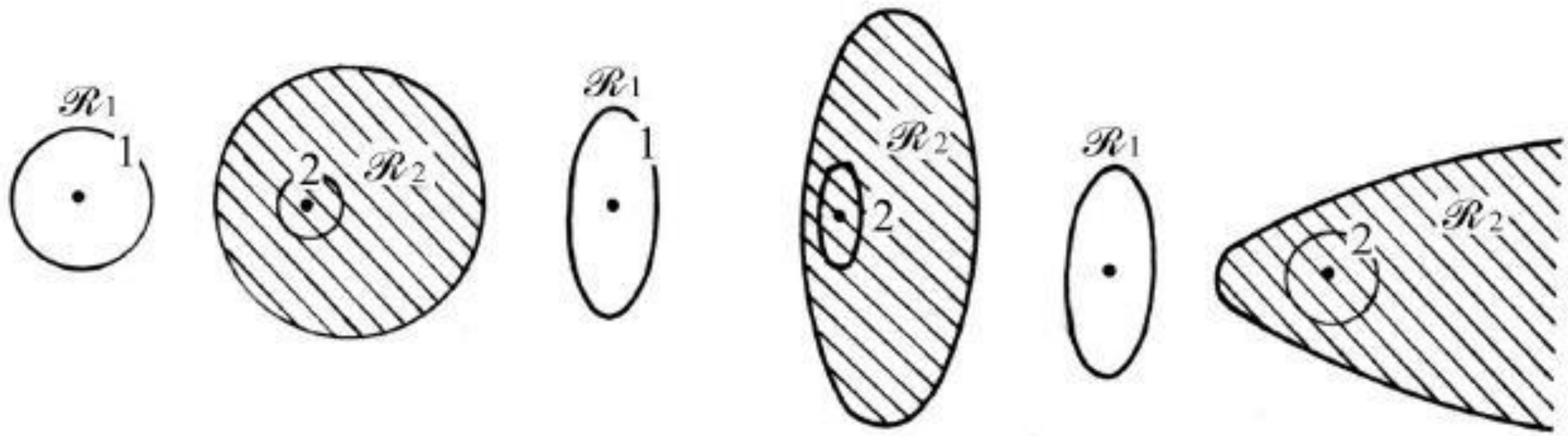
$$\omega = \Sigma^{-1}(\mu_i - \mu_j)$$

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

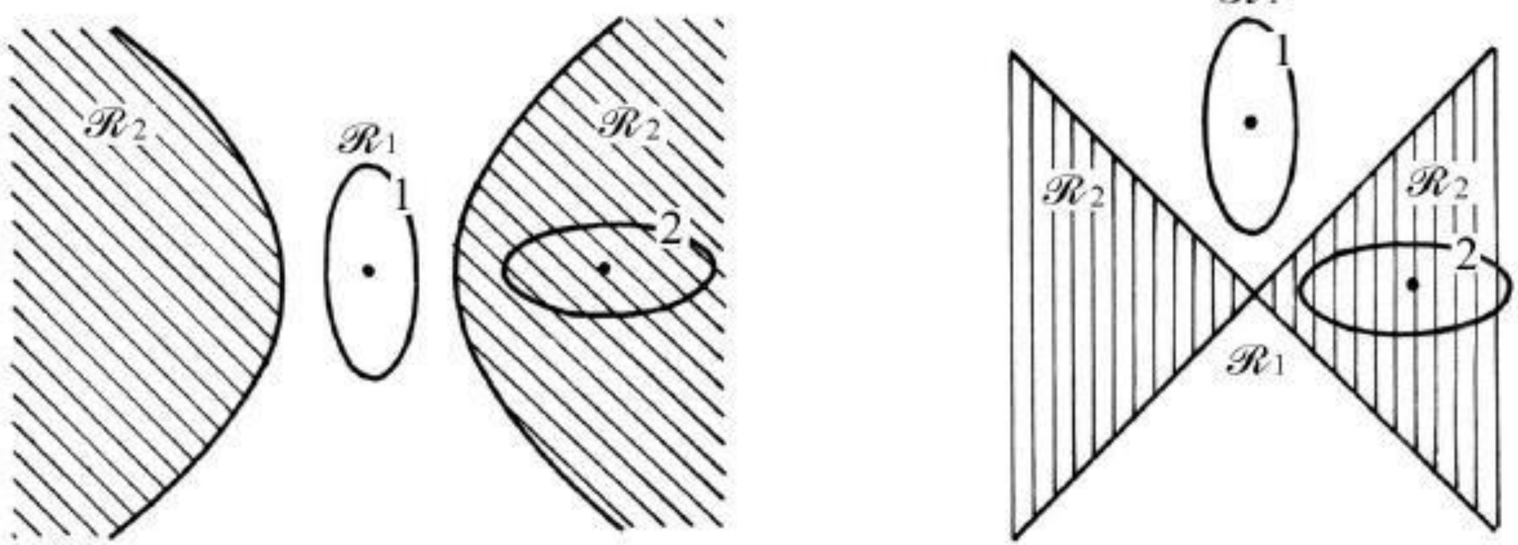

x_2
$\mathscr{R}_i$
$\mathscr{R}_j$
μ_i
μ_j
x_1

第三种情况：各类的协方差阵不相等

- 请看图



(a) 圆　　(b) 椭圆　　(c) 抛物线



(d) 双曲线　　(e) 直线

- 关于分类器的错误率问题
- 当分类器确定后，其错误率也就确定了。
- 三种方法：

  按照公式计算

  计算错误率的上界

  实验估计