模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 2 (Part 2): Bayesian Decision Theory

(Sections 2.3-2.5)

本讲内容

- 最小化错误率分类
- 分类器, 判别函数，及决策面
- 正态密度

最小化错误率分类
Minimum-Error-Rate Classification

行为是在类别决定之后的决策

若采取行为 α_i，而真实状态是 ω_j 则:

若 $i = j$，决策正确；相反，

若 $i \neq j$，决策错误。

目标是找到一个决策规则，可以最小化错误概率（*probability of error*，或称错误率 *error rate*），

0-1损失函数（zero-one loss function）

- 考虑下面的对称0-1损失函数:

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

由此，条件风险为:

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

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

相应于该损失函数的风险是平均概率误差

0-1损失（Zero-One Loss）

- 最小化风险要求最大化 $P(\omega_i \mid x)$

  因为 $R(\alpha_i \mid x) = 1 - P(\omega_i \mid x)$

- 对于最小化错误率

  若$P(\boldsymbol{\omega_i} \mid x) > P(\boldsymbol{\omega_j} \mid x) \quad \forall j \neq i$，则决策为$\boldsymbol{\omega_i}$ 。

0-1损失

0-1函数的决策区域

$$Let\ \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\cdot\frac{P(\omega_2)}{P(\omega_1)}=\theta_\lambda\ then\ decide\ \omega_1\ \ if:\ \frac{P(x|\omega_1)}{P(x|\omega_2)}>\theta_\lambda$$

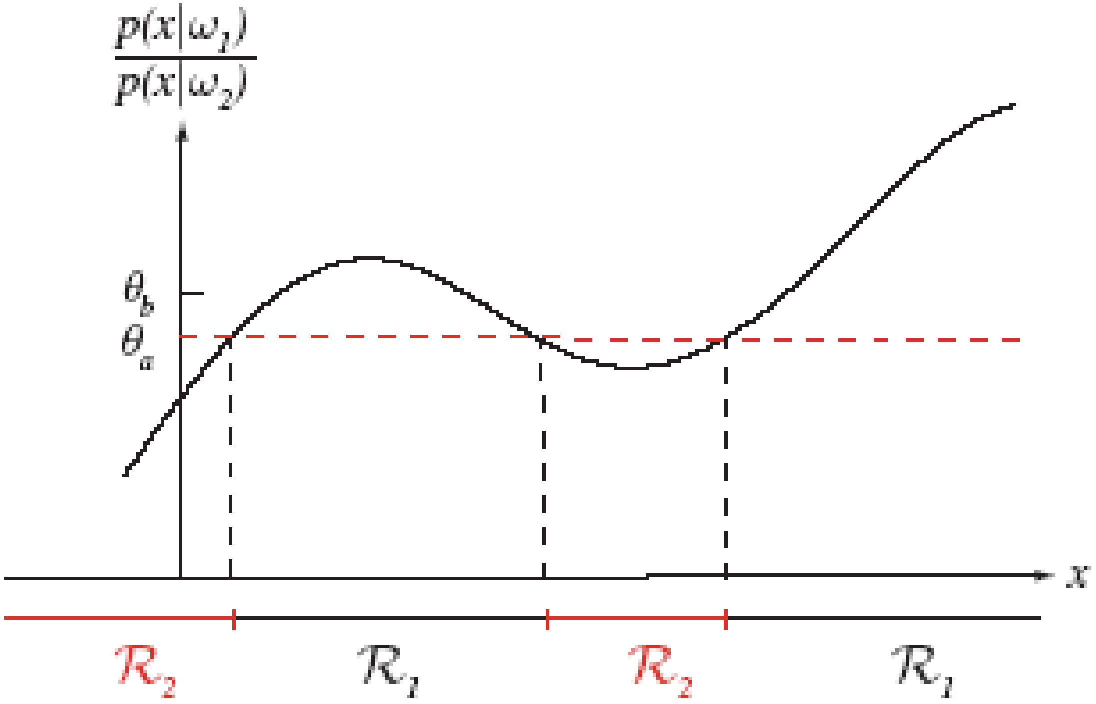
若λ是0-1损失函数，即意谓着:

$$\lambda=\begin{pmatrix}0&1\\1&0\end{pmatrix}$$

$$then\ \ \theta_\lambda=\frac{P(\omega_2)}{P(\omega_1)}=\theta_a$$

$$if\ \ \lambda=\begin{pmatrix}0&2\\1&0\end{pmatrix}then\ \ \theta_\lambda=\frac{2P(\omega_2)}{P(\omega_1)}=\theta_b$$

最小损失似然率决策示意图



图Fig. 2.1显示出了似然率$p(x|\omega_1)/p(x|\omega_2)$的分布。

- 若采用0-1损失分类, 决策边界由阈值θ_a确定。若将ω_2误分类为ω_1时的损失比将ω_1误分类为ω_2时的损失大，可以通过将阈值调整为较大的θ_b，这样可使R_1的区间变小。

模式分类器：Pattern Classifier

如何表示一个模式分类器?

最常用的方法是通过判别函数（**discriminant functions**）：
若用 $\{\omega_1,\omega_2,\cdots,\omega_c\}$ 标记可能的类别状态.

对每一个类，建立一个判别函数 $g_i(x)$

分类器是将样本指派一个到类 ω_i，若

$$g_i(x) > g_j(x) \quad for\ all \quad j \neq i$$

- 因此，分类器（**classifier**)是可以计算个c 个判别函数的一个网络（**network** ），或者一个机器（ **machine** ）

判别函数：Discriminant Functions

- 对多类别情况

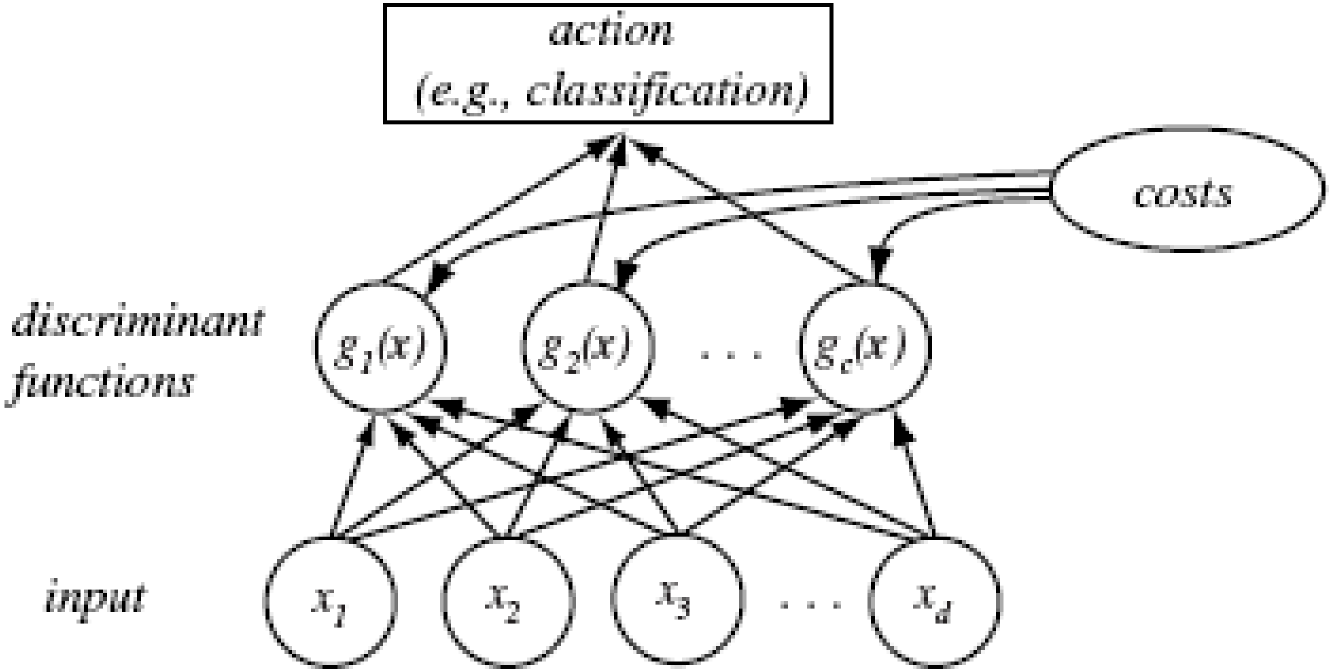
判别函数集合：

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

分类器将特征向量 $\boldsymbol{x}$ 指派到类 $\boldsymbol{\omega_i}$ 若:

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

模式分类器



具有d个输入和c个判别函数$g_i(\mathbf{x})$的统计模式分类其的函数结构

- 后续的步骤是对最大判决值进行决策，并将输入分类到对应模式类中。箭头表示了信息流的方向。

判别函数的简化

注意：假设 $f(\cdot)$ 是一个单调增函数(monotonically increasing function)，则将所有 $g_i(x)$ 替换为 $f(g_i(x))$ 时，可以得到相同的决策结果。

对应地:

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

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

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

(这里 ln 是自然对数函数)

- 分类器网络的作用是将特征空间划分为 c 个区域（每个类对应一个区域）. 称这些区域为决策区域（***decision regions***），区域边界称为决策边界（***decision boundaries***）。

判别公式：Discriminant

- 令 $g_i(x) = - R(\alpha_i \mid x)$, 对应与最小错误率

  (此时， 最大判别对应于最小风险!)

- 令 $g_i(x) = P(\omega_i \mid x)$

  (此时， 最大判别对应于最大后验概率!)

$$g_i(x) \equiv P(x \mid \omega_i)\, P(\omega_i)$$

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

判别公式

- 将特征空间划分为 c 个决策区域

若 $g_i(x) > g_j(x) \ \forall j \neq i$ 则 x 在区域 R_i 中

(R_i 表示将 x 指定为 ω_i 类)

两类的情况

此时的分类器成为二分器（**dichotomizer**），有两个判决函数 g_1 及 g_2

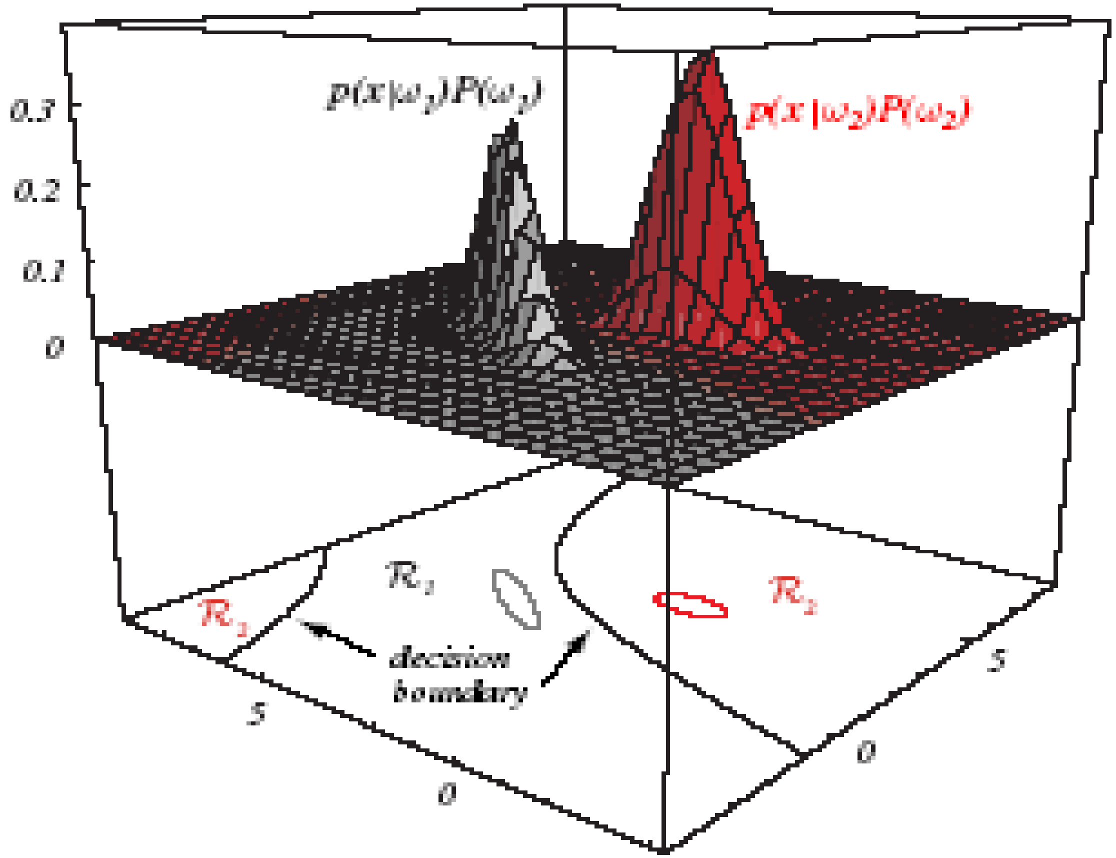
令 $g(x) \equiv g_1(x) - g_2(x)$

若$g(x) > 0$，决策为 ω_1；否则，决策为 ω_2

决策面：Decision Surfaces

$g(x)$的计算

$$g(x) = P(\omega_1 \mid x) - P(\omega_2 \mid x)$$
$$= \ln \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$



- 在图中所示的2维两类分类器中，概率密度是Gaussian型的，决策边界由两条双曲线构成。这时，决策区域 R_2 不是单联通的，图中标记的椭圆是密度为峰值分布的1/ e倍时的地方。

正态密度（The Normal Density）

回顾Bayes规则，一个决策函数需要计算两个概率密度:

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

如何建立条件概率密度函数模型?

大多数时候将其假设为一个多元正态密度(**Multivariate normal density**)。
单变量正态函数为:

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

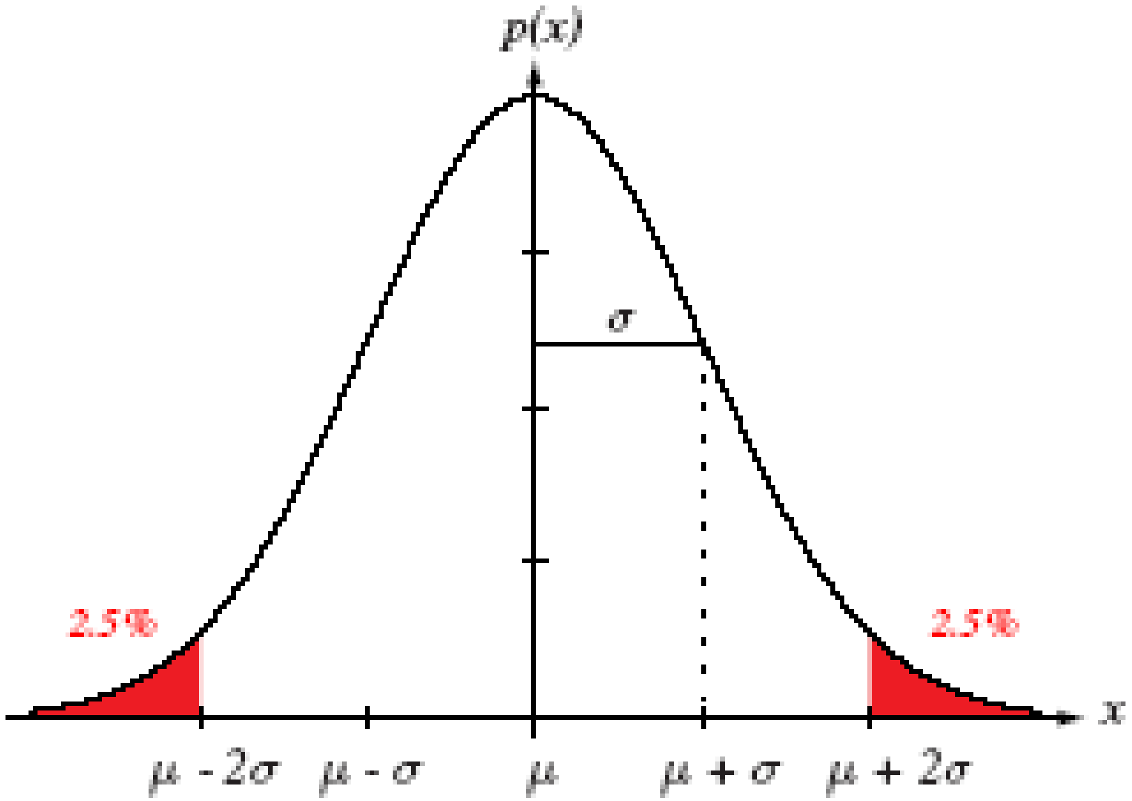
$$\mu \equiv E[x] = \int_{-\infty}^{\infty} xp(x)dx$$

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

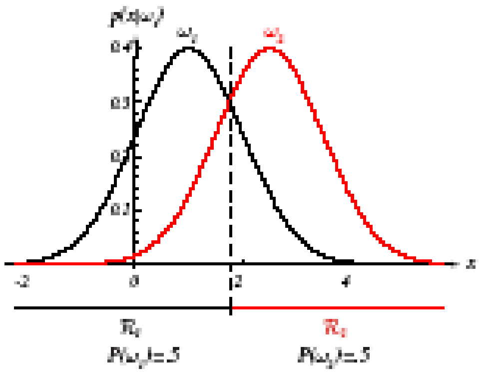
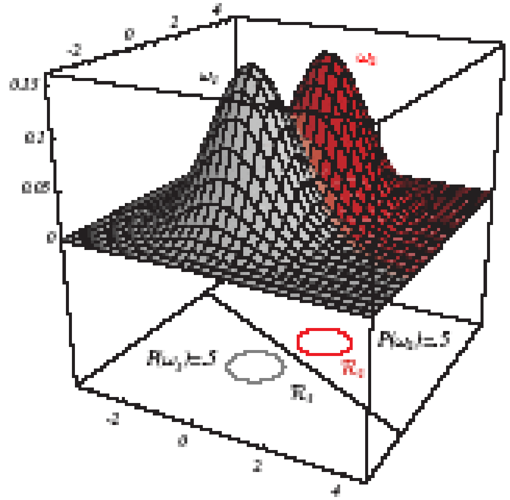
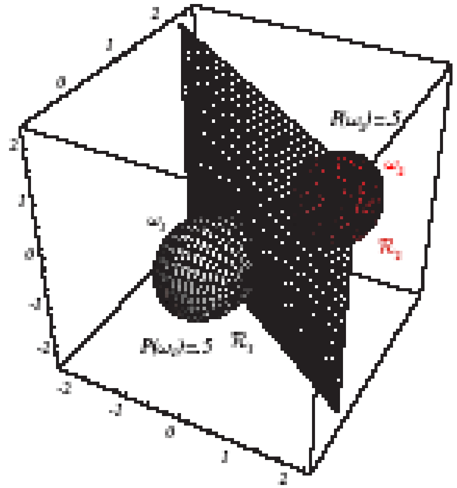
这里:

$\mu = x$的均值

$\sigma^2 = x$的方差



- 如图所示单变量正态分布中，95%的面积处在$|x-\mu|\le 2\sigma$范围之中。分布的峰值为$p(\mu)=1/\sqrt{2\pi}\sigma$.

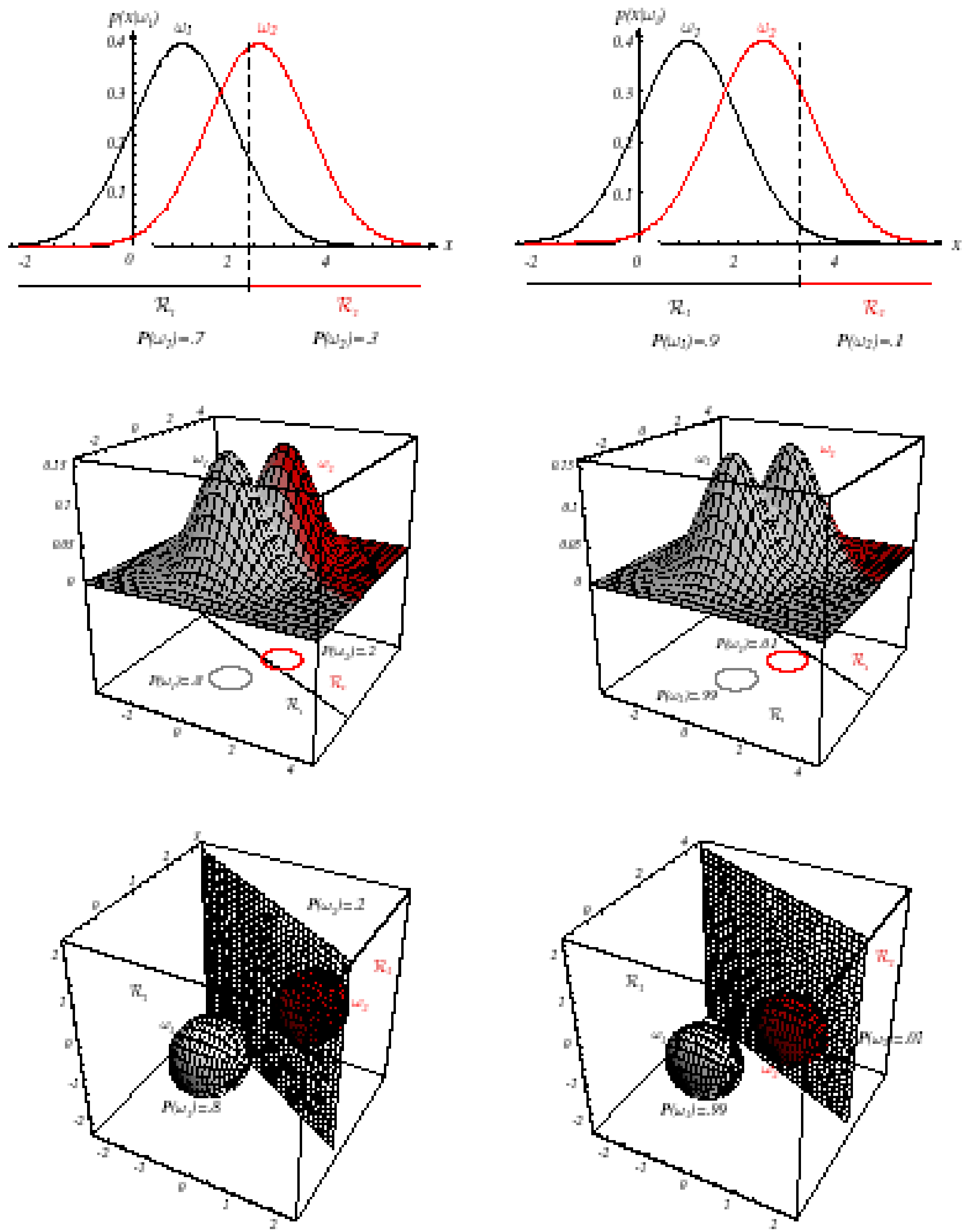
  

- 若两个分布的协方差矩阵相等且均为数量矩阵，则它们在d维空间的分布是球形，而边界是d−1维的广义超平面, 并且垂直于两个分布的均值之间的连线，位于该连线的中间位置。
- 在上图1、2、3维的例子中，画出的是在$P(\omega_1) = P(\omega_2)$条件下的$p(\mathbf{x}|\omega_i)$及分类边界 , 在3维情况下，网格面将$R1$和$R2$划分开来。

当先验概率变化时，决策边界发生平移;

在先验概率差异足够大时，在1、2、3维高斯球分布的分类边界将不位于分布的均值之间.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. Copyright c_ 2001 by John Wiley & Sons, Inc.



信号检测（Signal Detection）理论

假设，我们将从一个信号中检测出一个单脉冲。并且假设信号受到了随机噪声的干扰。

当脉冲出现时，观察到的测量值服从均值为 μ_2 的正态分布

当信号不出现时，观察到的测量值服从均值为 μ_1 的正态分布

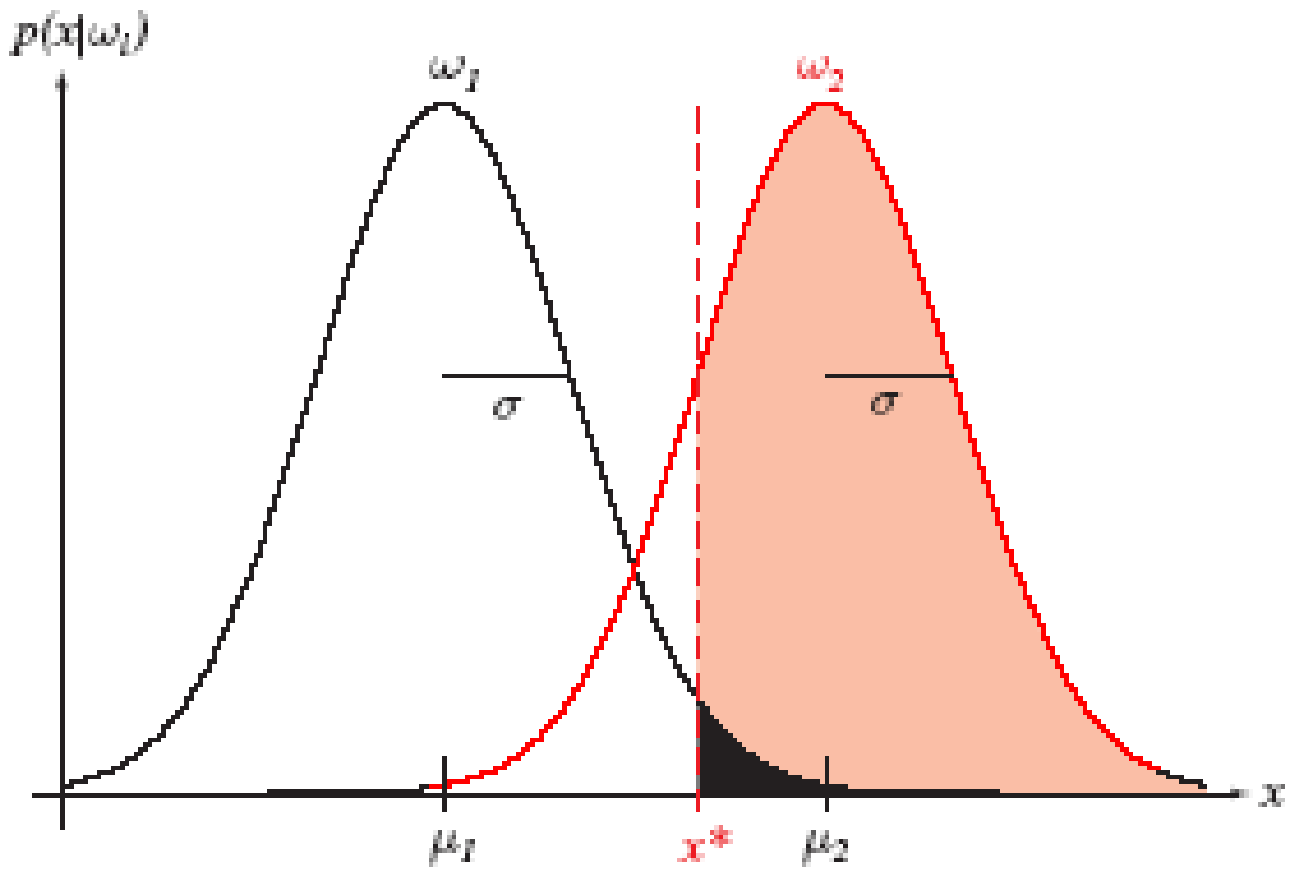
假设上述正态分布具有相同的方差.

我们能否对判别能力进行度量？

该度量是否独立于判别阈值 x^* ？

判别能力:

$$d' = \frac{|\mu_2 - \mu_1|}{\sigma}$$



- 在没有外部脉冲出现的任何瞬时时刻, 内部信号的概率密度是正态的，即, p($x|\omega_1$) ~ N(μ_1, σ_2); 而当外部脉冲出现时， 该密度为 p(x|$\omega 2$) ~ N($\mu 2$, $\sigma 2$).
- 决策阈值 $x*$ 决定了检测（a hit）概率(大于$x*$的曲线ω_2下的粉色面积) 以及虚警（a false alarm ）概率 (小于$x*$的曲线ω_1下的黑色面积).

信号检测理论

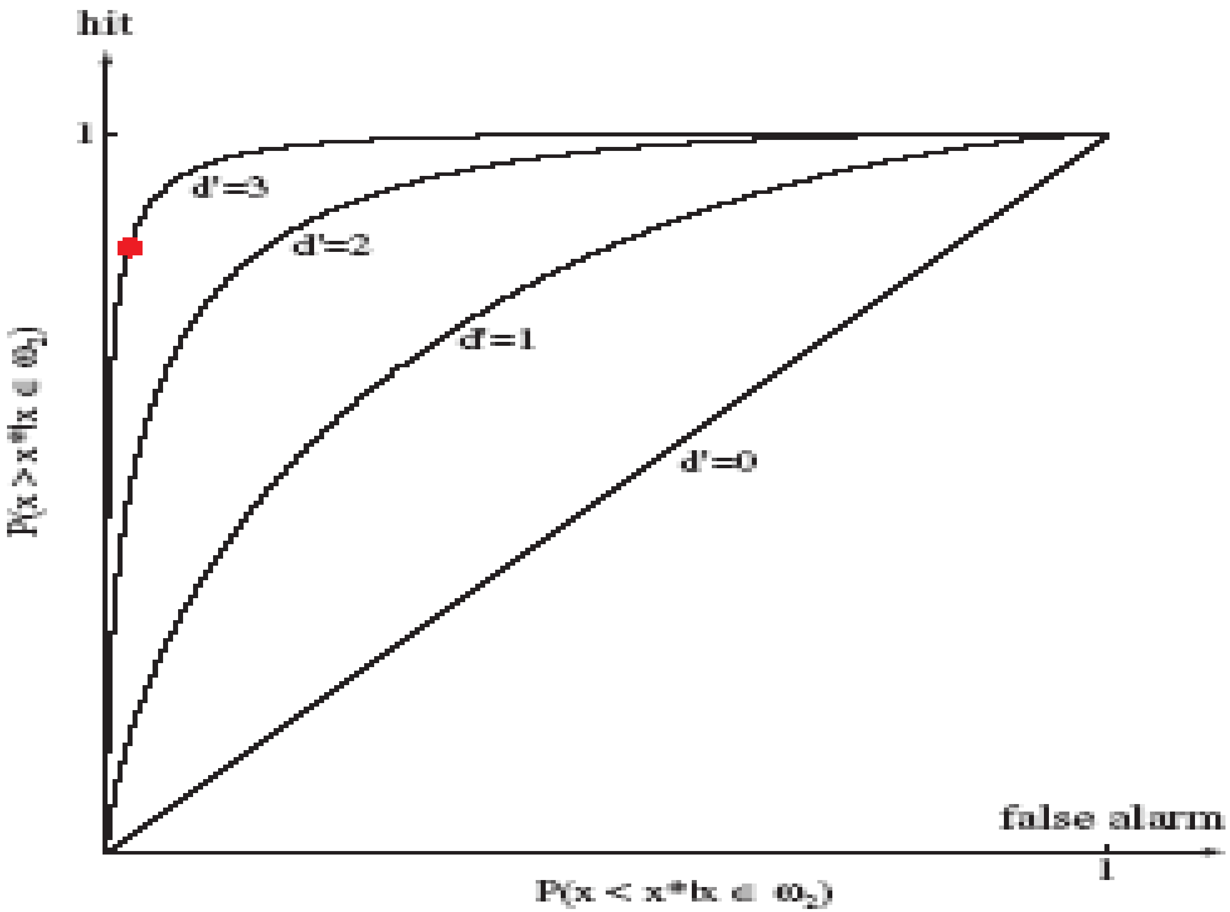
在 μ_1 ，μ_2 ，或 x^* 未知时，如何确定概率 *P*?

由数据可算出:

1. $P\left(x > x^* \middle| x \in \omega_2\right)$ 检测概率（a hit）.
2. $P\left(x > x^* \middle| x \in \omega_1\right)$ 虚警概率（a false alarm）.
3. $P\left(x < x^* \middle| x \in \omega_2\right)$ 漏检概率（a miss）.
4. $P\left(x < x^* \middle| x \in \omega_1\right)$ 正确拒识概率（a correct rejection）.

- 做出检测率（hit）和虚警率（false alarm）的比率曲线，称该曲线为接收机工作特性曲线（ROC，receiver operating characteristic）

利用ROC，可以有效区分判别能力（discriminability）和偏差（bias）.



- 在ROC曲线中, 横坐标是虚警率, $P(x>x*|\ x\in\omega_1)$, 而纵坐标是检测率, $P(x>x*|\ x\in\omega_2)$.
- 通过测量检测率和虚警率的比率，(这里对应于上页中的$x*$，图中的红点), 可以推导出 d = 3.