模式识别第二章贝叶斯决策理论

郭园方北京航空航天大学计算机学院

内容

- *引言
- * 几种常见的决策规则
- * 正态分布时的统计决策

引言

- * 贝叶斯决策理论是统计模式识别中最基本的方法,利用此方法进行分类时要求:
 - * 各类别总体的概率分布是已知的
 - * 待决策分类的类别数是一定的
- *统计决策理论——根据每一类总体的概率分布决定决策边界

*在连续的情况下,假设对要识别的物理对象 有d种特征观察量

$$X_1, X_2, ..., X_d$$

* 这些特征的所有可能的取值范围构成了d维空间 $x = [x_1, x_2, ..., x_d]^T$

* 我们又称x为d维特征向量

- * 假设要分类的问题有 \mathbf{c} 个类别,各类别状态用 ω_i 来表示,其中 i=1,2,...,c
- * 且对应于各类别的 ω_i 出现的先验概率 $P(\omega_i)$ 及类条件概率密度 $p(x|\omega_i)$ 已知
- *如果在特征空间已经观察到某一个向量x,应该把x分到哪一类?

- *例: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。(两分类问题)
- *根据以往医生的经验知道:
 - * 患病的人, 白细胞的浓度与正常人不同
 - *一般人群中,患病的人数比例为0.1。
- *如果已知一个人的白细胞浓度, 医生应该做出怎样的判断?

- *对该问题的数学表示如下:
- *用 Ω 表示"类别"这一随机变量, ω_1 表示正常, ω_2 表示患病
- *X表示"白细胞浓度"这个随机变量
- * X表示浓度值

* 医生的先验概率:

$$P(\omega_1) = 0.9$$

$$P(\omega_2) = 0.1$$

*观测数据白细胞浓度分别在两种情况下的类条件分布: $p(x|\omega_1)$

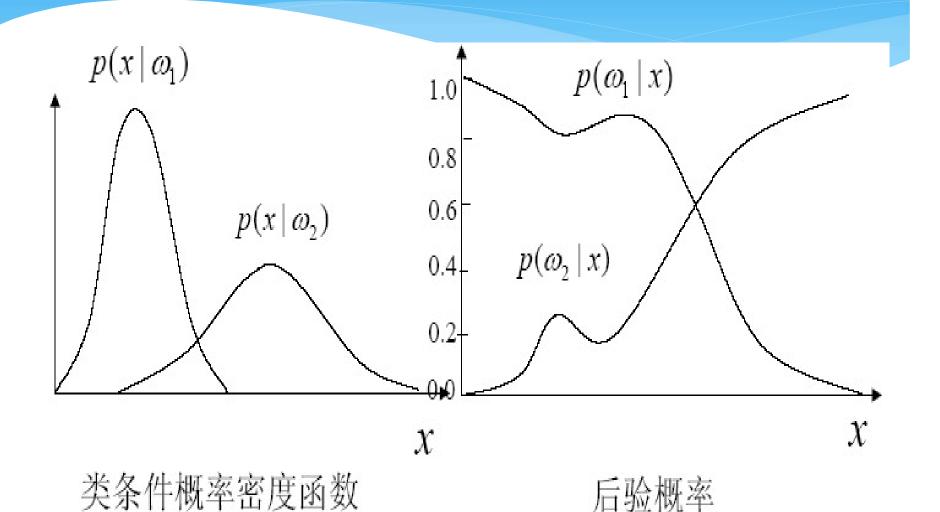
$$p(x \mid \omega_2)$$

- *决策:根据观测到的x,利用先验概率和类条件概率,决定x属于哪一类
- *决策是从样本空间到决策空间的一个映射
- * 评价决策有多种标准,对于同一个问题,采 用不同的标准会得到不同意义下"最优"的 决策。

- * Bayes决策是所有识别方法的一个基准
- * Bayes决策两种常用的准则:
 - * 最小错误概率准则
 - * 最小风险准则

*前面的问题,已知先验概率和类条件概率密度,利用贝叶斯公式,求得后验概率

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



- *决策规则
- * (1) $P(\omega_i|x) = max_{j=1,2}P(\omega_j|x), x \in \omega_i$

- * 其他等价形式
- * (2) $p(x|\omega_i)P(\omega_i) = max_{j=1,2}P(x|\omega_j)P(\omega_i)$

* (3) 若
$$l(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} \ge \frac{P(\omega_2)}{P(\omega_1)}$$
则 $x \in \frac{\omega_1}{\omega_2}$

*(4)对上式取自然对数的负值,可写为 $h(x) = -\ln(l(x))$ $= -\ln p(x|\omega_1) + \ln p(x|\omega_2) \ge \ln \frac{P(\omega_1)}{P(\omega_2)}$ 则 $x \in \frac{\omega_1}{\omega_2}$

- * 例题: 前面讲到的白血病检验问题
- *现有一待识别细胞,观察值为x,从类条件概率密度分布曲线上查得

$$p(x|\omega_1) = 0.2 p(x|\omega_2) = 0.4$$

*对该细胞进行分类

*利用贝叶斯公式,分别计算后验概率

$$P(\omega_{1}|x) = \frac{p(x | \omega_{1})P(\omega_{1})}{\sum_{j=1}^{2} p(x | \omega_{j})P(\omega_{j})}$$

$$= \frac{0.2 * 0.9}{0.2 * 0.9 + 0.4 * 0.1} = 0.818$$

$$P(\omega_{2}|x) = 1 - P(\omega_{1}|x) = 0.182$$

$$P(\omega_{1}|x) > P(\omega_{2}|x)$$

* 故该细胞是正常的

- *下面以一维情况证明上述贝叶斯决策规则使错误率最小
- *错误率:此处为平均错误率
- * $P(e) = \int_{-\infty}^{+\infty} P(e, x) dx = \int_{-\infty}^{+\infty} P(e|x) p(x) dx$

- 条件错误率 P(e|x) 的计算:
 - 以两类问题为例,当获得观测值 x 后,有两种决策可能:决定 $x \in \omega_1$,或者 $x \in \omega_2$. 条件错误率为:

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

- Bayes最小错误率决策:
 - 选择后验概率 $P(\omega_1|x)$, $P(\omega_2|x)$ 中大的 ω 作为决策,使得在观测值x下的条件错误率最小。

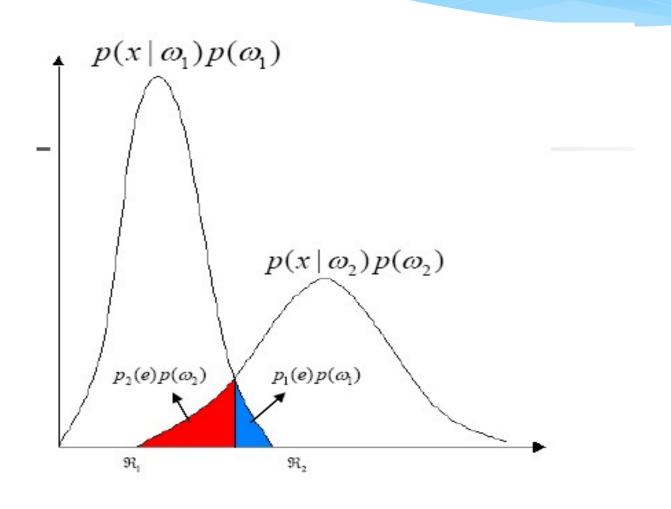
$$D(x) = \arg\max_{i} P(\omega_i \mid x).$$

■ 条件错误率:

$$P(e \mid x) = 1 - \max_{i} P(\omega_{i} \mid x).$$

■ 错误率:

$$P(e) = E(P(e \mid x)).$$



$$p(e) = p(x \in \Re_1, \omega_2) + p(x \in \Re_2, \omega_1)$$

$$== p(x \in \Re_1 \mid \omega_2) p(\omega_2) + p(x \in \Re_2 \mid \omega_1) p(\omega_1)$$

$$= p(\omega_2) p_2(e) + p(\omega_1) p_1(e)$$

- *此决策实际上对每个x都使P(e|x)取最小
- * 从而保证上式最小
- *由此证明了最小错误率贝叶斯规则确实使错误率最小

Bayes最小错误率决策不仅保证了错误率 (条件错误率的期望)最小,而且保证 每个观测值下的条件错误率最小,Bayes 决策是一致最优决策。

多类识别问题的Bayes最小错误率决策:

在观测值x下的每个决策的条件错误率为:

$$P(e \mid x) = 1 - P(\omega_i \mid x)$$
 若决定 $x \in \omega_i$

■ 决策为:

$$D(x) = \arg\max_{i} P(\omega_{i} \mid x).$$

后验概率 $P(\omega_i|x)$ 的计算:

■ Bayes公式:

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

$$= \frac{p(x \mid \omega_i)P(\omega_i)}{\sum_i p(x \mid \omega_i)P(\omega_i)}.$$

■ $P(\omega_i)$: 先验概率; $p(x|\omega_i)$: 类条件概率密度。

$$P(e) = [P(\vec{x} \in \Re_{2} \mid \omega_{1}) + P(\vec{x} \in \Re_{3} \mid \omega_{1}) + ... + P(\vec{x} \in \Re_{c} \mid \omega_{1})]P(\omega_{1}) + [P(\vec{x} \in \Re_{1} \mid \omega_{2}) + P(\vec{x} \in \Re_{3} \mid \omega_{2}) + ... + P(\vec{x} \in \Re_{c} \mid \omega_{2})]P(\omega_{2}) + ... + [P(\vec{x} \in \Re_{1} \mid \omega_{c}) + P(\vec{x} \in \Re_{2} \mid \omega_{c}) + ... + P(\vec{x} \in \Re_{c-1} \mid \omega_{c})]P(\omega_{c}) + ... + P(\vec{x} \in \Re_{c-1} \mid \omega_{c})]P(\omega_{c})$$

$$= \frac{1}{4\pi} e^{-1\pi}$$

$$= \sum_{i=1}^{c} \sum_{j=1, j\neq i}^{c} [P(\vec{x} \in \mathfrak{R}_{j} \mid \omega_{i})] P(\omega_{i})$$

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

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

*每个决策的风险是不同的,我们在决策过程中应该考虑决策风险

* 以疾病诊断为例, 风险大不相同

- *如何在决策中考虑风险?
- *引入风险函数(损失函数),通常是决策和自然状态的函数,可用决策表来表示

■ N类问题: (λ_{i,j})_{N*N}

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & \lambda_{14} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & & \cdots & & \lambda_{2n} \\ & & \ddots & & & & \\ \vdots & & \lambda_{ij} & & \vdots \\ \lambda_{(n-1)1} & & & & \lambda_{n1} \\ \lambda_{n1} & \lambda_{n2} & & \cdots & & \lambda_{nn} \end{pmatrix}$$
 表示当我们给出的决策为ip的损失值。

- * 以数学形式表示决策问题:
- *(1)观察x是d维随机向量

$$x = [x_1, x_2, ..., x_d]^T$$

*(2)状态空间Ω由c个自然状态组成

$$\Omega = [\omega_1, \omega_2, ..., \omega_c]^T$$

- *(3) 决策空间由a个决策组成 $A = \{a_1, a_2, ..., a_a\}$
- *(4)损失函数为\lambda

*根据贝叶斯公式,

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

* 其中

$$p(x) = \sum_{i=1}^{c} p(x|\omega_i)P(\omega_i)$$

- * 在引入损失函数后,对应于决策 a_i , λ 可以在c $\Lambda(a_i, \omega_j)$, j = 1, 2, ..., c 中任意选取一个,相应的概率为 $P(\omega_i|x)$
- *故采用决策ai时的条件期望损失是
- * $R(a_i|x) = E[\lambda(a_i, \omega_j)] =$ $\sum_{j=1}^{c} \lambda(a_i, \omega_j) P(\omega_j|x), i = 1, 2, ..., a$

*X是随机变量,采用X不同的观测值,产生的条件风险不同,决策a可以看成是X的函数, 我们定义期望风险

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

*条件风险对应的是x,期望风险对应的是a(x)

- * 在考虑错判带来的损失时,我们希望损失最小,如果在采取每个决策,都使其条件风险最小,则对所有的x作出决策时,其期望风险也必然最小。这样的决策就是最小风险贝叶斯决策。
- * 最小风险贝叶斯决策规则为:

如果
$$R(a_k|x) = min_{i=1,...,a}R(a_i|x)$$
 则 $a = a_k$

- * 具体计算步骤如下:
- *(1)计算后验概率
- *(2) 根据决策风险表,计算出采取 a_i 的条件风险 $R(a_i|x) = E[\lambda(a_i,\omega_j)]$ $= \sum_{i=1}^{c} \lambda(a_i,\omega_j) P(\omega_j|x), i = 1,2,...,a$

*(3)对a个条件风险值进行比较,找出使条件风险最小的决策a

$$R(a_k|x) = \min_{i=1,\dots,a} R(a_i|x)$$

*则ak就是最小风险贝叶斯决策

- * 以前面例题为例
- * 已知:

$$P(\omega_1) = 0.9, P(\omega_2) = 0.1$$

 $P(x|\omega_1) = 0.2, P(x|\omega_2) = 0.4$
 $\lambda_{11} = 0, \lambda_{12} = 6$
 $\lambda_{21} = 1, \lambda_{22} = 0$

*利用贝叶斯公式,分别计算后验概率

$$P(\omega_{1}|x) = \frac{p(x | \omega_{1})P(\omega_{1})}{\sum_{j=1}^{2} p(x | \omega_{j})P(\omega_{j})}$$

$$= \frac{0.2 * 0.9}{0.2 * 0.9 + 0.4 * 0.1} = 0.818$$

$$P(\omega_{2}|x) = 1 - P(\omega_{1}|x) = 0.182$$

$$P(\omega_{1}|x) > P(\omega_{2}|x)$$

*引入风险函数后,计算条件风险

$$R(a_1|x) = \sum_{j=1}^{c} \lambda_{1j} P(\omega_j|x) = \lambda_{12} P(\omega_2|x)$$

= 1.092

$$R(a_2|x) = \lambda_{21}P(\omega_1|x) = 0.818$$

- * 因为 $R(a_1|x) > R(a_2|x)$
- *故选择决策行动 a2 即该细胞为异常

*小作业:最小错误率的贝叶斯决策和最小风险贝叶斯决策的关系? (证明)

几种常见的决策规则分类器设计

- *应用最小错误率贝叶斯决策准则的分类器设计
- * 分类器设计问题为确定判别函数和决策面
- *对于c类分类问题,按照决策规则可以把d维特征空间分成c个决策域,我们将划分决策域的边界面称为决策面,在数学上用解析形式可以表示成决策面方程,用于表达决策规则的某些函数称为判别函数。

几种常见的决策规则 多类情况的贝叶斯决策规则

- * (1) $P(\omega_i|x) > P(\omega_j|x), j = 1, 2, ..., c, j \neq i$ $\text{ If } x \in \omega_i$
- * (2) $p(x \mid \omega_i)P(\omega_i) > p(x \mid \omega_j)P(\omega_j), j = 1, 2, ..., c, j \neq i, \quad \emptyset \mid x \in \omega_i$

几种常见的决策规则 多类情况的贝叶斯决策规则

* (3)
$$l(x) = \frac{p(x|\omega_i)}{p(x|\omega_j)} > \frac{P(\omega_j)}{P(\omega_i)}, j = 1, 2, ..., c, j \neq i$$
, $\lim_{x \to \omega_i} x \in \omega_i$

* (4)
$$\ln \mathbf{p}(x|\omega_i) + \ln \mathbf{P}(\omega_i) > \ln \mathbf{p}(x|\omega_j) + \ln \mathbf{P}(\omega_j), j = 1, 2, ..., c, j \neq i, \quad \emptyset \mid x \in \omega_i$$

几种常见的决策规则判别函数

- * 判別函数: 通常定义一组判别函数 $g_i(x)$, i = 1, 2, ..., c, 用来表示多类决策规则:
- * 如果使得 $g_i(x) > g_j(x)$ 对于一切的 $j \neq i$ 均成立,则将x归于 ω_i 类。

几种常见的决策规则判别函数

- *相对应于贝叶斯决策的判别函数
- * (1) $g_i(x) = P(\omega_i|x)$
- * (2) $g_i(x) = p(x \mid \omega_i)P(\omega_i)$
- * (3) $g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$

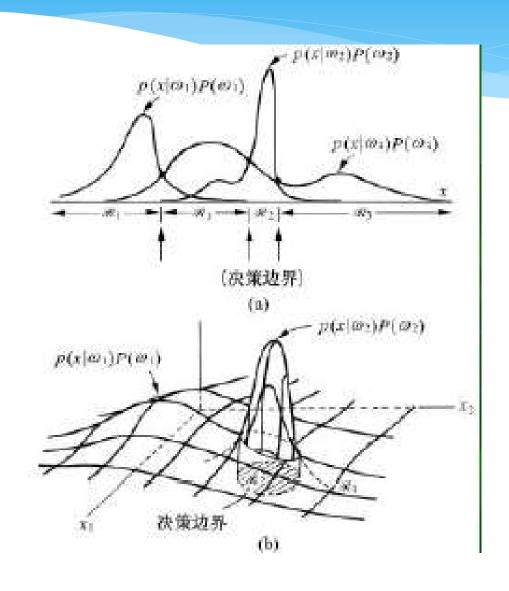
几种常见的决策规则 决策面方程&分类器设计

*决策面方程

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

- * 分类器设计原则
 - * 选取最大的g(x)

几种常见的决策规则 决策面方程&分类器设计



正态分布时的统计决策

- * 在统计决策分析中经常假定正态分布, 因为:
 - * 物理上的合理性
 - * 正态分布的一些性质使在数学上计算简单

正态分布时的统计决策

- * 物理上的合理性示例
 - *人的身高、寿命等
 - *考试成绩
 - * 医学参考值
 - *信道噪声

*单变量正态分布

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

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

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

*概率密度函数应满足下面关系:

$$p(x) \ge 0(-\infty < x < +\infty)$$
$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

* 多元正态分布
$$p(x)$$
 = $\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$ 其中 $x = [x_1, x_2, ..., x_d]^T$ 是d维列向量, $\mu = [\mu_1, \mu_2, ..., \mu_d]^T$ 是d维均值向量, Σ 是d*d维的协方差矩阵

$$\mu = E\{x\}$$

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

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

$$\sigma_{11}^2 \quad \sigma_{12}^2 \quad \sigma_{1d}^2$$

$$\Sigma = \sigma_{21}^2 \quad \sigma_{22}^2 \quad \sigma_{2d}^2$$

$$\sigma_{d1}^2 \quad \sigma_{d2}^2 \quad \sigma_{dd}^2$$

*多元正态概率模型下的最小错误率贝叶斯判别函数和决策面

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

* 正态分布时判别函数为

$$g_{i}(x)$$

$$= \frac{-1}{2} (x - \mu_{i})^{T} \Sigma_{i}^{-1} (x - \mu_{i}) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_{i}|$$

$$+ \ln \mathbf{P}(\boldsymbol{\omega}_{i})$$

*决策面方程

$$g_{i}(x) = g_{j}(x)$$

$$\frac{-1}{2} [(x - \mu_{i})^{T} \Sigma_{i}^{-1} (x - \mu_{i})$$

$$- (x - \mu_{j})^{T} \Sigma_{j}^{-1} (x - \mu_{j})] - \frac{1}{2} \ln \frac{|\Sigma_{i}|}{|\Sigma_{j}|} + \ln \frac{\mathbf{P}(\boldsymbol{\omega}_{i})}{\mathbf{P}(\boldsymbol{\omega}_{j})}$$

$$= \mathbf{0}$$

*特殊情况1:

$$\Sigma_i = \sigma^2 I$$
, $i = 1, 2, ..., c$

- *每类协方差矩阵相等,类内各特征相互独立,方差相等
- * 若先验概率 $P(\omega_i) \neq P(\omega_i)$

$$\Sigma_i = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \dots & \sigma^2 & \dots, |\Sigma_i| = \sigma^{2d}, \Sigma_i^{-1} = \frac{1}{\sigma^2} I \\ 0 & \dots & \sigma^2 \end{bmatrix}$$

*判别函数为:

$$g_{i}(x)$$

$$= \frac{-1}{2\sigma^{2}}(x - \mu_{i})^{T}(x - \mu_{i}) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^{2d}$$

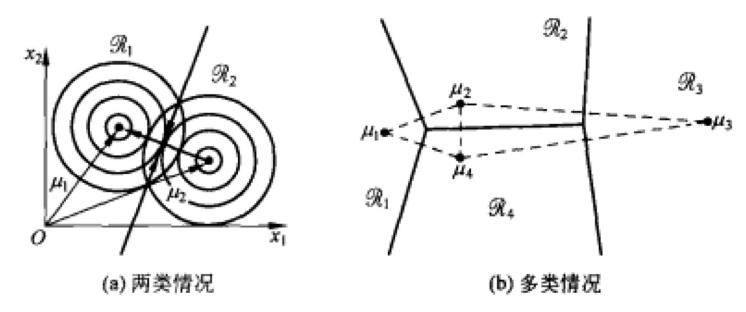
$$+ \ln \mathbf{P}(\boldsymbol{\omega}_{i})$$

$$\mathbb{P}g_{i}(x) = \frac{-1}{2\sigma^{2}}(x - \mu_{i})^{T}(x - \mu_{i}) + \ln \mathbf{P}(\boldsymbol{\omega}_{i})$$

* 若先验概率
$$\mathbf{P}(\boldsymbol{\omega_i}) = \mathbf{P}(\boldsymbol{\omega_j})$$

$$\boldsymbol{g_i}(\boldsymbol{x}) = \frac{-1}{2\sigma^2} (\boldsymbol{x} - \mu_i)^T (\boldsymbol{x} - \mu_i)$$

等价于最小距离分类器



正态分布且 $P(\omega_i) = P(\omega_j), \Sigma_i = \sigma^i I$ 时的决策面

*特殊情况2:

$$\Sigma_i = \Sigma, i = 1, 2, ..., c$$

* 每类协方差矩阵相等

$$g_{i}(x)$$

$$= \frac{-1}{2\sigma^{2}}(x - \mu_{i})^{T}(x - \mu_{i}) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^{2d}$$

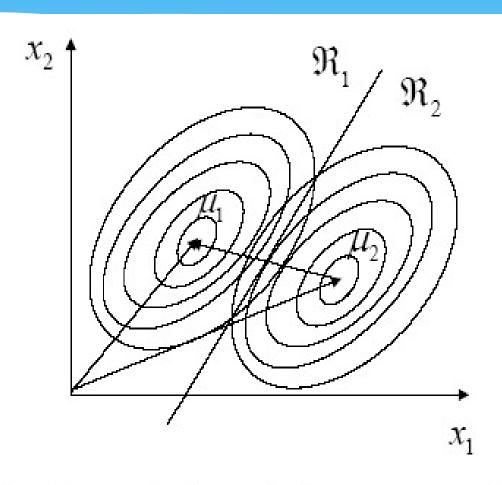
$$+ \ln \mathbf{P}(\boldsymbol{\omega}_{i})$$

可简化为
$$\mathbf{g}_i(\mathbf{x}) = \frac{-1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i) + \ln \mathbf{P}(\boldsymbol{\omega}_i)$$

* 当先验概率相等时

$$g_i(x) = \gamma^2 = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)$$

即为马氏距离



正态分布且 $P(\omega_1) = P(\omega_2), \Sigma_1 = \Sigma_2$ 时的决策面

*特殊情况3:

$$\Sigma_i \neq \Sigma_j$$

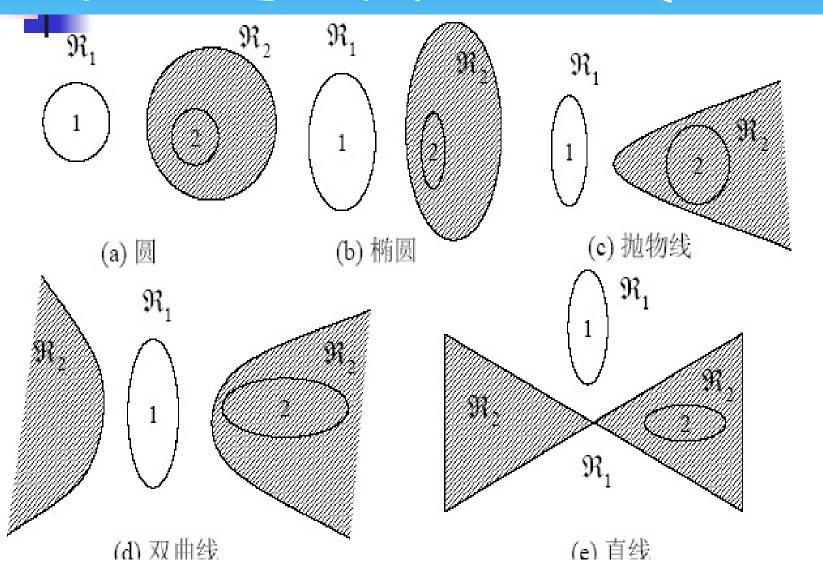
* 各类协方差矩阵不相等

$$g_{i}(x)$$

$$= \frac{-1}{2} (x - \mu_{i})^{T} \Sigma_{i}^{-1} (x - \mu_{i}) - \frac{1}{2} \ln |\Sigma_{i}|$$

$$+ \ln \mathbf{P}(\boldsymbol{\omega}_{i})$$

* 决策面: $g_i(x) - g_j(x) = 0$



谢谢