模式识别-复习课
Pattern Recognition

2021.5

概　念

- 模式识别：确定一个样本的类别属性（模式类）的过程，即把某一样本归属于多个类型中的某个类型。模式分类的过程。
- 样本（Sample)：一个具体的研究（客观）对象。如患者，某人写的一个汉字，一幅图片等。
- 模式（Pattern）：对客体（研究对象）特征的描述（定量的或结构的描述），是取自客观世界的某一样本的测量值的集合（或综合）。
- 特征（Features）：能描述模式特性的量（测量值）。

  在统计模式识别方法中，通常用一个矢量表示，称之为特征矢量，记为 $\vec{x}=(x_1,x_2,\cdots,x_n)'$
- 模式类（Class）：具有某些共同特性的模式的集合。

1.1.3 模式识别系统

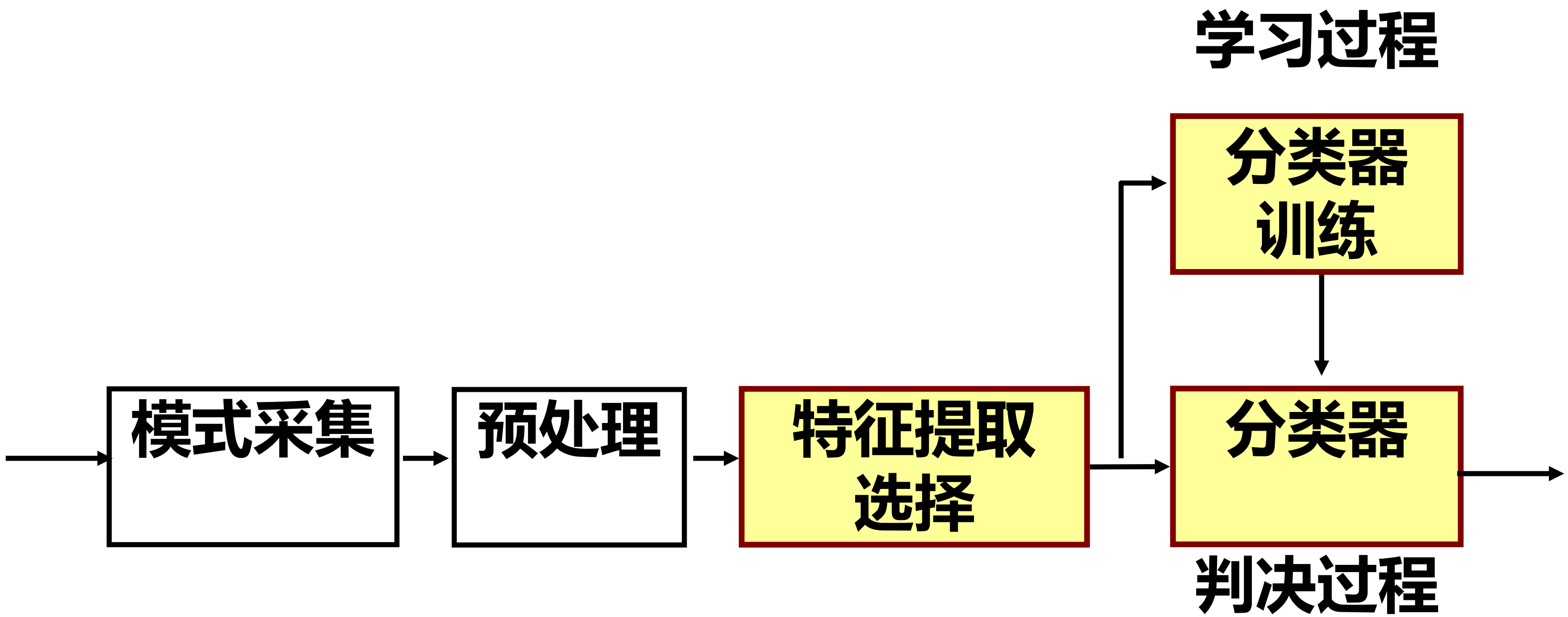


图1 模式识别系统框图

（1）统计判决

- 理论基础：概率论，数理统计
- 模式描述方法：特征向量 $\vec{x} = (x_1, x_2, \cdots, x_n)'$
- 模式判定

  是一个聚合类。用条件概率分布$p(\vec{x} \mid \omega_i)$表示,m类就有m个分布，然后判定未知模式属于哪一个分布。
- 主要方法

  几何分类：线性分类，非线性分类

  统计分类：Bayes决策

  无教师的分类：聚类分析
- 主要优点

  1）比较成熟

  2）能考虑干扰噪声等影响

  3）识别模式基元能力强
- 主要缺点

  1）对结构复杂的模式抽取特征困难

  2）不能反映模式的结构特征，难以描述模式的性质

  3）难以从整体角度考虑识别问题

第二章　聚类分析 总结

- 内容：
  - 聚类的基本概念；
  - 相似性测度、类间距离 、聚类准则；
  - 简单聚类、层次聚类；
  - 动态聚类。
- 要求：

  重点:相似性测度、C均值聚类和层次聚类算法。

  难点:聚类准则函数。

第二章　聚类分析 总结

方法：根据待分类的模式属性或特征相似程度进行分类，相似的模式归为一类，不相似的模式分划到不同类中。

第二章　聚类分析 总结

一、影响分类的因数

（1）分类准则；（2）特征量的选择；（3）量纲。

二、模式相似性测度

（一） 距 离 测 度

(1) 欧氏距离

(2) 马氏距离 $d^2(\vec{x}_i,\vec{x}_j)=(\vec{x}_i-\vec{x}_j)'V^{-1}(\vec{x}_i-\vec{x}_j)$

对坐标系平移、旋转、比例不变（非奇异线性变换不变）。

（二） 相 似 测 度

相关系数 （特征矢量的方向） $r(\vec{x},\vec{y})=\dfrac{(\vec{x}-\bar{\vec{x}})'(\vec{y}-\bar{\vec{y}})}{[(\vec{x}-\bar{\vec{x}})'(\vec{x}-\bar{\vec{x}})(\vec{y}-\bar{\vec{y}})'(\vec{y}-\bar{\vec{y}})]^{1/2}}$

对坐标系平移、旋转、比例不变。

（三） 匹 配 测 度

2.3 聚类算法

（一）简单聚类

①最邻近规则试探法

给定阈值T，聚类到z_i

②最大最小距离法

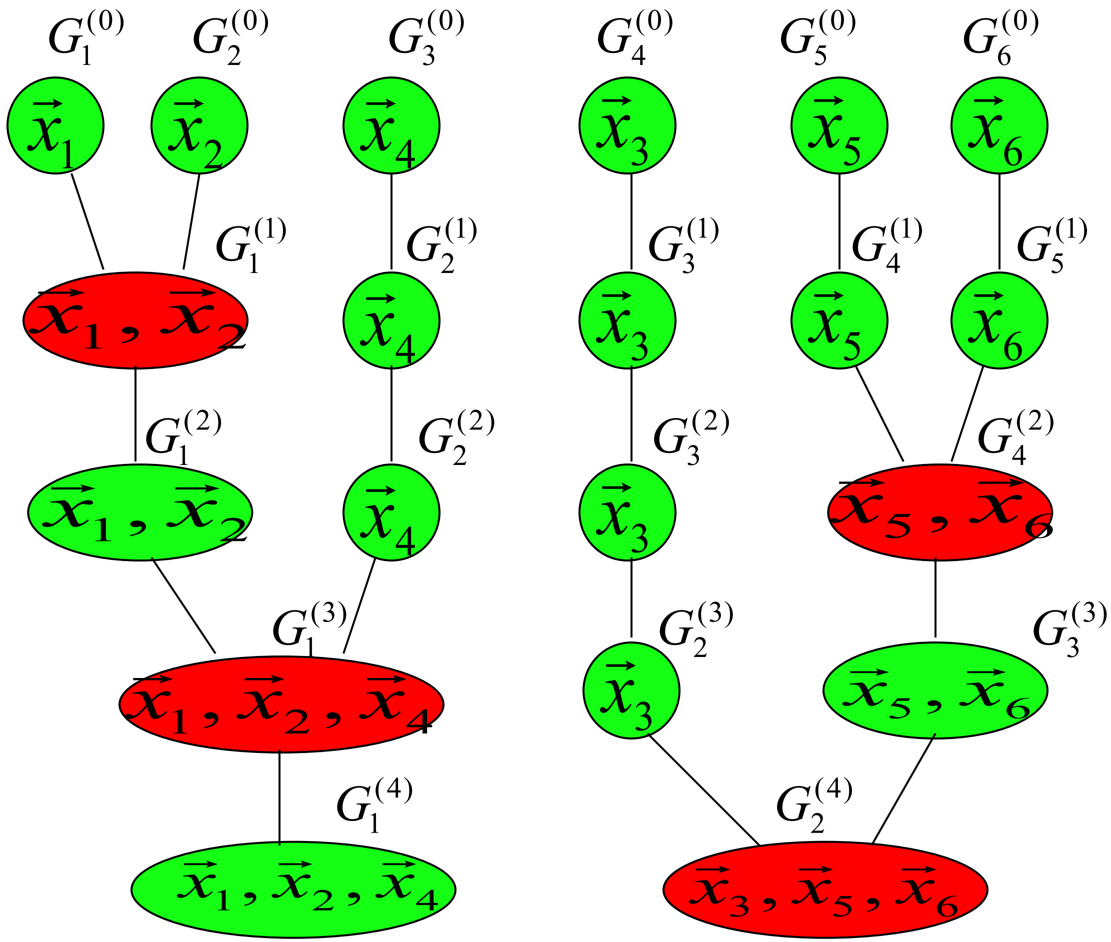
给定$0<\theta<1$，距当前类心最大者为新的聚类中心，按最小距离原则聚类。

（二）层次聚类

初始每个样本点为一类（N类），将类间距离最小者合并为一类，逐级进行。

类间距离可用：最小、最大、中间、重心、平均距离等。

层次聚类示意图



停止条件：类间距离>门限T 或/和 **类别数目**=C

（三）动态聚类算法

- **C-均值算法（适用于团状分布的情况）**

$$\forall c > 0, \quad \vec{z}_i(1) = \vec{x}_{i_1}, \ i = 1, 2, \ldots, c;$$

$$\vec{z}_j(k) = \frac{1}{N_j} \sum_{i=1}^{N_j} \vec{x}_i^{(j)}, \quad \mathrm{N} = \sum_{\mathrm{j}=1}^{\mathrm{c}} \mathrm{N}_\mathrm{j}, \quad \vec{x}_i^{(j)} \in \omega_\mathrm{j}$$

重新聚类

- **ISODATA算法**

c（预期类数），N_c（初始类心个数），θ_N（各类最小样本数），

θ_s（类中样本特征分量标准差上限），$\varepsilon\sigma_{jmax}$，

θ_D（聚合中心最小间距），L，I

C-均值算法步骤

(1) 任选个模式特征矢量作为初始聚类中心,令$k=0$：

$$\vec{z}_1^{(0)}, \vec{z}_2^{(0)}, \cdots, \vec{z}_c^{(0)}$$

(2) 将待分类的模式特征矢量集中的模式逐个按最小距离原则分划给类中的某一类，即

$$\text{IF} \;\; d_{il}^{(k)} = \min_{\text{j}}[d_{ij}^{(k)}], \quad i=1,2,\cdots,c. \;\; d_{ij}^{(k)} = (\vec{x}_i - \vec{z}_j^{(k)})'(\vec{x}_i - \vec{z}_j^{(k)})$$

$$\text{then} \;\; \vec{x}_i \in \omega_l^{(k+1)}$$

得到新的类：$\omega_j^{(k+1)}, j=1,2,\cdots,c$

(3) 计算重新分类后的各类心

$$\vec{z}_j^{(k+1)} = \frac{1}{n_j^{(k+1)}} \sum_{\vec{x}_i \in \omega_j^{(k+1)}} \vec{x}_i \quad , \quad j=1,2,\cdots,c$$

(4) $\text{IF} \;\; (\vec{z}_j^{(k+1)} = \vec{z}_j^{(k)} \;, \; j=1,2,\cdots,c)$

$\text{then} \;\; \text{End}$

$\text{else} \quad k = k+1; \;\; \text{goto (2)};$

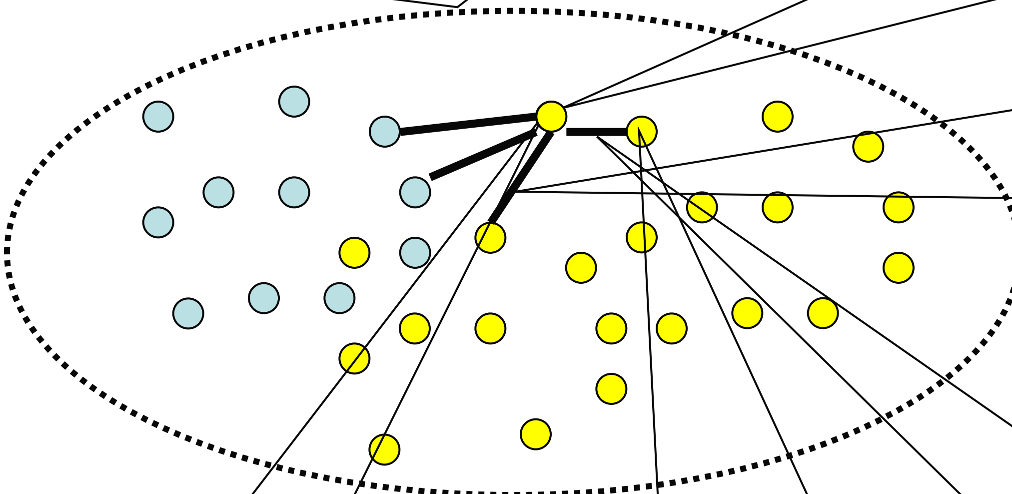
C-均值算法性能

- 算法简单，收敛。如模式呈现类内团状分布，效果很好，故应用较多。
- 能使各模式到其所判属类别中心距离平方之和为最小。
- 以确定的类数 c、模式输入次序、及选定的初始聚类中心为前提，受此限制结果只是局部最优。

2.6.1 最近邻决策规则—1-NN

(1)已知N个已知类别样本X

(2)输入未知类别样本x



(3)计算x到$x_i \in X$，$(i=1,2,\ldots,N)$的距离$d_i(x)$

(4)找出最小距离$d_m(x)=\min\{d_i(x)\}$

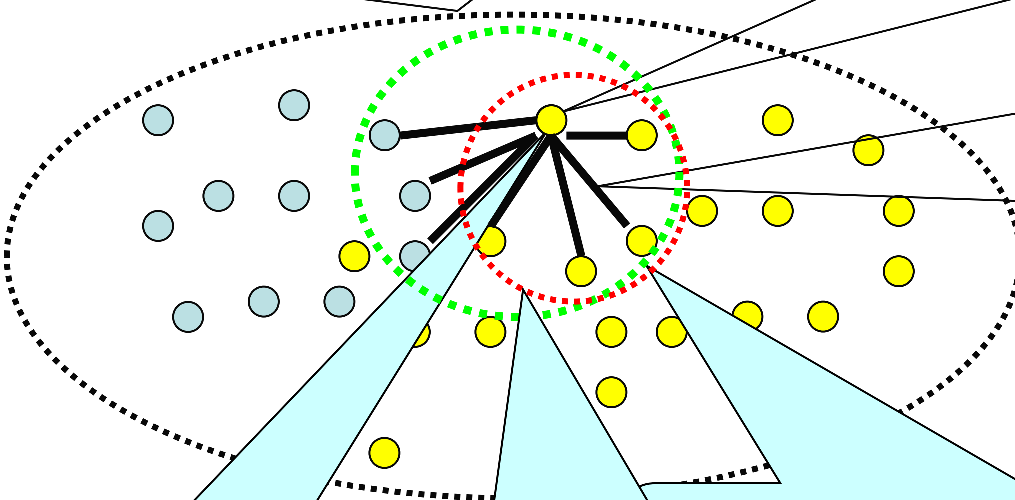
(5)看x_m属于哪一类：$x_m \in \omega_2$

(6) 判$x \in \omega_2$

2.6.1 最近邻决策规则—k-NN

(1)已知N个已知类别样本X

(2)输入未知类别样本x



(3)计算x到$x_i \in X$，$(i=1,2,\ldots,N)$的距离$d_i(x)$

(4)找出x的k个最近邻元$X_k=\{x_i, i=1,2,\ldots,k\}$

(5)看X_k中属于哪一类的样本最多$k_1=3<k_2=4$

(6) 判$x \in \omega_2$

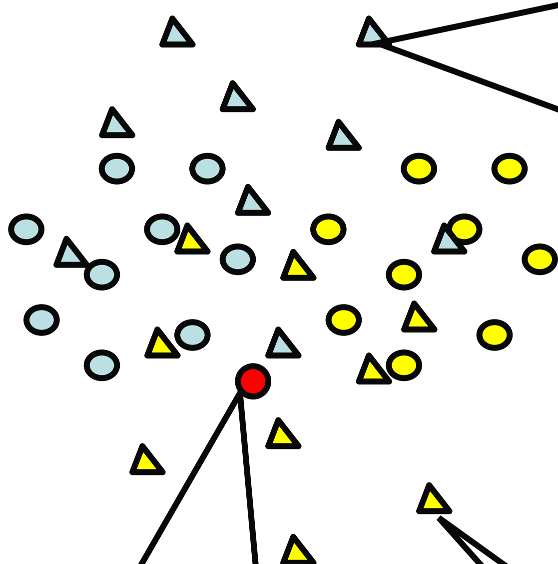
剪辑1-NN最近邻方法

□ $\in \omega_1$

■ $\in \omega_2$

○ $\in X^{(NR)}$

△ $\in X^{(NT)}$



用$X^{(NR)}$中的样本采用最近邻规则对$X^{(NT)}$中的每个样本分类，剪辑掉$X^{(NT)}$中被错误分类的样本。

余下判决正确的样本组成剪辑样本集$X^{(NTE)}$。

用$X^{(NTE)}$对输入的未知样本做1-NN分类。

1-NN

判决函数 $d_i(\vec{x}) = \underbrace{\min}_{j=1,2,\cdots,N_i} \left\| \vec{x} - \vec{x}_j^{(i)} \right\| \quad i = 1, 2, \cdots, c$

判决规则：如果 $d_m(\vec{x}) = \underbrace{\min}_{i=1,2,\cdots,c} d_i(\vec{x})$ 则 $\vec{x} \in \omega_m$

k-NN

判决函数 $d_i(\vec{x}) = k_i$

判决规则：如果 $d_m(\vec{x}) = \underbrace{\max}_{i=1,2,\cdots,c} d_i(\vec{x})$ 则 $\vec{x} \in \omega_m$

第三章 类域界面方程法 总结

分类⇒特征空间的分划⇒子空间的界面：判别函数

$$d(\vec{x})=0 \qquad d(\vec{x})=\vec{w}'\vec{x}$$

求解 $\vec{w}=(w_1,w_2,\cdots,w_n,w_{n+1})'$

3·2 线性判别函数

$$d(\vec{x})=w_1x_1+w_2x_2+\cdots+w_nx_n+w_{n+1} \triangleq \vec{w}_0'\vec{x}+w_{n+1}$$

式中 $\vec{w}_0=(w_1,w_2,\cdots,w_n)'$**,称为权矢量或系数矢量。写成矢量形式**

$$d(\vec{x}) \triangleq \vec{w}'\vec{x}$$

这里 $\vec{x}=(x_1,x_2,\cdots,x_n,1)'$ **,** $\vec{w}=(w_1,w_2,\cdots,w_n,w_{n+1})'$ **，称为增广特征矢量和增广权矢量。增广特征矢量的全体称为增广特征空间。**

判别规则：

$$d(\vec{x})=\vec{w}'\vec{x}\begin{cases}>0\Rightarrow\vec{x}\in\omega_1\\<0\Rightarrow\vec{x}\in\omega_2\\=0\Rightarrow\vec{x}\in\omega_i\text{或拒判}\end{cases}$$

解多类问题的两分法:

(1) $\omega_i/\bar{\omega}_i$ 两分法

有不确定区域

$$if\ \begin{cases}d_i(\vec{x})>0\\d_j(\vec{x})\le 0,\ \forall j\ne i\end{cases}\ then\ \ \vec{x}\in\omega_i$$

(2) ω_i/ω_j 两分法

$$if\ \ d_{ij}(\vec{x})>0,\forall j\ne i,\ \ then\ \ \vec{x}\in\omega_i$$

(3) 没有不确定区的 ω_i/ω_j 两分法

令

$$d_{ij}(\vec{x})=d_i(\vec{x})-d_j(\vec{x})=(\vec{w}_i-\vec{w}_j)'\vec{x}$$

$$if\ \ \ d_i(\vec{x})>d_j(\vec{x}),\forall j\ne i\ \ \ then\ \ \vec{x}\in\omega_i$$

$$or$$

$$if\ \ \ d_i(\vec{x})=\max_j[d_j(\vec{x})]\ \ \ then\ \ \vec{x}\in\omega_i$$

3·3　判别函数值的鉴别意义、权空间及解空间

（1）系数矢量 $\vec{w}_0$ 是超平面 $d(\vec{x})=0$ 的法矢量；

（2）$d(\vec{x})$ 的绝对值 $\left|d(\vec{x})\right|$ 正比于 $\vec{x}$ 到超平面的距离；

（3）$d(\vec{x})$ 的正（负）反映 $\vec{x}$ 在超平面 $d(\vec{x})=0$ 的正（负）侧。

第四章　统计判决 总结

概念:

后验概率：$P(\omega_i|\vec{x})$

类概密：　$p(\vec{x}|\omega_i)$

先验概率：$P(\omega_i)$

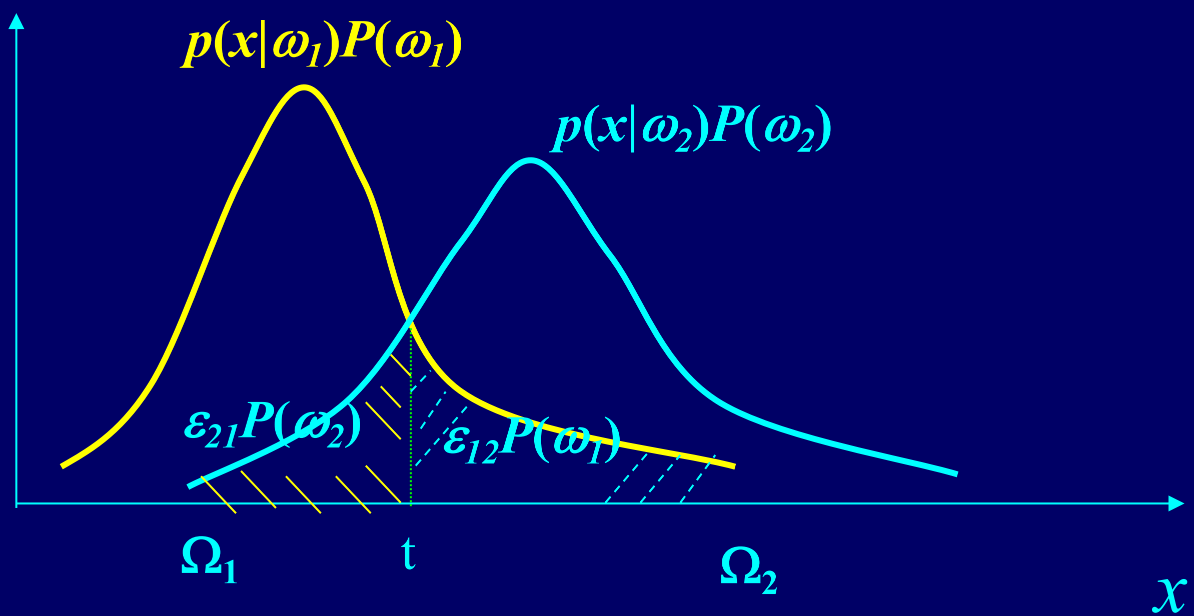
条件期望：$E_i[g(\vec{x})]=\int_{X^n} g(\vec{x})p(\vec{x}|\omega_i)d\vec{x}$

4·1最小误判概率准则判决

总的误判概率

$$\varepsilon_{12} = \int_{\Omega_2} p(\vec{x}|\omega_1)d\vec{x} \quad \varepsilon_{21} = \int_{\Omega_1} p(\vec{x}|\omega_2)d\vec{x}$$

$$P(e) = P(\omega_1)\varepsilon_{12} + P(\omega_2)\varepsilon_{21}$$
$$= P(\omega_1)\int_{\Omega_2} p(\vec{x}|\omega_1)d\vec{x} + P(\omega_2)\int_{\Omega_1} p(\vec{x}|\omega_2)d\vec{x}$$



多类问题，最小误判概率准则有如下几种等价的判决规则

(1) $\text{if } P(\omega_i|\vec{x}) > P(\omega_j|\vec{x}), \forall j \neq i \text{ then } \vec{x} \in \omega_i$

$\text{if } P(\omega_i|\vec{x}) = \max_j [P(\omega_j|\vec{x})] \text{ then } \vec{x} \in \omega_i$

(2) $\text{if } p(\vec{x}|\omega_i)P(\omega_i) > p(\vec{x}|\omega_j)P(\omega_j), \forall j \neq i \text{ then } \vec{x} \in \omega_i$

$\text{if } p(\vec{x}|\omega_i)P(\omega_i) = \max_j [p(\vec{x}|\omega_j)P(\omega_j)] \text{ then } \vec{x} \in \omega_i$

(3) $\text{if } l_{ij}(\vec{x}) = \dfrac{p(\vec{x}|\omega_i)}{p(\vec{x}|\omega_j)} > \dfrac{P(\omega_j)}{P(\omega_i)} = \theta_{ij}, \forall j \neq i \text{ then } \vec{x} \in \omega_i$

(4) $\text{if } \ln p(\vec{x}|\omega_i) + \ln P(\omega_i) > \ln p(\vec{x}|\omega_j) + \ln P(\omega_j), \forall j \neq i \text{ then } \vec{x} \in \omega_i$

正态模式分类的误判概率

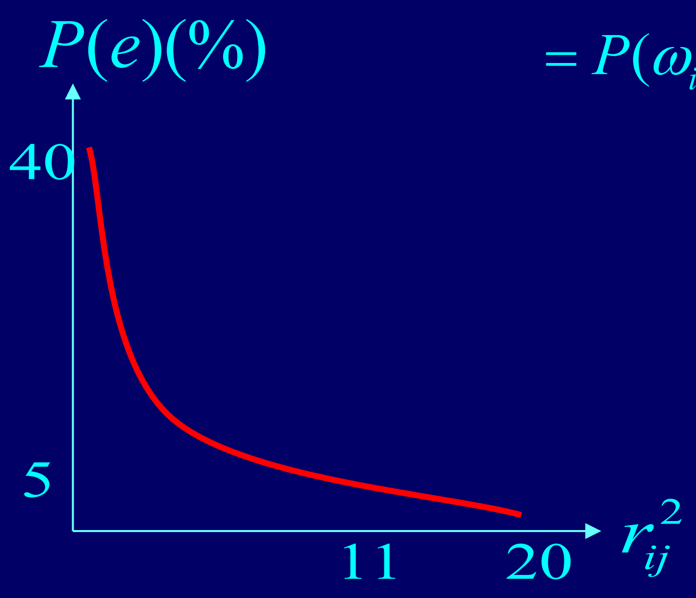
对数似然比 $$L_{ij}(\vec{x}) \triangleq \ln l_{ij}(\vec{x}) = \ln p(\vec{x}|\omega_i) - \ln p(\vec{x}|\omega_j)$$

$$= \vec{x}'\Sigma^{-1}(\vec{\mu}_i - \vec{\mu}_j) - \frac{1}{2}(\vec{\mu}_i + \vec{\mu}_j)'\Sigma^{-1}(\vec{\mu}_i - \vec{\mu}_j)$$

总的误判概率为

$$P(e) = P(\omega_i)P(L_{ij} < \alpha|\omega_i) + P(\omega_j)P(L_{ij} > \alpha|\omega_j)$$

$$= P(\omega_i)\Phi\left[\frac{\alpha - \frac{1}{2}r_{ij}^2}{r_{ij}}\right] + P(\omega_j)\left\{1 - \Phi\left[\frac{\alpha + \frac{1}{2}r_{ij}^2}{r_{ij}}\right]\right\}$$



4·2 最小损失准则判决

- 对一个实属ω_i类的模式采用了决策α_j所造成损失记为损失函数 $\lambda(\alpha_j/\omega_i) \triangleq \lambda_{ij}$
- 条件平均损失 $R_j(\vec{x}) = \sum_{i=1}^{c} \lambda_{ij} P(\omega_i \mid \vec{x})$
- （总的）平均损失

$$R = \int_{\Omega} R_j(\vec{x}) p(\vec{x}) d\vec{x} = \sum_{j=1}^{c} \int_{\Omega_j} R_j(\vec{x}) p(\vec{x}) d\vec{x}$$
$$= \sum_{i=1}^{c} P(\omega_i) E_i\left[\lambda(\alpha(\vec{x}) \mid \omega_i)\right]$$
$$= E\left[\lambda(\alpha(\vec{x}))\right]$$

- [定理]：使<u>条件平均损失</u>最小的判决也必然使总的<u>平均损失</u>最小。
- 最小损失准则判决：$if \quad R_j(\vec{x}) = \min_i \left[R_i(\vec{x})\right] \quad then \quad \vec{x} \in \omega_j$
- 两类问题

$$\frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} \gtrless \frac{P(\omega_2)(\lambda_{21}-\lambda_{22})}{P(\omega_1)(\lambda_{12}-\lambda_{11})} \Rightarrow \vec{x} \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$$

4·3 最小最大损失准则

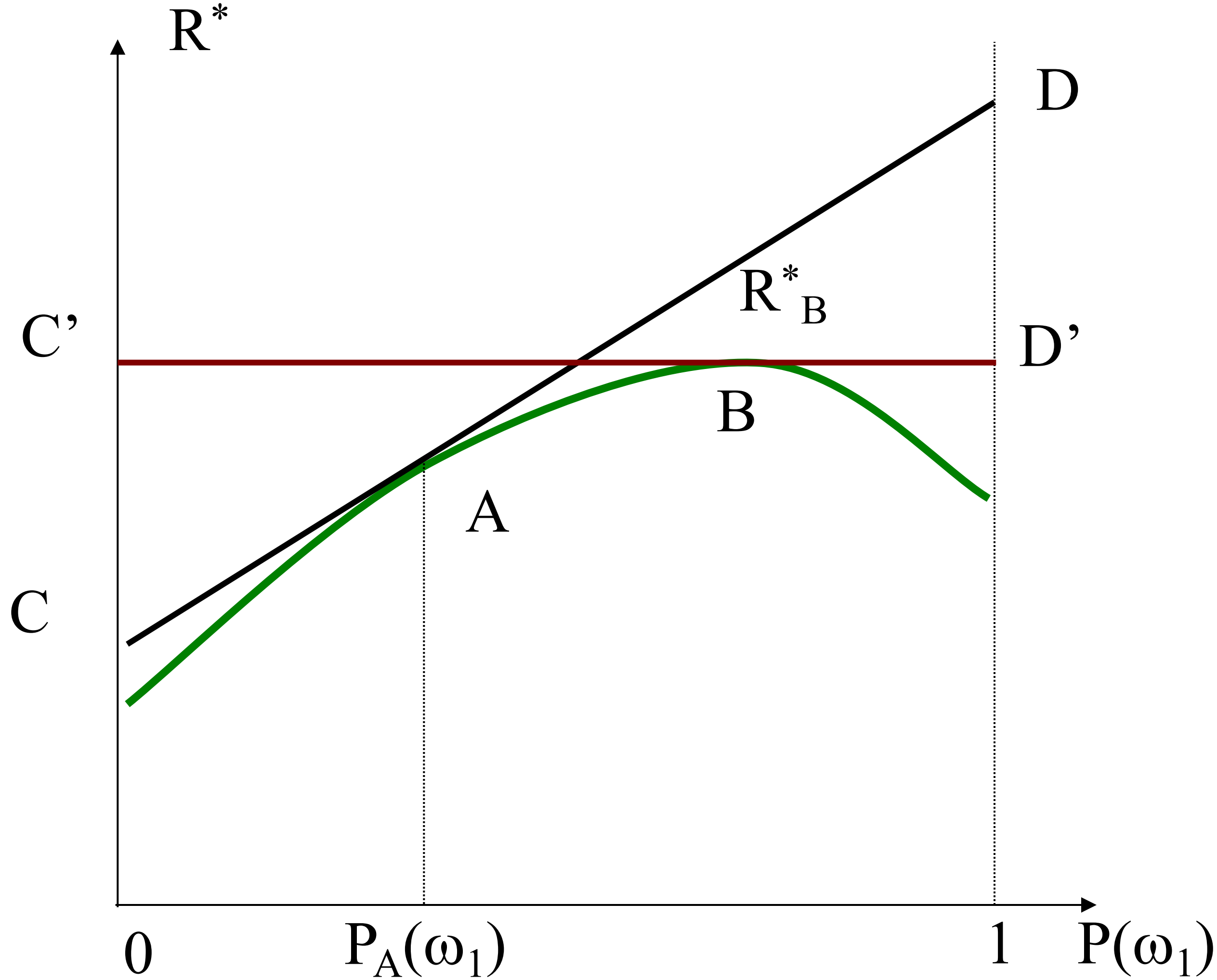
适用于$P(\omega_i)$未知或是变动的情况。

当类概密已知，损失函数λ_{ij}、类域Ω_j取定后，平均损失R是$P(\omega_i)$的线性函数

$$R=\lambda_{22}+(\lambda_{21}-\lambda_{22})\int_{\Omega_1}p(\vec{x}|\omega_2)d\vec{x}$$
$$+P(\omega_1)[(\lambda_{11}-\lambda_{22})+(\lambda_{12}-\lambda_{11})\int_{\Omega_2}p(\vec{x}|\omega_1)d\vec{x}+(\lambda_{22}-\lambda_{21})\int_{\Omega_1}p(\vec{x}|\omega_2)d\vec{x}]$$
$$\triangleq a+bP(\omega_1)$$

方法：按最小损失准则，对$P(\omega_i)\in(0,1)$，计算R ~ $P(\omega_i)$曲线，找出使R取最大值的$P^*(\omega_i)$，然后最小损失准则对具体的模式分类识别。最小最大损失判决规则为

$$if\quad \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)}\gtrless\frac{(\lambda_{21}-\lambda_{22})(1-P^*(\omega_1))}{(\lambda_{12}-\lambda_{11})P^*(\omega_1)}\Rightarrow\vec{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$



$$dR(P(\omega_1))/dP(\omega_1) = b = 0$$

$$R_B^* = a = \lambda_{22} + (\lambda_{21} - \lambda_{22})\int_{\Omega_1} p(\vec{x} \mid \omega_2)d\vec{x}$$

各种判决适用情况

- 在某些实际问题中，可能存在以下几种情况：
  - (1) 不知道各类的先验概率$P(\omega_i)$；
  - (2) 难于确定误判的代价λ_{ij} ；
  - (3) 某一种错误较另一种错误更为重要。
- 针对(1)，可以采用最小最大损失准则或令各类概率相等的办法克服；
- 针对(2)，如果允许的话，可以避开使用损失函数而采用最小误判概率准则 ；
- 针对(3)，可以采用N-P判决。

统计判决要点总结

$$l_{12}(\vec{x}) = \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} \gtrless \frac{P(\omega_2)}{P(\omega_1)} \quad \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

最小误判概率准则

$$l_{12}(\vec{x}) = \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} \gtrless \frac{P(\omega_2)(\lambda_{21}-\lambda_{22})}{P(\omega_1)(\lambda_{12}-\lambda_{11})} \quad \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

最小损失准则

$$l_{12}(\vec{x}) = \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} \gtrless \frac{(\lambda_{21}-\lambda_{22})(1-P^*(\omega_1))}{(\lambda_{12}-\lambda_{11})P^*(\omega_1)} \quad \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

最小最大损失判决准则

$$l_{12}(\vec{x}) = \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} \gtrless \lambda \quad \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

N-P准则

第五章 统计决策中的训练、学习与错误率测试、估计

本章目的：

已知类别的样本（训练样本）→学习或训练→类概密 $p(\vec{x}|\omega_i)$

5·1 统计推断概述

参数估计

如果已知ω_i类的概密$p(\vec{x}|\omega_i)$的函数类型，但不知道其中的参数或参数集 $\{\theta_i\} \triangleq \vec{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)'$

→可采用参数估计确定未知参数 $\vec{\theta}$

(1)将参数作为非随机量处理，如矩法估计、最大似然估计；

(2)将参数作为随机变量，如贝叶斯估计。

非参数估计

当不知道类的概型时，可采用非参数估计方法，也称为总体推断，这类方法有

(1) p-窗法；

(2) 有限项正交函数级数逼近法。

5.2.2 最大似然估计

(Maximum Likelihood Estimate)

- **似然函数** $p(\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_N;\vec{\theta})$

  称为相对于 $X^{(N)}=\{\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_N\}$ 的 $\vec{\theta}$ 的似然函数。

  如果各个 $\vec{x}_j\,(j=1,2,\cdots,N)$ 是独立抽取的，则有

$$p(X^{(N)}\mid\vec{\theta})=p(\vec{x}_1\mid\vec{\theta})p(\vec{x}_2\mid\vec{\theta})\cdots p(\vec{x}_N\mid\vec{\theta})=\prod_{j=1}^{N}p(\vec{x}_j\mid\vec{\theta})$$

- **最大似然估计**

取 $p(X^{(N)}\mid\vec{\theta})=\max_{\vec{\theta}}\left[p(X^{(N)}\mid\vec{\theta})\right]$ 的 $\hat{\vec{\theta}}$ 作为未知参数集的估计值。

根据已知样本 $X^{(N)}$ 和下列似然方程组或对数似然方程组,可得 $\hat{\vec{\theta}}$

$$\frac{\partial p(X^{(N)}\mid\vec{\theta})}{\partial\vec{\theta}}=0 \qquad \frac{\partial[\ln p(X^{(N)}\mid\vec{\theta})]}{\partial\vec{\theta}}=\sum_{j=1}^{N}\frac{\partial}{\partial\vec{\theta}}\ln p(\vec{x}_j\mid\vec{\theta})=0$$

5.2.3贝叶斯估计

（1）确定未知参数集$\vec{\theta}$的先验概密 $p(\vec{\theta}) = N(\vec{\theta}_0, C_0)$；

（2）由样本集$X^{(N)}$ 求 $p(X^{(N)} | \vec{\theta})$

$$p(X^{(N)} | \vec{\theta}) = p(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N | \vec{\theta}) = \prod_{j=1}^{N} p(\vec{x}_j | \vec{\theta}) \qquad (5\text{-}2\text{-}17)$$

这里 $p(\vec{x}_j | \vec{\theta})$ 的概型是已知的；

（3）计算

$$p(\vec{\theta} \,|\, X^{(N)}) = p(\vec{\theta} \,|\, \vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N) = \alpha^{-1}(X^{(N)}) \prod_{j=1}^{N} p(\vec{x}_j \,|\, \vec{\theta}) p(\vec{\theta}) \qquad (5\text{-}2\text{-}19)$$

$$\alpha(X^{(N)}) = \int_{\Theta} p(X^{(N)} | \vec{\theta}) p(\vec{\theta}) d\vec{\theta} = \int_{\Theta} \prod_{j=1}^{N} p(\vec{x}_j | \vec{\theta}) p(\vec{\theta}) d\vec{\theta}$$

（4）计算

$$\hat{\vec{\theta}}(X^{(N)}) = \int_{\Theta} \vec{\theta} p(\vec{\theta} \,|\, X^{(N)}) d\vec{\theta} = E\left[\vec{\theta} \,\middle|\, X^{(N)}\right] \qquad (5\text{-}2\text{-}16)$$

5·3 贝叶斯学习

贝叶斯学习与贝叶斯估计前提条件是相同的，不同的是，贝叶斯学习不是进行概密的参数估计，而是进行总体概密的估计以获得 $p(\vec{x}|X^{(N)})$。

在前一节贝叶斯估计的四个步骤中，贝叶斯学习在执行完了前三个步骤得到未知参数的 后验概密 $p(\vec{\theta}|X^{(N)})$ 之后不是去求 $\hat{\vec{\theta}}$，而是直接求总体的后验概密 $p(\vec{x}|X^{(N)})$。

$$p(\vec{x}\mid X^{(N)})=\int_{\Theta} p(\vec{x},\vec{\theta}\mid X^{(N)})d\vec{\theta}=\int_{\Theta} p(\vec{x}\mid \vec{\theta},X^{(N)})p(\vec{\theta}\mid X^{(N)})d\vec{\theta}$$

$$=\int_{\Theta} p(\vec{x}\mid \vec{\theta})p(\vec{\theta}\mid X^{(N)})d\vec{\theta}$$