

1.什么是模式及模式识别？模式识别的应用领域主要有哪些？

模式：存在于时间，空间中可观察的事物，具有时间或空间分布的信息；

模式识别：用计算机实现人对各种事物或现象的分析，描述，判断，识别。

模式识别的应用领域：（1）字符识别；（2）医疗诊断；（3）遥感；

（4）指纹识别 脸形识别；（5）检测污染分析，大气，水源，环境监测；

（6）自动检测；（7）语声识别，机器翻译，电话号码自动查询，侦听，机器故障判断；

（8）军事应用。

2.模式识别系统的基本组成是什么？

（1）信息的获取：是通过传感器，将光或声音等信息转化为电信息；

（2）预处理：包括 A/D,二值化，图象的平滑，变换，增强，恢复，滤波等，主要指图象处理；

（3）特征抽取和选择：在测量空间的原始数据通过变换获得在特征空间最能反映分类本质的特征；

（4）分类器设计：分类器设计的主要功能是通过训练确定判决规则，使按此类判决规则分类时，错误率最低。把这些判决规则建成标准库；

（5）分类决策：在特征空间中对被识别对象进行分类。

3.模式识别的基本问题有哪些？

（1）模式(样本)表示方法：（a）向量表示；（b）矩阵表示；（c）几何表示；（4）基元(链码)表示；

（2）模式类的紧致性：模式识别的要求：满足紧致集，才能很好地分类；如果不满足紧致集，就要采取变换的方法，满足紧致集

（3）相似与分类；（a）两个样本 x_i, x_j 之间的相似度量满足以下要求：

应为非负值

样本本身相似性度量应最大

度量应满足对称性

在满足紧致性的条件下，相似性应该是点间距离的

单调函数

（b）用各种距离表示相似性

（4）特征的生成：特征包括：（a）低层特征；（b）中层特征；（c）高层特征

（5）数据的标准化：（a）极差标准化；（b）方差标准化

4.线性判别方法

（1）两类：二维及多维判别函数，判别边界，判别规则

二维情况：（a）判别函数： $g(x) = w_1 x_1 + w_2 x_2 + w_3$ （ w 为参数， x_1, x_2 为坐标向量）

（b）判别边界： $g(x)=0$;

（c）判别规则：

$$g_i(x) = \begin{cases} > 0, X \in \omega_1 \\ < 0, X \in \omega_2 \end{cases}$$

n 维情况：（a）判别函数： $g(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_{n+1}$

也可表示为： $g(x) = W^T X$

$W = (w_1, w_2, \dots, w_n, w_{n+1})^T$ 为增值权向量，

$X = (x_1, x_2, \dots, x_n, x_{n+1})^T$ 为增值模式向量。

(b) 判别边界： $g_1(x) = W^T X = 0$

(c) 判别规则：

$$g_i(x) = W_i^T X \begin{cases} > 0, X \in \omega_i \\ < 0, \text{其它}, i = 1, 2, \dots, M. \end{cases}$$

(2) 多类：3种判别方法（函数、边界、规则）

(A) 第一种情况：(a) 判别函数：M类可有M个判别函数

$$g_i(x) = W_i^T X$$

式中 $W_i = (w_{i1}, w_{i2}, \dots, w_{in}, w_{in+1})^T$ 为第i个判别函数的权向量。

(b) 判别边界： i ($i=1, 2, \dots, n$) 类与其它类之间的边界由 $g_i(x)=0$ 确定

(c) 判别规则：

$$g_i(x) = W_i^T X \begin{cases} > 0, X \in \omega_i \\ < 0, \text{其它}, i = 1, 2, \dots, M. \end{cases}$$

(B) 第二种情况：(a) 判别函数：有 $M(M-1)/2$ 个判别平面

$$g_{ij}(x) = W_{ij}^T X$$

(b) 判别边界： $g_{ij}(x) = 0$

(c) 判别规则：

$$g_{ij}(x) \begin{cases} > 0 \rightarrow \text{当 } x \in \omega_i \\ < 0 \rightarrow \text{当 } x \in \omega_j \end{cases} \quad i \neq j$$

(C) 第三种情况：(a) 判别函数： $g_k(x) = W_k^T X$

(b) 判别边界：

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

(c) 判别规则：

$$g_i(x) = W_k^T X \begin{cases} \text{最大, 当 } x \in \omega_i \\ \text{小, 其它} \end{cases}$$

5. 什么是模式空间及加权空间，解向量及解区？

(1) 模式空间：由 $X = (x_1, x_2, x_3, \dots, x_n)^T$ 构成的 n 维欧氏空间；

(2) 加权空间：以 w_1, w_2, \dots, w_n 为变量构成的欧氏空间；

(3) 解向量：分界面为 H ， W 与 H 正交， W 称为解向量；

(4) 解区：解向量的变动范围称为解区。

6. 超平面的四个基本性质是什么？

性质： W 与 H 正交；

性质：

$$\|r\| = \frac{g(x)}{\|W\|}$$

其中， $\|r\|$ 为 x 矢量到 H 的正交投影；

性质：

$$\|q\| = \frac{W_{n+1}}{\|W\|}, \text{原点到 } H \text{ 的距离与 } W_{n+1} \text{ 成正比}$$

性质：

若 $W_{n+1} > 0$, 则 H 在原点正侧, 若 $W_{n+1} < 0$, 则 H 在原点负侧。

若 $W_{n+1} = 0$, 则 $g(x) = W^T x$, 说明超平面 H 通过原点。

7. 二分法能力如何表示？

N 个样品线性可分数目 (条件: 样本分布良好) :

$$D(N, n) = \begin{cases} 2^N, & \text{若 } N \leq n+1 \\ 2 \sum_{k=0}^n C_{N-1}^k, & \text{若 } N > n+1 \end{cases}$$

其中 $C_{N-1}^k = \frac{(N-1)!}{k!(N-k-1)!}$, N 为样本数, n 为特征数

线性可分概率：

$$P(N, n) = \frac{D(N, n)}{2^N} = \begin{cases} 1, & \text{若 } N \leq n+1 \\ 2^{1-N} \sum_{k=0}^n C_{N-1}^k, & \text{若 } N > n+1 \end{cases}$$

(a): 当 n 由 $1 \rightarrow \infty$ 时, 曲线急剧下降, 在 $\lambda = 2$ 处出现明显的门限效应。

(b): 对于任意 n 值, $\lambda = 2$ 时, 即 $N = 2(n+1)$ 时, 线性可分概率为 $P(N, n) = \frac{1}{2}$ 。

(c): 在 $\lambda < 2$ 范围, 即 $N < 2(n+1)$, $P(N, n) \approx 1$, 说明样本少时二分能力强。

(d): 在 $\lambda > 2$ 范围, 即 $N > 2(n+1)$, 线性可分概率急剧下降, 说明样品越多线性可分能力越差。

(e): 对 N 个样品的线性可分性 (二分能力) 的估计: $N_0 = 2(n+1)$, 即 $\lambda = 2$ 是最好情况。

8. 广义线性判别方法

(1) 非线性 线性

一个非线性判别函数通过映射, 变换成线性判别函数:

$$g(x) = \sum_{i=1}^{k+1} w_i f_i(x) \xrightarrow{x \text{ 空间} \rightarrow \text{变换} \rightarrow y \text{ 空间}} W^T Y = g(Y)$$

$$\text{其中: } W = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{bmatrix}, (\text{广义权向量}) \quad Y = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \dots \\ f_k(x) \end{bmatrix}, (\text{增广模式向量})$$

(2) 线性判别

判别平面： $W^T Y = 0$

$$W^T Y = g(Y) \begin{cases} > 0, x \in \omega_1 \\ < 0, x \in \omega_2 \end{cases}$$

9. 分段线性判别方法

1) 基于距离：(1) 子类，类判别函数

(2) 判别规则

(1) 子类：把 ω_i 类可以分成 l_i 个子类：

$\omega_i = (\omega_i^1, \omega_i^2, \dots, \omega_i^{l_i})$ 分成 l_i 个子类。

子类判别函数：

$$g_i(x) = \min_{j=1,2,\dots,l_i} \|x - \mu_j^i\|$$

在同类的子类中找最近的均值

(2) 判别规则：

$$g_j(x) = \min g_i(x), i = 1, 2, \dots, M$$

这是在 M 类中找最近均值。则把 x 归于 j 类完成分类

2) 基于函数：(1) 子类，类判别函数

(2) 判别规则

(1) 子类别判别函数：对每个子类定义一个线性判别函数为：

$$g_i^j(x) = w_i^j x, \text{ 其中 } w_i^j \text{ 为 } \omega_i^j \text{ 子类的权向量。}$$

(2) 判别规则：在各子类中找最大的判别函数作为此类的代表，则对于 M 类，可定义

M 个判别函数 $g_i(x), i=1, 2, \dots, M$ 。因此，决策规则

$$g_j(x) = \max_{i=1,2,\dots,M} g_i(x), \text{ 则 } x \in \omega_j$$

3) 基于凹函数的并：(1) 析取范式，合取范式，凹函数

(2) 判别规则

(1)

析取范式： $P = (L_{11} \vee L_{12} \vee \dots \vee L_{1m})$

$\dots (L_{q1} \vee L_{q2} \vee \dots \vee L_{qm})$

合取范式： $Q = (L_{11} \wedge L_{12} \wedge \dots \wedge L_{1m}) \wedge \dots \wedge (L_{q1} \wedge L_{q2} \wedge \dots \wedge L_{qm})$

凹函数： $P_i = L_{i1} \vee L_{i2} \vee \dots \vee L_{im}$

(2)

判别规则：设第一类有 q 个峰，

则有 q 个凹函数。

即 $P = P_1 \vee P_2 \vee \dots \vee P_q$

$$L_{ij} = w_{ij} x \begin{cases} > 0, x \in \omega_{1,i}, i = 1, 2, \dots, q \text{ 子类。} \\ < 0, x \in \omega_{2,j}, j = 1, 2, \dots, m \text{ 每个子类的判别函数数。} \end{cases}$$

$$\text{判别规则：} \begin{cases} P > 0, \text{ 则 } x \in \omega_1 \\ P \leq 0, \text{ 则 } x \in \omega_2 \end{cases}$$

10. 非线性判别方法

(1) ω_1 集中, ω_2 分散

定义 ω_1 判别函数:

$$g(x) = k^2 - (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1), k \text{ 的大小, 决定超平面的大小。}$$

其中: μ_1 为 ω_1 均值, Σ_1 为 ω_1 协方差

$$\text{判别规则: } \begin{cases} g(x) > 0, x \in \omega_1 \\ g(x) < 0, x \in \omega_2 \end{cases}$$

判别平面: $g_1(x) = 0$ 是个超球面由 k 控制大小

(2) ω_1, ω_2 均集中

如果 ω_1, ω_2 都比较集中, 那么定义两个判别函数:

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

其中: μ_i 为 ω_1, ω_2 均值, Σ_i 为 ω_1, ω_2 协方差

$$\begin{aligned} \text{判别平面方程: } g(x) &= g_1(x) - g_2(x) \\ &= -x^2 (\Sigma_1^{-1} - \Sigma_2^{-1}) x + 2(\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1}) x - \\ &(\mu_1^T \Sigma_1^{-1} \mu_1 - \mu_2^T \Sigma_2^{-1} \mu_2) + (k_1^2 - k_2^2) = 0 \end{aligned}$$

$$\text{判别规则: } g(x) \begin{cases} > 0, x \in \omega_1 \\ < 0, x \in \omega_2 \end{cases}$$

k_1, k_2 可用来调整二类错误率。

11. 分类器的设计

(1) 梯度下降法 (迭代法): 准则函数, 学习规则

(a) 准则函数: $J(W) = J(W) + \frac{1}{2} (W - W_k)^T D (W - W_k)$

其中 D 为当 $W = W_k$ 时 $J(W)$ 的二阶偏导数矩阵

(b) 学习规则: 从起始值 W_1 开始, 算出 W_1 处目标函数的梯度矢量 $J(W_1)$, 则下一步的 w 值为: $W_2 = W_1 - \eta J(W_1)$ 其中 W_1 为起始权向量, η 为迭代步长, $J(W_1)$ 为目标函数, $J(W_1)$ 为 W_1 处的目标函数的梯度矢量

在第 K 步的时候

$$W_{k+1} = W_k - \eta_k J(W_k) \quad \text{最佳步长为 } \eta_k = \|J\|^2 / J^T D J$$

这就是梯度下降法的迭代公式。

(2) 感知器法: 准则、学习规则 (批量, 样本)

(a) 准则函数: $J(W) = \sum_{X \in X_0} (-W^T X)$ 其中 x_0 为错分样本

(b) 学习规则:

1. 错误分类修正 w_k

$$\text{如 } w_k^T x < 1 \quad \text{并且 } x \in X_1 \quad W_{k+1} = W_k + x$$

$$\text{如 } w_k^T x > 2 \quad \text{并且 } x \in X_2 \quad W_{k+1} = W_k - x$$

2. 正确分类, w_k 不修正

如 $w_k^T x > 0$ 并且 $x \in \Omega_1$

如 $w_k^T x < 0$ 并且 $x \in \Omega_2$

$$w_{k+1} = w_k$$

(3) 最小平方误差准则法 (MSE 法) (非迭代法) : 准则、权向量解

(a) 准则函数 :

$$J(W) = \|e\|^2 = \|XW - b\|^2 = \sum_{i=1}^N (W^T x_i - b_i)^2$$

(b) 权向量解 :

$$W = (X^T X)^{-1} X^T b = X^+ b$$

其中 $X^+ = (X^T X)^{-1} X^T$ 称为 X 的伪逆 (规范矩阵)

(4) 韦—霍氏法 (LMS 法) (迭代法) : 准则, 学习规则

(a) 准则函数 :

$$J(W) = \|e\|^2 = \|XW - b\|^2 = \sum_{i=1}^N (W^T x_i - b_i)^2$$

(b) 学习规则 : W_1 任意, $W_{k+1} = W_k + \rho_k (b - W_k^T X) X^k$ 取 $\rho_k = \frac{\rho_1}{k}$
 ρ_k 随迭代次数 k 而减少, 以保证算法收敛于满意的 W 值

(5) 何—卡氏法 (H-K 法) (迭代法) : 准则, b , W 的学习规则

(a) 准则 :

$$J(W) = \|e\|^2 = \|XW - b\|^2 = \sum_{i=1}^N (W^T x_i - b_i)^2$$

它的解为 : $W = (X^T X)^{-1} X^T b = X^+ b$

(b) b , W 的学习规则 :

对 b 前后两次迭代后, $b^{k+1} = b^k + \delta b^k$ 其中 δb^k 为 b 的增量

$$\delta b^k = C[e_k^+ | e_k^-]$$

其中 c 为矫正系数, e_k 为误差矢量, $e_k = XW_k - b_k$

$$W_{k+1} = X^+ b_{k+1} = X^+ [b_k + \delta b^k] = X^+ b_k + X^+ \delta b^k = W_k + c X^+ [e_k^+ | e_k^-]$$

初始条件 $W_1 = X^+ b_1$ 并且 $b_1 > 0$

迭代时检测

如果 $e_k \geq 0$ 时, $XW > b$, 系统线性可分, 迭代收敛

如果 $e_k < 0$ 时, $XW < b$, 系统线性不可分, 迭代不收敛

(6) Fisher 分类法 : 准则函数的建立, W 权值计算, W_0 的选择

(a) 准则函数的建立 : 投影样本之间的类间分离性越大越好, 投影样本的总离散度越小越好。所以 Fisher 准则函数有 $J(W) = \frac{|Y_1 - Y_2|^2}{(\sigma_1^2 + \sigma_2^2)}$
 即可表示为 :

$$J(W) = \frac{W^T S_b W}{W^T S_w W}$$

其中 S_w 为类内散布矩阵, S_b 为类间散布矩阵

$$S_w = S_1 + S_2 \quad S_1 = \sum_{X \in N_1} (X - \bar{X}_1)(X - \bar{X}_1)^T \quad S_2 = \sum_{X \in N_2} (X - \bar{X}_2)(X - \bar{X}_2)^T$$

$$S_b = (\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)^T$$

对 $J(W)$ 求极值得 $W = S_w^{-1}(\bar{X}_1 - \bar{X}_2)$

(b) W 权值计算：

(c) W_0 的选择：

$$1. W_0 = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

$$2. W_0 = \frac{N_1 \bar{Y}_2 + N_2 \bar{Y}_1}{N_1 + N_2} = \frac{N_1 W^T \bar{X}_1 + N_2 W^T \bar{X}_2}{N_1 + N_2}$$

$$3. W_0 = Y_1 + (Y_2 - Y_1) \frac{\sum_{k=1}^{N_1} (Y_{k1} - \bar{Y}_1)^2}{\sum_{k=1}^{N_1} (Y_{k1} - \bar{Y}_1)^2 + \sum_{k=1}^{N_2} (Y_{k2} - \bar{Y}_2)^2}$$

Y_{ki} 表示第 i 类中第 k 个样本的投影值

N_1 为 $_1$ 样本数

N_2 为 $_2$ 样本数

(7) 电位函数分类器：电位函数，累积电位的计算

(a) 电位函数：电位分布函数有如下三种形式：

$$1. K(X, X_k) = \exp\{-\alpha \|x - x_k\|^2\}$$

$$2. K(X, X_k) = \frac{1}{1 + \alpha \|x - x_k\|^2}$$

$$3. K(X, X_k) = \left| \frac{\sin \alpha \|x - x_k\|^2}{\alpha \|x - x_k\|^2} \right|$$

为系数 x_k 为某一特定点

(b) 累计电位的计算： $K_{k+1}(x) = K_k(x) + r_{k+1} K(x, x_k)$

其中： $x_{k+1}?$ $_1$ 并且 $K_k(x_{k+1}) > 0$ 时 $r_{k+1} = 0$

$x_{k+1}?$ $_1$ 并且 $K_k(x_{k+1}) < 0$ 时 $r_{k+1} = 1$

$x_{k+1}?$ $_2$ 并且 $K_k(x_{k+1}) < 0$ 时 $r_{k+1} = 0$

$x_{k+1}?$ $_2$ 并且 $K_k(x_{k+1}) > 0$ 时 $r_{k+1} = -1$

12. 1) 二类问题的贝叶斯判别

(1) 判别函数的四种形式

(2) 决策规则

(3) 决策面方程

(4) 决策系统的结构

(1) 判别函数的四种形式：

$$(A) g(x) = P(\omega_1/x) - P(\omega_2/x), (\text{后验概率})$$

$$(B) g(x) = P(x/\omega_1)P(\omega_1) - P(x/\omega_2)P(\omega_2), (\text{类条件概率密度})$$

(2) 判别规则： $(C) g(x) = \frac{P(x/\omega_1)}{P(x/\omega_2)} - \frac{P(\omega_2)}{P(\omega_1)}$, (似然比形式)

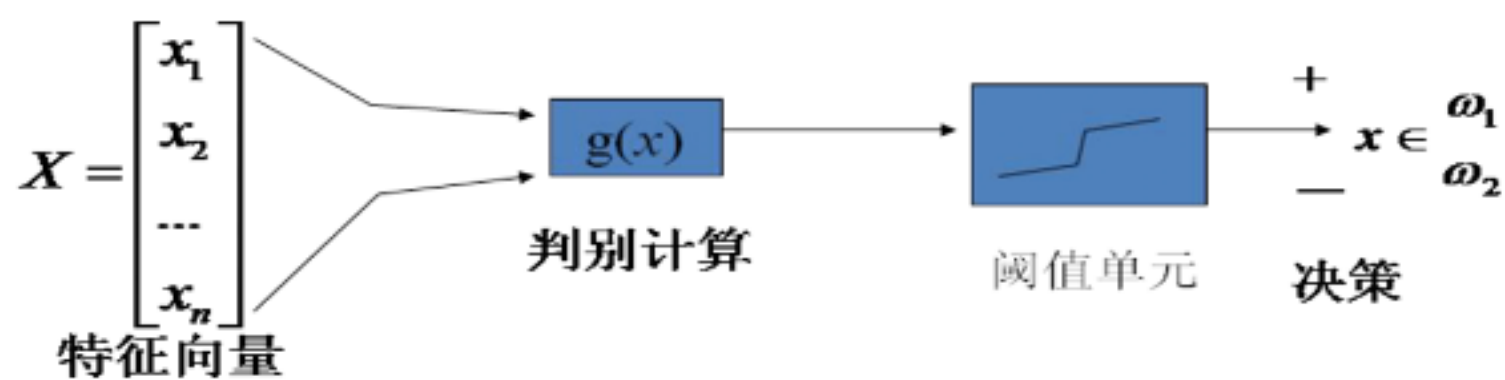
$$(D) g(x) = \ln \frac{P(x/\omega_1)}{P(x/\omega_2)} - \ln \frac{P(\omega_2)}{P(\omega_1)}, (\text{取对数方法})$$

$$\begin{aligned}
 (A) & P(\omega_1/x) > P(\omega_2/x) \Rightarrow x \in \omega_1 \\
 (B) & P(x/\omega_1)P(\omega_1) > P(x/\omega_2)P(\omega_2) \Rightarrow x \in \omega_1 \\
 (C) & \frac{P(x/\omega_1)}{P(x/\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_1 \\
 (D) & g(x) = \ln \frac{P(x/\omega_1)}{P(x/\omega_2)} > \ln \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow x \in \omega_1
 \end{aligned}$$

(3) 决策面方程: $g(x) = 0$

(4) 决策系统的结构

(A) 向量特征 (B) 判别计算 (C) 阈值单元 (D) 决策



2) 多类问题的贝叶斯判别

(1) 判别函数的四种形式

(2) 决策规则

(3) 决策面方程

(4) 决策系统的结构

(1) 判别函数的四种形式: M 类有 M 个判别函数 $g_1(x), g_2(x), \dots, g_M(x)$.

$$(A) g(x) = P(\omega_1/x) - P(\omega_2/x), (\text{后验概率})$$

$$(B) g(x) = P(x/\omega_1)P(\omega_1) - P(x/\omega_2)P(\omega_2), (\text{类条件概率密度})$$

$$(C) g(x) = \frac{P(x/\omega_1)}{P(x/\omega_2)} - \frac{P(\omega_2)}{P(\omega_1)}, (\text{似然比形式})$$

$$(D) g(x) = \ln \frac{P(x/\omega_1)}{P(x/\omega_2)} - \ln \frac{P(\omega_2)}{P(\omega_1)}, (\text{取对数方法})$$

(2) 决策规则:

$$\begin{aligned}
 g_i(x) &= P(x/\omega_i)P(\omega_i) \\
 &= \max_{1 \leq j \leq M} P(x/\omega_j)P(\omega_j) \Rightarrow x \in \omega_i, (i = 1, 2, \dots, M)
 \end{aligned}$$

另一种形式:

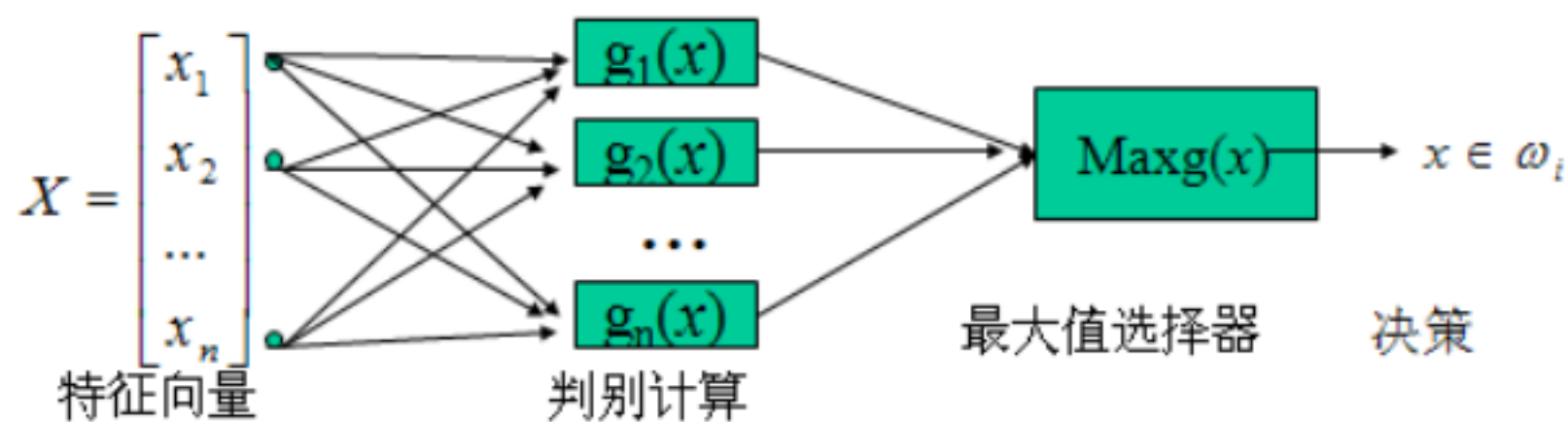
$$\begin{aligned}
 g_i(x) &= \ln P(x/\omega_i) + \ln P(\omega_i) \\
 &= \max_{1 \leq j \leq M} \{ \ln P(x/\omega_j) + \ln P(\omega_j) \} \Rightarrow x \in \omega_i
 \end{aligned}$$

(3) 决策面方程:

$$g_i(x) = g_j(x), \text{ 即 } g_i(x) - g_j(x) = 0$$

(4) 决策系统的结构：

(a)特征向量； (b)判别计算； (c)最大选择器； (d)决策



13. 三种最小错误率贝叶斯分类器（正态分布）：判别函数，判别规则，决策面方程

(1) 第一种情况：各个特征统计独立，且同方差情况。（最简单情况）

(a)判别函数： $g_i(x) = w_i^T x + w_{i0}$, (线性判别函数)

$$\text{其中： } w_i = \frac{1}{2\delta^2} \mu_i, w_{i0} = -\frac{1}{2\delta^2} \mu_i^T \mu_i + \ln P(\omega_i)$$

(b)判别规则： $g_i(x) = w_i^T x + w_{i0} = \max_{1 \leq j \leq M} \{w_j^T x + w_{j0}\} \Rightarrow x \in \omega_i$

(c)决策面方程：

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

$$W(x - x_0) = 0$$

$$\text{其中 } W = \mu_i - \mu_j$$

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

(2) 第二种情况： $\delta_i = \delta_j$ 相等，即各类协方差相等。

(a)判别函数： $g_i(x) = W_i^T x + w_{i0}$, (线性函数)

$$\text{其中 } W_i = \sum \Sigma^{-1} \mu_i$$

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

(b)判别规则： $g_i(x) = W_i^T x + w_{i0} = \max_{1 \leq j \leq M} \{W_j^T x + w_{j0}\} \Rightarrow x \in \omega_i$

(c)决策面方程：若 ω_i 与 ω_j 相邻 $\therefore g_i(x) - g_j(x) = 0$

$$\therefore W^T(x - x_0) = 0, \text{ 其中 } W = \sum \Sigma^{-1}(\mu_i - \mu_j)$$

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

(3) 第三种情况（一般情况）： Σ_i 为任意，各类协方差矩阵不等，二次项 $x^T \Sigma_i x$ 与 i 有关。所以判别函数为二次型函数。

(a)判别函数：

$$g_i(x) = x^T W_i x + W_i^T x + w_{i0}, \text{ 其中 } W_i = -\frac{1}{2} \Sigma_i^{-1}, (n \times n \text{ 矩阵})$$

$$W_i = \sum \Sigma_i^{-1} \mu_i (n \text{ 维列向量}), w_{i0} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)$$

(b)判别规则：

$$g_i(x) = x^T W_i x + W_i^T x + w_{i0}$$

$$= \max_{1 \leq j \leq M} \{x^T \bar{W}_j x + W_j^T x + w_{j0}\} \Rightarrow x \in \omega_i$$

(c)决策面方程： $g_i(x) - g_j(x) = 0$

14. 最小风险贝叶斯分类器：判别函数，判别规则

(1) 判别函数：

条件风险：

$$R(\alpha_i/x) = E[\lambda(\alpha_i/\omega_j)] = \sum_{j=1}^M \lambda(\alpha_i/\omega_j) P(\omega_j/x) \quad i=1,2,\dots,a. (a < M)$$

α_i ：表示把模式 x 判决为 ω_i 类的一次动作

期望风险： $R = \int R(\alpha(x)) x P(x) dx$, (平均风险)

(2) 判别规则：若 $R(\alpha_k/x) = \min_{i=1,2,\dots,M} R(\alpha_i/x)$ 则 $x \in \omega_k$:

15. 聂曼—皮尔逊判决：(二类)：准则，判别规则，阈值的确定

(1) 准则：在取 ε_2 为常数时， $\varepsilon_2 = \varepsilon_0$, 使 ε_1 最小，

(2) 判别规则：

$$\frac{P(x/\omega_1)}{P(x/\omega_2)} > T \Rightarrow x \in \omega_1 \therefore \text{皮尔逊规则归结为找阈值 } T.$$

(3) 阈值的确定：

当 $\frac{P(x/\omega_1)}{P(x/\omega_2)} = T$ 时, T 作 Ω_1, Ω_2 的分界线.

$$\varepsilon_2 = \int_{-\infty}^T P(x/\omega_2) dx, \therefore T \text{ 为 } \varepsilon_2 \text{ 的函数在取 } \varepsilon_2 \text{ 为常数时, } T \text{ 可确定,}$$

这时 ε_2 一定 ε_1 最小

16. 最小最大损失准则判决(二类)：准则，判别规则， $P^*(\omega_1)$ 的确定

(1) 准则：讨论在 $P(\omega_1)$ 变化时如何使最大可能风险最小；

(2) 判别规则：风险 $R = a + bP(\omega_1)$

$$\text{其中: } a = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\Omega_2} P(x/\omega_2) dx$$

$$b = (\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{\Omega_2} P(x/\omega_1) dx - (\lambda_{12} - \lambda_{22}) \int_{\Omega_2} P(x/\omega_2) dx$$

通过最小风险与先验概率的关系曲线，确定最大风险，使最大风险最小。

(3) $P^*(\omega_1)$ 的确定：

如果选择 Ω_1, Ω_2 使 $b = 0$, R 与 $P(\omega_1)$ 无关.

$$\text{即 } (\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{\Omega_1} P(x/\omega_1) dx - (\lambda_{12} - \lambda_{22}) \int_{\Omega_2} P(x/\omega_2) dx = 0$$

这时候最大风险为最小, $R = a = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\Omega_2} P(x/\omega_2) dx$

17. 什么是序贯分类?

序贯: 随着时间的推移可以得到越来越多的信息。

序贯分类决策规则

$$\begin{cases} \frac{P_i(X/\omega_1)}{P_i(X/\omega_2)} \geq A = \frac{1 - P_1(e)}{P_2(e)} \text{ 时 } X \in \omega_1 \\ \frac{P_i(X/\omega_1)}{P_i(X/\omega_2)} \leq B = \frac{P_1(e)}{1 - P_2(e)} \text{ 时 } X \in \omega_2 \\ B < \frac{P_i(X/\omega_1)}{P_i(X/\omega_2)} < A \text{ 时, 继续观察} \end{cases}$$

18. 什么是参数估计, 非参数估计, 监督学习, 无监督学习?

参数估计: 先假定研究的问题具有某种数学模型, 如正态分布, 二项分布, 再用已知类别的学习样本估计里面的参数;

非参数估计: 不假定数学模型, 直接用已知类别的学习样本的先验知识直接估计数学模型;

监督学习: 在已知类别样本指导下的学习和训练, 参数估计和非参数估计都属于监督学习。

无监督学习: 不知道样本类别, 只知道样本的某些信息去估计, 如: 聚类分析。

19. (1) 最大似然估计算法思想: 准则, 求解过程

(1) 准则: 第 i 类样本的类条件概率密度:

$$P(X^i/\omega_i) = P(X^i/\omega_i, \theta^i) = P(X^i/\theta^i)$$

原属于 i 类的学习样本为 $X^i = (X_1, X_2, \dots, X_N)^T$ $i=1, 2, \dots, M$

求 θ^i 的最大似然估计就是把 $P(X^i/\theta^i)$ 看成 θ^i 的函数, 求出使它最大时的 θ^i 值

(2) 求解过程:

学习样本独立从总体样本集中抽取的

$$P(X^i/\omega_i, \theta^i) = P(X^i/\theta^i) = \prod_{k=1}^N P(X_k/\theta^i)$$

N 个学习样本出现概率的乘积

$$\text{取对数: } \log \prod_{k=1}^N P(X_k/\theta^i) = \sum_{k=1}^N \log P(X_k/\theta^i)$$

对 θ^i 求导, 并令它为 0:

$$\begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \dots \\ \frac{\partial}{\partial \theta_p} \end{bmatrix} \sum_{k=1}^N \log P(X_k/\theta^i) = 0$$

$$\begin{cases} \sum_{k=1}^N \frac{\partial}{\partial \theta_1} \log P(X_k | \theta^i) = 0 \\ \dots\dots\dots \\ \sum_{k=1}^N \frac{\partial}{\partial \theta_p} \log P(X_k | \theta^i) = 0 \end{cases}$$

利用上式求出 θ^i 的估值 $\hat{\theta}$, 即为 $\theta^i = \hat{\theta}$

(2) 正态分布情况下: μ, Σ 的计算

已知, μ 未知, 估计 μ

$$\mu = \hat{\mu} = \frac{1}{N} \sum_{k=1}^N X_k$$

, μ 均未知

A 一维情况: $n=1$ 对于每个学习样本只有一个特征的简单情况:

$$\hat{\theta}_1 = \hat{\mu}_1 = \frac{1}{N} \sum_{k=1}^N X_k \quad \text{即学习样本的算术平均}$$

$$\hat{\theta}_2 = \hat{\sigma}_1^2 = \frac{1}{N} \sum_{k=1}^N (X_k - \hat{\mu})^2 \quad \text{样本方差}$$

B 多维情况: n 个特征

估计值:

$$\hat{\theta}_1 = \hat{\mu} = \frac{1}{N} \sum_{k=1}^N X_k$$

$$\hat{\theta}_2 = \hat{\Sigma} = \frac{1}{N} \sum_{k=1}^N (X_k - \hat{\mu})(X_k - \hat{\mu})^T$$

20. (1) 贝叶斯估计算法思想: 准则, 求解过程

(A) 准则: 通过对第 i 类学习样本 X^i 的观察, 使概率密度分布 $P(X^i | \theta)$ 转化为后验概率 $P(\theta | X^i)$, 再求贝叶斯估计;

(B) 求解过程: 确定 θ 的先验分布 $P(\theta)$ 待估参数为随机变量。

用第 i 类样本 $X^i = (x_1, x_2, \dots, x_N)^T$ 求出样本的联合概率密度分布 $P(X^i | \theta)$ 它是 θ 的函数。

$$\text{利用贝叶斯公式, 求 } \theta \text{ 的后验概率 } P(\theta | X^i) = \frac{P(X^i | \theta) \cdot P(\theta)}{\int_{\theta} P(X^i | \theta) P(\theta) d\theta}$$

$$\text{求贝叶斯估计 } \hat{\theta} = \int_{\theta} \theta P(\theta | X^i) d\theta$$

(2) 正态分布情况下: μ 的计算

对 μ 的估计为

$$\hat{\mu}_N = \mu_N = \frac{\sigma_0^2}{N\sigma_0^2 + \sigma^2} \sum_{k=1}^N X_k + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0$$

若令 $P(\mu) = N(\mu_0, \sigma_0^2) = N(0, 1)$

$$\hat{\mu}_N = \frac{1}{N+1} \sum_{k=1}^N X_k$$

21 . (1) 贝叶斯学习概念

求出 μ 的后验概率之后，直接去推导总体分布即

$$P(X | x^i) = \int P(X | \theta) P(\theta | x^i) d\theta = \int P(X | \mu) P(\mu | x^i) d\mu$$

当 N , μ_N 就反映了观察到 N 个样本后对 μ 的最好推测，而 σ_N^2 反映了这种推测的不确定性，

N , σ_N^2 , σ^2 随观察样本增加而单调减小，且当 $N \rightarrow \infty$, $\sigma_N^2 \rightarrow 0$

当 $N \rightarrow \infty$, $P(\mu | x^i)$ 越来越尖峰突起； $N \rightarrow \infty$, $P(\mu | x^i)$ 函数，这个过程成为贝叶斯学习。

(2) 正态分布情况下 $P(x | x^i)$ 的计算

(A)一维正态：已知 σ^2 , μ 未知

$$\begin{aligned} P(x | x^i) &= \int P(x | \theta) \cdot P(\theta | x^i) d\theta = \int P(x | \mu) \cdot P(\mu | x^i) d\mu \\ &= \int \frac{1}{\sqrt{2\pi} \sigma} \exp\left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi} \sigma_N} \exp\left[-\frac{1}{2} \left(\frac{\mu - \mu_N}{\sigma_N}\right)^2\right] d\mu \\ &= \frac{1}{2\pi \sigma \sigma_N} \exp\left[-\frac{1}{2} \left(\frac{x - \mu_N}{\sigma_N^2 + \sigma^2}\right)^2\right] \int \exp\left[-\frac{1}{2} \frac{\sigma_N^2 + \sigma^2}{\sigma_N^2 \sigma^2} \left(\mu - \frac{\sigma_N^2 x + \sigma^2 \mu_N}{\sigma_N^2 + \sigma^2}\right)^2\right] d\mu \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\sigma_N^2 + \sigma^2}} \exp\left[-\frac{1}{2} \left(\frac{x - \mu_N}{\sqrt{\sigma_N^2 + \sigma^2}}\right)^2\right] \\ &= N(\mu_N, \sigma_N^2 + \sigma^2) \text{ 为正态函数} \end{aligned}$$

(B)多维正态 (已知 Σ , 估计 μ):

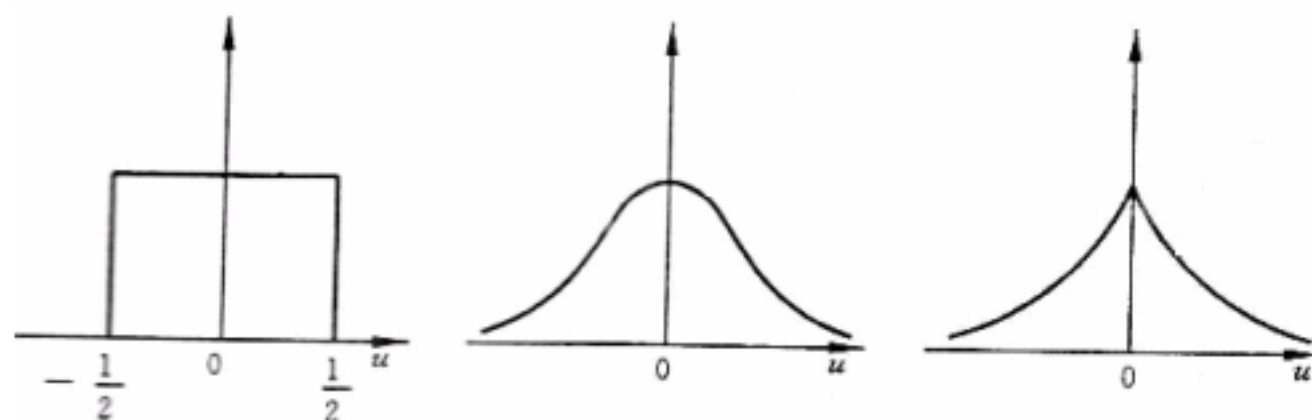
$$\mu_N = \Sigma_0 \left(\Sigma_0 + \frac{1}{N} \Sigma \right)^{-1} \left(\frac{1}{N} \sum_{k=1}^N x_k \right) + \frac{1}{N} \Sigma (\Sigma_0 + \frac{1}{N} \Sigma)^{-1} \mu_0$$

将 μ_N 代入 $P(x | x^i) = \int P(x | \mu) P(\mu | x^i) d\mu$ 就可以设计 Bayes分类器

22 . 非参数估计的条件密度计算公式

(1) Parzen 窗口估计的三种形式，条件密度的计算

(A) 窗口的选择： (A) 方窗函数； (B) 正态窗函数； (C) 指数窗函数



$$\varphi(u) = \begin{cases} 1, & |u| \leq \frac{1}{2} \\ 0, & \text{其他} \end{cases} \quad \varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} u^2\right\} \quad \varphi(u) = \exp\{-|u|\}$$

(B) 条件密度的计算：

$$P_N(x) = \frac{K_N/N}{V_N} = \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \phi\left(\frac{|x - x_i|}{h_N}\right)$$

(2) K-近邻估计的基本思想及用 K-近邻法作后验概率估计的方法

(A) 基本思想：以 x 为中心建立空胞，使 v ，直到捕捉到 K_N 个样本为止。

(B) 用 K-近邻法作后验概率估计的方法：由 K_N 近邻估计知 N 个已知类别样本落入 V_N 内为 K_N 个样本的概率密度估计为

$$P_N(x) = \frac{K_N/N}{V_N}$$

N 个样本落入 V_N 内有 K_N 个， K_N 个样本内有 K_i 个样本属于 ω_i 类，则联合概率密度：

$$P_N(x, \omega_i) = \frac{K_i/N}{V_N} = P(x|\omega_i)P(\omega_i)$$

根据 Bayes 公式可求出后验概率：

$$P_N(\omega_i|x) = \frac{P(x|\omega_i) \cdot P(\omega_i)}{\sum_{i=1}^M P(x|\omega_i) \cdot P(\omega_i)} = \frac{P_N(x, \omega_i)}{\sum_{j=1}^M P_N(x, \omega_j)}$$

后验概率的估计： $P_N(\omega_i|x) = \frac{K_i}{K_N}$

23. 分类与聚类的区别是什么？

分类：用已知类别的样本训练集来设计分类器（监督学习）；

聚类（集群）：用事先不知样本的类别，而利用样本的先验知识来构造分类器（无监督学习）。

24. (1) 聚合聚类（系统聚类）的算法

思想：先把每个样本作为一类，然后根据它们间的相似性和相邻性聚合。

若有 n 个样本：(A) 设全部样本分为 n 类；

(B) 作距离矩阵 $D(0)$ ；

(C) 求最小元素；

(D) 将距离平方最小的元素归为一类；

(E) 以新类从新分类，作距离矩阵 $D(1)$ ；

(F) 若合并的类数没有达到要求，转 (C)，否则停止。

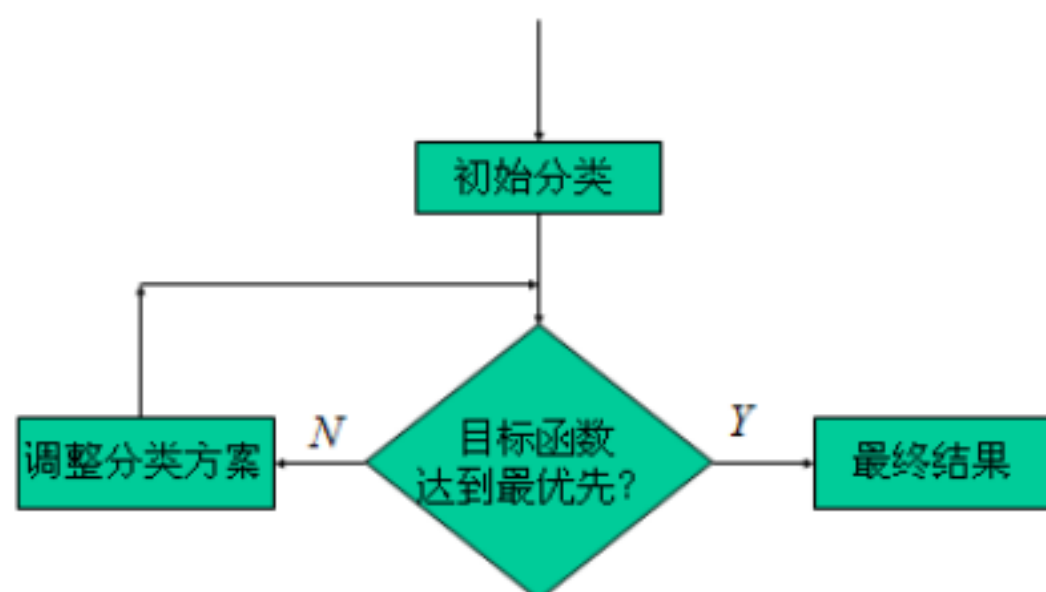
(2) 分解聚类的算法

思想：把全部样本作为一类，然后根据相似性、相邻性分解。

目标函数：两类均值方差

$$E = \frac{N_1 N_2}{N} (\bar{x}_1 - \bar{x}_2)^T (\bar{x}_1 - \bar{x}_2)$$

N ：总样本数， N_1 ：1类样本数 N_2 ：2类样本数， \bar{x}_1, \bar{x}_2 ：两类均值



(3) 动态聚类的算法 (K-均值算法)

先选定某种距离作为样本间的相似性的度量 ;

确定评价聚类结果的准则函数 ;

给出某种初始分类, 用迭代法找出使准则函数取极值的最好的聚类结果。

25. (1) 什么是模糊集, α -水平截集?

(A) 模糊集: 假设论域 $E=\{x\}$ (讨论的区间), 模糊集 A 是由隶属函数 $\mu_A(x)$ 描述。其中, $\mu_A(x)$ 是定义在 E 上在闭区间 $\{0, 1\}$ 中取值的一个函数, 反映 x 对模糊集的隶属程度。

(B) α -水平截集: 设 A 为 $E=\{x\}$ 中的模糊集, 则 $A_\alpha=\{x | \mu_A(x) \geq \alpha\}$ 称为模糊集 A 的 α -水平集, 为阈值在 $(0, 1)$ 间取值。

(2) 什么是模糊集的并, 交, 补运算?

设: A, B 为 $E=\{x\}$ 上的两个模糊集

并集: $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$

交集: $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$

补集: $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$, $\mu_A(x)$, $\mu_B(x)$ 分别为 A 、 B 的隶属函数

(3) 什么是模糊关系及其变换运算?

(A) 模糊关系: 设 U, V 为两个模糊集, 则 u, v 的笛卡儿乘积集记为: $U \times V = \{(u, v) | u \in U, v \in V\}$, (u, v) 是 U, V 元素间的一种无约束搭配, 若把这种搭配加某种限制, U, V 间的这种特殊关系叫模糊关系 R 。

(B) 变换运算: $\tilde{R} = (r_{ij})$ 是 $n \times m$ 维模糊矩阵; $\tilde{S} = (s_{jk})$ 是 $m \times r$ 维模糊矩阵

令 $t_{ik} = \bigvee_{j=1}^m [r_{ij} \wedge s_{jk}]$, $(i = 1, 2, \dots, n; k = 1, 2, \dots, r)$; 式中 “ \vee ” 表示求最大值, “ \wedge ” 表示求最小值

上式表示 \tilde{R} 与 \tilde{S} 的最大 - 最小合成关系。

$\tilde{T} = (t_{ik})$ 为 \tilde{R} 对 \tilde{S} 的复合矩阵, 记作 $\tilde{T} = \tilde{R} \circ \tilde{S}$

(4) 什么是相似关系, 等价关系?

(A) 相似关系: 具有自反性对称性的模糊关系称为相似关系 (或类似关系) ;

(B) 等价关系: 具有自反性、对称性、传递性的模糊关系称为等价关系。

26. 模糊识别方法

(1) 隶属原则识别法的基本思想

设： A_1, A_2, \dots, A_n 是 E 中的 n 个模糊子集， x_0 为 E 中的一个元素，若有隶属函数 $\mu(x_0) = \max(\mu_1(x_0), \mu_2(x_0), \dots, \mu_n(x_0))$ ，则 $x_0 \in \mu_0$ 。则 $x_0 \in A_i$ 。若有了隶属函数 $\mu(x)$ ，我们把隶属函数作为判别函数使用即可。

(2) 择近原则识别法的基本思想

(a) 贴近度的计算：

$$(\tilde{A} \bullet \tilde{B}) = \frac{1}{2} [\tilde{A} \circ \tilde{B} + (1 - \tilde{A} \cup \tilde{B})]$$

式中， $\tilde{A} \circ \tilde{B} = \bigvee_{x \in E} (\tilde{A}(x) \wedge \tilde{B}(x))$ ， $\tilde{A} \cup \tilde{B} = \bigvee_{x \in E} (\tilde{A}(x) \vee \tilde{B}(x))$

分别称为 \tilde{A} 与 \tilde{B} 的内积和外积。

符号“ \vee ”表示求最大，“ \wedge ”表示求最小。

(b) 设： E 上有 n 个模糊子集 $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ 及另一模糊子集 \tilde{B} 。若贴近度

$$(\tilde{B} \bullet \tilde{A}_i) = \max_{1 \leq i \leq n} (\tilde{B} \bullet \tilde{A}_i)$$

则称 \tilde{B} 与 \tilde{A}_i 最贴近。则 $\tilde{B} \in \tilde{A}_i$ 类，这就是择近原则识别方法。

27. 模糊聚类分析方法

1) 基于等价关系

(1) α -水平截阵

(2) 等价划分

(1) 水平截阵： $R_\alpha = [x | \mu(x) \geq \alpha]$

(2) 等价划分：若 R 是 E 上的一个等价关系。则对任意阈值 α ($0 \leq \alpha \leq 1$) 则模糊水平集 R_α 也是 E 上的一个等价关系；由小到大选取阈值 α ($0 \leq \alpha \leq 1$)，将矩阵中相同的行的特征归为一类，得到分类；逐渐增大阈值，则分类增多，知道满足分类数目为止。

2) 基于相似关系

(1) 求传递闭包 \rightarrow 等价

(2) 利用等价关系聚类

(1) 把相似关系（相似矩阵） R 变成等价关系方法为：

取 R 的乘幂为 R^2, R^4, R^8, \dots

若在某一步有 $R^k = R^{2k} = R$ 。

则 R 就是模糊等价关系。且 $R^2 = R \circ R$

$R^4 = R^2 \circ R^2, R^8 = R^4 \circ R^4$

(2) 选择适当 α 值，取等价关系 R 的 α -水平集，根据水平集确定样本的类别

另：

1. 所有作业涉及的计算问题！

2. 分段线性判别方法中的基于凸函数的交方法？

3. 结构模式识别中的形式语言、文法推断、句法分析、自动机理论等问题！

