

2013 模式识别练习题

一、填空题

- 1、模式识别系统的基本构成单元包括：模式采集、特征的选择和提取和模式分类。
- 2、统计模式识别中描述模式的方法一般使用特征矢量；句法模式识别中模式描述方法一般有串、树、网。
- 3、影响层次聚类算法结果的主要因素有计算模式距离的测度、聚类准则、类间距离阈值、预定的类别数目。
- 4、线性判别函数的正负和数值大小的几何意义是正负表示样本点位于判别界面法向量指向的正负半空间中，绝对值正比于样本点与判别界面的距离。
- 5、感知器算法1，H-K 算法2。
(1) 只适用于线性可分的情况；(2) 线性可分、不可分都适用。
- 6、在统计模式分类问题中，聂曼-皮尔逊判决准则主要用于某一种判别错误较另一种判别错误更为重要的情况；最小最大判别准则主要用于先验概率未知的情况。
- 7、“特征个数越多越有利于分类”这种说法正确吗？错误。
特征选择的主要目的是从 n 个特征中选取最有利于分类的 m 个特征 ($m < n$)，以降低特征维数。一般在可分性判断对特征个数具有单调性 和 ($C_n^m \gg n$) 的条件下，可以使用分支定界法以减少计算量。
- 8、散度 J_{ij} 越大，说明 ω_i 类模式与 ω_j 类模式的分布差别越大；
当 ω_i 类模式与 ω_j 类模式的分布相同时， $J_{ij} = 0$ 。

二、选择题

- 1、影响聚类算法结果的主要因素有 (B、C、D)。
A. 已知类别的样本质量；B. 分类准则；C. 特征选取；D. 模式相似性测度
- 2、模式识别中，马式距离较之于欧式距离的优点是 (C、D)。
A. 平移不变性；B. 旋转不变性；C. 尺度不变性；D. 考虑了模式的分布
- 3、影响基本 K-均值算法的主要因素有 (ABD)。
A. 样本输入顺序；B. 模式相似性测度；C. 聚类准则；D. 初始类中心的选取
- 4、位势函数法的积累势函数 $K(x)$ 的作用相当于 Bayes 判决中的 (B D)。

A. 先验概率; B. 后验概率; C. 类概率密度; D. 类概率密度与先验概率的乘积

5、在统计模式分类问题中, 当先验概率未知时, 可以使用 (BD)。

A. 最小损失准则; B. 最小最大损失准则; C. 最小误判概率准则; D. N-P 判决

6、散度 J_D 是根据 (C) 构造的可分性判据。

A. 先验概率; B. 后验概率; C. 类概率密度; D. 信息熵; E. 几何距离

7、似然函数的概型已知且为单峰, 则可用 (ABCDE) 估计该似然函数。

A. 矩估计; B. 最大似然估计; C. Bayes 估计; D. Bayes 学习; E. Parzen 窗法

8、KN 近邻元法较之 Parzen 窗法的优点是 (B)。

A. 所需样本数较少; B. 稳定性较好; C. 分辨率较高; D. 连续性较好

9、从分类的角度讲, 用 **DKLT** 做特征提取主要利用了 **DKLT** 的性质: (A C)。

A. 变换产生的新分量正交或不相关; B. 以部分新的分量表示原矢量均方误差最小; C. 使变换后的矢量能量更集中

10、如果以特征向量的相关系数作为模式相似性测度, 则影响聚类算法结果的主要因素有 (BC)。

A. 已知类别样本质量; B. 分类准则; C. 特征选取; D. 量纲

11、欧式距离具有 (AB); 马式距离具有 (ABCD)。

A. 平移不变性; B. 旋转不变性; C. 尺度缩放不变性; D. 不受量纲影响的特性

12、聚类分析算法属于 (A) ; 判别域代数界面方程法属于 (C) 。

A. 无监督分类; B. 有监督分类; C. 统计模式识别方法; D. 句法模式识别方法

13、若描述模式的特征量为 0-1 二值特征量, 则一般采用 (D) 进行相似性度量。

A. 距离测度; B. 模糊测度; C. 相似测度; D. 匹配测度

14、下列函数可以作为聚类分析中的准则函数的有 (ACD) 。

$$A. J = Tr[S_W^{-1} S_B]; \quad B. J = |S_W S_B^{-1}|; \quad C. J = \sum_{j=1}^c \sum_{i=1}^{n_j} \|\bar{x}_i^{(j)} - \bar{m}_j\|^2; \quad D. J = \sum_{j=1}^c (\bar{m}_j - \bar{m})' (\bar{m}_j - \bar{m})$$

15、Fisher 线性判别函数的求解过程是将 N 维特征矢量投影在 (B) 中进行。

A. 二维空间; B. 一维空间; C. N-1 维空间

16、用 parzen 窗法估计类概率密度函数时, 窗宽过窄导致波动过大的原因是 (BC)。

A. 窗函数幅度过小; B. 窗函数幅度过大; C. 窗口中落入的样本数过少; D. 窗口中落入的样本数过多。

17、如下聚类算法中，属于静态聚类算法的是（AB）。

A. 最大最小距离聚类； B. 层次聚类； C. c-均值聚类。

18、一般，k-NN 最近邻方法在（B）的情况下效果较好。

A. 样本较多但典型性不好； B. 样本较少但典型性好； C. 样本呈团状分布； D. 样本呈链状分布

19、影响 c 均值聚类算法效果的主要因素之一是初始类心的选取，相比较而言，（C）c 个样本作为初始类心较好。

A. 按输入顺序选前； B. 选相距最远的； C. 选分布密度最高处的； D. 随机挑选。

20、类域界面方程法中，能求线性不可分情况下分类问题近似或精确解的方法是（BCD）。

A. 感知器算法； B. 伪逆法； C. 基于二次准则的 H-K 算法； D. 势函数法。

三、简答题

1、试说明 Mahalanobis 距离平方的定义，到某点的 Mahalanobis 距离平方为常数的轨迹的几何意义，它与欧氏距离的区别与联系。

答：Mahalanobis 距离的平方定义为： $r^2(x, u) = (x - u)^T \Sigma^{-1} (x - u)$

其中 x, u 为两个数据， Σ^{-1} 是一个正定对称矩阵（一般为协方差矩阵）。根据定义，距某一点的 Mahalanobis 距离相等点的轨迹是超椭圆，如果是单位矩阵 Σ ，则 Mahalanobis 距离就是通常的欧氏距离。

2、试说明用监督学习与非监督学习两种方法对道路图像中道路区域的划分的基本做法，以说明这两种学习方法的定义与它们间的区别。

答：监督学习方法用来对数据实现分类，分类规则通过训练获得。该训练集由带分类号的数据集组成，因此监督学习方法的训练过程是离线的。

非监督学习方法不需要单独的离线训练过程，也没有带分类号（标号）的训练数据集，一般用来对数据集进行分析，如聚类，确定其分布的主分量等。

就道路图像的分割而言，监督学习方法则先在训练用图像中获取道路像素与非道路像素集，进行分类器设计，然后用所设计的分类器对道路图像进行分割。

使用非监督学习方法，则依据道路路面象素与非道路象素之间的聚类分析进行聚类运算，以实现道路图像的分割。

3、已知一组数据的协方差矩阵为 $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，试问

- (1) 协方差矩阵中各元素的含义。
- (2) 求该数组的两个主分量。
- (3) 主分量分析或称 K-L 变换，它的最佳准则是什么？
- (4) 为什么说经主分量分析后，消除了各分量之间的相关性。

答：协方差矩阵为 $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，则

- (1) 对角元素是各分量的方差，非对角元素是各分量之间的协方差。

- (2) 主分量，通过求协方差矩阵的特征值，用 $\begin{pmatrix} \lambda - 1 & -\frac{1}{2} \\ -\frac{1}{2} & \lambda - 1 \end{pmatrix} = 0$ 得 $(\lambda - 1)^2 = 1/4$ ，则

$$\lambda = \begin{cases} 1/2 \\ 3/2 \end{cases}, \text{ 相应地: } \lambda = 3/2, \text{ 对应特征向量为 } \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \lambda = 1/2, \text{ 对应 } \begin{pmatrix} 1 \\ -1 \end{pmatrix}。$$

这两个特征向量，即为主分量。

- (3) K-L 变换的最佳准则为：

对一组数据进行按一组正交基分解，在只取相同数量分量的条件下，以均方误差计算截尾误差最小。

- (4) 在经主分量分解后，协方差矩阵成为对角矩阵，因而各主分量间相关性消除。

4、试说明以下问题求解是基于监督学习或是非监督学习：

- (1) 求数据集的主分量
- (2) 汉字识别
- (3) 自组织特征映射
- (4) CT 图像的分割

答：(1) 求数据集的主分量是非监督学习方法；

(2) 汉字识别：对待识别字符加上相应类别号——有监督学习方法；

(3) 自组织特征映射——将高维数组按保留近似度向低维映射——非监督学习；

(4) CT 图像分割——按数据自然分布聚类——非监督学习方法；

5、试列举线性分类器中最著名的三种最佳准则以及它们各自的原理。

答：线性分类器三种最优准则：

Fisher 准则：根据两类样本一般类内密集，类间分离的特点，寻找线性分类器最佳的法线向量方向，使两类样本在该方向上的投影满足类内尽可能密集，类间尽可能分开。

这种度量通过类内离散矩阵 S_w 和类间离散矩阵 S_b 实现。

感知准则函数：准则函数以使错分类样本到分界面距离之和最小为原则。

其优点是通过错分类样本提供的信息对分类器函数进行修正，这种准则是人工神经网络多层感知器的基础。

支持向量机：基本思想是在两类线性可分条件下，所设计的分类器界面使两类之间的间隔为最大，它的基本出发点是使期望泛化风险尽可能小。

6、试分析五种常用决策规则思想方法的异同。

答、五种常用决策是：

1. **基于最小错误率的贝叶斯决策**，利用概率论中的贝叶斯公式，得出使得错误率最小的分类规则。
2. **基于最小风险的贝叶斯决策**，引入了损失函数，得出使决策风险最小的分类。当在 0-1 损失函数条件下，基于最小风险的贝叶斯决策变成基于最小错误率的贝叶斯决策。
3. 在限定一类错误率条件下使另一类错误率最小的**两类别决策**。
4. **最大最小决策**：类先验概率未知，考察先验概率变化对错误率的影响，找出使最小贝叶斯奉献最大的先验概率，以这种最坏情况设计分类器。
5. **序贯分类方法**，除了考虑分类造成的损失外，还考虑特征获取造成的代价，先用一部分特

征分类，然后逐步加入性特征以减少分类损失，同时平衡总的损失，以求得最有效益。

7、 1. 什么是特征选择？

2. 什么是 Fisher 线性判别？

答：1. 特征选择就是从一组特征中挑选出一些最有效的特征以达到降低特征空间维数的目的。

2. Fisher 线性判别：可以考虑把 d 维空间的样本投影到一条直线上，形成一维空间，即把维数压缩到一维，这在数学上容易办到，然而，即使样本在 d 维空间里形成若干紧凑的互相分得开的集群，如果把它们投影到一条任意的直线上，也可能使得几类样本混在一起而变得无法识别。但是在一般情况下，总可以找到某个方向，使得在这个方向的直线上，样本的投影能分开得最好。问题是如何根据实际情况找到这条最好的、最易于分类的投影线，这就是 Fisher 算法所要解决的基本问题。

8、写出两类和多类情况下最小风险贝叶斯决策判别函数和决策面方程。

两类问题：判别函数 $g_1(x) = \lambda_{11}p(w_1|x) + \lambda_{12}p(w_2|x)$

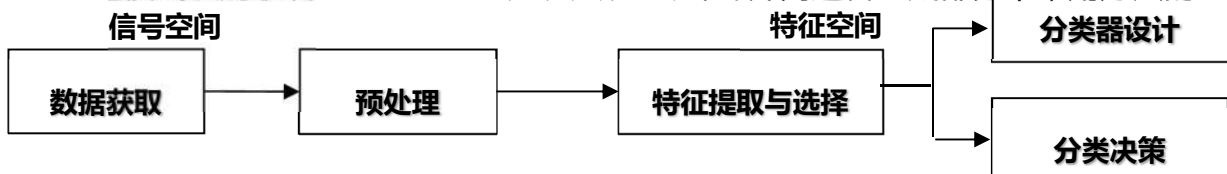
$$g_2(x) = \lambda_{21}p(w_1|x) + \lambda_{22}p(w_2|x)$$

决策面方程： $g_1(x) = g_2(x)$

C 类问题：判别函数 $g_i(x) = \sum_{j=1}^c \lambda_{ij}p(w_j|x)$, $i = 1, 2, \dots, c$

决策面方程： $g_i(x) = g_j(x)$, $i \neq j$, $i = 1, 2, \dots, c$, $j = 1, 2, \dots, c$

9、请论述模式识别系统的主要组成部分及其设计流程，并简述各组成部分中常用方法的主要思想。



信息获取：通过测量、采样和量化，可以用矩阵或向量表示二维图像或波形。

预处理：去除噪声，加强有用的信息，并对输入测量仪器或其他因素造成的退化现象进行复原。

特征选择和提取：为了有效地实现分类识别，就要对原始数据进行变换，得到最能反映分类本质的特征。

分类决策：在特征空间中**用统计方法把识别对象归为某一类。**

10、简述支持向量机的基本思想。

答：SVM 从线性可分情况下的最优分类面发展而来。

最优分类面就是要求分类线不但能将两类正确分开(训练错误率为 0)，且使分类间隔最大。

SVM 考虑寻找一个满足分类要求的超平面，并且使训练集中的点距离分类面尽可能的远，也就是寻找一个分类面使它两侧的空白区域(margin)最大。

过两类样本中离分类面最近的点，且平行于最优分类面的超平面上 H_1 , H_2 的训练样本就叫支持向量。

四、计算题

1、设两类样本的类内散布矩阵分别为 $S_1 = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$, $S_2 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$ ，两类的类心分别为

$m_1 = (2, 0)^T$, $m_2 = (2, 2)^T$ ，试用 fisher 准则求其决策面方程。

解： $S_w = \frac{1}{2}(S_1 + S_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $S_w^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$d(\vec{x}) = (\vec{m}_1 - \vec{m}_2)^T S_w^{-1} (\vec{x} - \frac{\vec{m}_1 + \vec{m}_2}{2}) = (0, -2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (x_1 - 2, x_2 - 1)^T = -2x_2 + 2$$

$$d(\vec{x}) \geq 0 \Rightarrow \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

或 写出决策面方程 $d(\vec{x}) = -x_2 + 1 = 0$

2、已知两个一维模式类别的类概率密度函数为：

$$p(x/\omega_1) = \begin{cases} x & 0 \leq x < 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{其它} \end{cases} \quad p(x/\omega_2) = \begin{cases} x-1 & 1 \leq x < 2 \\ 3-x & 2 \leq x \leq 3 \\ 0 & \text{其它} \end{cases}$$

先验概率 $P(\omega_1)=0.6$, $P(\omega_2)=0.4$,

(1) 求 0-1 代价 Bayes 判决函数;

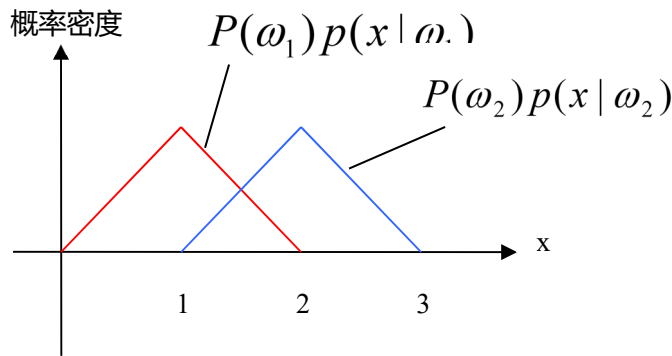
(2) 求总错误概率 $P(e)$;

(3) 判断样本 $\{x_1=1.35, x_2=1.45, x_3=1.55, x_4=1.65\}$ 各属于哪一类别。

答: (1) 基于 0-1 代价 Bayes 判决函数为:

$$\text{当 } \frac{p(x|\omega_1)}{p(x|\omega_2)} \geq \frac{P(\omega_2)}{P(\omega_1)} = \frac{0.4}{0.6} = \frac{2}{3} \approx 0.67 \text{ 时, } x \in w_1, \text{ 否则 } x \in w_2,$$

(2) 总的误判概率 $P(e)$ 为:



$$\text{由 } \frac{2-x}{x-1} = \frac{2}{3}, \text{ 得: } x = \frac{8}{5} = 1.6$$

$$\begin{aligned} P(e) &= P(\omega_1) * \int_{D_2} p(x|\omega_1) dx + P(\omega_2) * \int_{D_1} p(x|\omega_2) dx \\ &= 0.6 * \int_{1.6}^2 (2-x) dx + 0.4 * \int_1^{1.6} (x-1) dx \\ &= 0.12 \end{aligned}$$

(3) $x_1 = 1.35$, $p(x/\omega_1)/p(x/\omega_2) = 0.65/0.35 \approx 1.86 > 0.67$, 所以 $x_1 \in w_1$

$x_2 = 1.45$, $p(x/\omega_1)/p(x/\omega_2) = 0.55/0.45 \approx 1.22 > 0.67$, 所以 $x_2 \in w_1$

$x_3 = 1.55$, $p(x/\omega_1)/p(x/\omega_2) = 0.45/0.55 \approx 0.82 > 0.67$, 所以 $x_3 \in w_1$

$x_4 = 1.65$, $p(x/\omega_1)/p(x/\omega_2) = 0.35/0.65 \approx 0.54 < 0.67$, 所以 $x_4 \in w_2$

3、假设在某个地区细胞识别中正常(w_1)和异常(w_2)两类先验概率分别为 $P(w_1) = 0.8$, $P(w_2) = 0.2$,

现有一待识别的细胞, 其观察值为 x , 从类条件概率密度分布曲线上查得 $P(x/w_1) = 0.25$,

$P(x/w_2) = 0.6$, 并且已知 $\lambda_{11} = 0$, $\lambda_{12} = 6$, $\lambda_{21} = 1$, $\lambda_{22} = 0$

试对该细胞 x 用一下两种方法进行分类：

- (1) 基于最小错误率的贝叶斯决策；
- (2) 基于最小风险的贝叶斯决策；

请分析两种结果的异同及原因。

答：(1) 利用贝叶斯公式，分别计算出 w_1 和 w_2 的后验概率：

$$p(w_1|x) = \frac{p(x|w_1)p(w_1)}{\sum_{j=1}^2 p(x|w_j)p(w_{j1})} = \frac{0.25 \times 0.8}{0.25 \times 0.8 + 0.6 \times 0.2} = 0.625$$

$$p(w_2|x) = 1 - p(w_1|x) = 0.375$$

根据贝叶斯决策规则： $p(w_1|x) = 0.625 > p(w_2|x) = 0.375$ ，所以把 x 归为正常状态。

(2) 根据条件和上面算出的后验概率，计算出条件风险：

$$R(\alpha_1|x) = \sum_{j=1}^2 \lambda_{1j} p(w_j|x) = \lambda_{11} p(w_1|x) + \lambda_{12} p(w_2|x) = 3.125$$

$$R(\alpha_2|x) = \sum_{j=1}^2 \lambda_{2j} p(w_j|x) = \lambda_{21} p(w_1|x) + \lambda_{22} p(w_2|x) = 1.75$$

由于 $R(\alpha_1|x) > R(\alpha_2|x)$ ，即决策为 w_2 的条件风险小于决策为 w_1 的条件风险，因此采取决策行动 α_2 ，即判断待识别的细胞 x 为 w_2 类----异常细胞。

将 (1) 与 (2) 相对比，其分类结果正好相反，这是因为这里影响决策结果的因素又多了一个，

即“损失”；而且两类错误决策所造成的损失相差很悬殊，因此“损失”起了主导作用。

4、有两类样本集 $x_1^1 = [0,0,0]^T$ ， $x_1^2 = [1,0,0]^T$ ， $x_1^3 = [1,0,1]^T$ ， $x_1^4 = [1,1,0]^T$

$x_2^1 = [0,0,1]^T$ ， $x_2^2 = [0,1,0]^T$ ， $x_2^3 = [0,1,1]^T$ ， $x_2^4 = [1,1,1]^T$

(1) 用 K-L 变换求其二维特征空间，并求出其特征空间的坐标轴；

(2) 使用 Fisher 线性判别方法给出这两类样本的分类面。

解： $E[XX^T] = \frac{1}{8} \sum_i x_i x_i^T$

$E[XX^T] = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$, 其对应的特征值和特征向量为:

$\Lambda = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $U = \begin{bmatrix} 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$, 降到二维空间, 取 $U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

对应的坐标: $w_1: \begin{bmatrix} 0 \\ \frac{1}{\sqrt{3}} \\ -2 \\ \frac{1}{\sqrt{6}} \end{bmatrix}$, $w_2: \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$

(2) $m_1 = \begin{bmatrix} 0.75 \\ 0.25 \\ 0.25 \end{bmatrix}$ $m_2 = \begin{bmatrix} 0.25 \\ 0.75 \\ 0.75 \end{bmatrix}$,

$S_1 = \sum_{i=1}^4 (x_i - m_1)(x_i - m_1)^T = \begin{bmatrix} \frac{9}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix} + \begin{bmatrix} \frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix} + \begin{bmatrix} \frac{1}{16} & -\frac{1}{16} & \frac{3}{16} \\ \frac{1}{16} & \frac{1}{16} & -\frac{3}{16} \\ \frac{3}{16} & -\frac{3}{16} & \frac{9}{16} \end{bmatrix}$

$+ \begin{bmatrix} \frac{1}{16} & \frac{3}{16} & -\frac{1}{16} \\ \frac{3}{16} & \frac{9}{16} & -\frac{1}{16} \\ \frac{1}{16} & -\frac{1}{16} & \frac{1}{16} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}$

$S_2 = \sum_{i=1}^4 (x_i - m_2)(x_i - m_2)^T = \begin{bmatrix} \frac{1}{16} & \frac{3}{16} & -\frac{1}{16} \\ \frac{3}{16} & \frac{9}{16} & -\frac{1}{16} \\ \frac{1}{16} & -\frac{1}{16} & \frac{1}{16} \end{bmatrix} + \begin{bmatrix} \frac{1}{16} & -\frac{1}{16} & \frac{3}{16} \\ -\frac{1}{16} & \frac{1}{16} & -\frac{3}{16} \\ \frac{3}{16} & -\frac{3}{16} & \frac{9}{16} \end{bmatrix}$

$+ \begin{bmatrix} \frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\ -\frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix} + \begin{bmatrix} \frac{9}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & \frac{1}{16} & \frac{1}{16} \\ \frac{3}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}$

$$S_w = \frac{1}{2}(S_1 + S_2) = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

$$w^* = S_w^{-1}(m_1 - m_2) = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}, \quad y_0 = \frac{w^{*T} m_1 + w^{*T} m_2}{2} = -1$$

$$\text{所以 判别函数为 } g(x) = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}^T x + 1$$

5、假设两类模式服从如下的正态分布：

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2.5 \end{bmatrix}, \quad P_1 = P_2 = 0.5,$$

求使 $tr(S_w^{-1}S_b)$ 最大化的一维特征空间的变换矢量。

答： $\mu = [0, 0]^T$, $S_w = \frac{1}{2}(\Sigma_1 + \Sigma_2) = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$, $S_w^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0.4 \end{bmatrix}$

$$S_b = \frac{1}{2} \sum_{i=1}^2 (\mu_i - \mu)(\mu_i - \mu)^T = \frac{1}{4}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad S_w^{-1}S_b = \begin{bmatrix} 1 & 1 \\ 0.4 & 0.4 \end{bmatrix}$$

因为 $S_w^{-1}S_b$ 的秩为 1，所以 $S_w^{-1}S_b$ 只有一个非零特征值， W 是 $D \times 1$ 矩阵，即 $W = w$ 。

为求解 $S_w^{-1}S_b$ 的特征值，解方程 $S_w^{-1}S_b w = \lambda_1 w$,

$$\text{即 } \frac{1}{4} S_w^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = \lambda_1 w$$

$$\text{因为 } \frac{1}{4}(\mu_1 - \mu_2)^T w \text{ 为标量, 所以, } w = S_w^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} 2 \\ 0.8 \end{bmatrix}.$$

6、现有样本集 $X=\{(0,0)^T, (0,1)^T, (2,1)^T, (2,3)^T, (3,4)^T, (1,0)^T\}$,

试用 K-means {C-均值} 算法进行聚类分析(类数 $C=2$), 初始聚类中心为 $(0, 0)^T$ 、 $(0, 1)^T$ 。

解:

第一步: 由题意知 $C = 2$, 初始聚类中心为 $Z_{10}=(0,0)^T$, $Z_{20}=(0,1)^T$

第二步: $\|x_1-Z_{10}\|=\|(0,0)^T-(0,0)^T\|=0$

$$\|x_1-Z_{20}\|=\|(0,0)^T-(0,1)^T\|=1$$

因为 $\|x_1-Z_{10}\|<\|x_1-Z_{20}\|$, 所以 $x_1 \in \omega_1$

$$\|x_2-Z_{10}\|=\|(0,1)^T-(0,0)^T\|=1$$

$$\|x_2-Z_{20}\|=\|(0,1)^T-(0,1)^T\|=0$$

因为 $\|x_2-Z_{10}\|>\|x_2-Z_{20}\|$, 所以 $x_2 \in \omega_2$

同理

因为 $\|x_3-Z_{10}\|=5/2>\|x_3-Z_{20}\|=2$, 所以 $x_3 \in \omega_2$

因为 $\|x_4-Z_{10}\|=8/2>\|x_4-Z_{20}\|=5/2$, 所以 $x_4 \in \omega_2$

$$x_5 \in \omega_2, \quad x_6 \in \omega_1$$

由此得到新的类: $\omega_1=\{x_1, x_6\}$ $N_1=2$, $\omega_2=\{x_2, x_3, x_4, x_5\}$ $N_2=4$

第三步: 根据新分成的两类计算新的聚类中心

$$Z_{11}=(x_1+x_6)/2=(0.5, 0)^T$$

$$Z_{21}=(x_2+x_3+x_4+x_5)/4=(1.75, 2.25)^T$$

第四步: 因为新旧聚类中心不等, 转第二步

第二步: 重新计算 $x_1, x_2, x_3, x_4, x_5, x_6$ 到 Z_{11} 、 Z_{21} 的距离, 把它们归为最近聚类中心, 重新分

为两类 $\omega_1=\{x_1, x_2, x_6\}$ $N_1=3$, $\omega_2=\{x_3, x_4, x_5\}$ $N_2=3$

第三步: 根据新分成的两类计算新的聚类中心

$$Z_{12}=(x_1+x_2+x_6)/3=(1/3, 1/3)^T$$

$$Z_{22}=(x_3+x_4+x_5)/3=(7/3, 8/3)^T$$

第四步: 因为新旧聚类中心不等, 转第二步

重新计算 $x_1, x_2, x_3, x_4, x_5, x_6$ 到 Z_{12}, Z_{22} 的距离, 把它们归为最近聚类中心, 重新分为两类

$$\omega_1 = \{x_1, x_2, x_6\}, N_1=3, \quad \omega_2 = \{x_3, x_4, x_5\}, N_2=3$$

和上次聚类结果一样, 计算结束。

7、已知二维样本: $\bar{x}_1=(-1,0)^T, \bar{x}_2=(0,-1)^T, \bar{x}_3=(0,0)^T, \bar{x}_4=(2,0)^T$ 和 $\bar{x}_5=(0,2)^T, \{\bar{x}_1, \bar{x}_2\} \in \omega_1, \{\bar{x}_3, \bar{x}_4, \bar{x}_5\} \in \omega_2$ 。试用感知器算法求出分类决策函数, 并判断 $\bar{x}_6=(2,1)^T$ 属于哪一类?

答: 将样本符号规范化得:

$$\bar{x}_1=(-1,0,1)^T, \bar{x}_2=(0,-1,1)^T, \bar{x}_3=(0,0,-1)^T, \bar{x}_4=(-2,0,-1)^T \text{ 和 } \bar{x}_5=(0,-2,-1)^T,$$

赋初值 $w(1) = (1, 1, 1)^T$, 取增量 $\rho = 1$, 置迭代步数 $k=1$, 则有:

$$k=1, \quad x_k = x_1, \quad d(x_k) = w^T(k)x_k = 0, \quad w(2) = w(1) + x_1 = (0, 1, 2)^T$$

$$k=2, \quad x_k = x_2, \quad d(x_k) = w^T(k)x_k = 1 > 0, \quad w(3) = w(2)$$

$$k=3, \quad x_k = x_3, \quad d(x_k) = w^T(k)x_k = -2 < 0, \quad w(4) = w(3) + x_3 = (0, 1, 1)^T$$

$$k=4, \quad x_k = x_4, \quad d(x_k) = w^T(k)x_k = -1 < 0, \quad w(5) = w(4) + x_4 = (-2, 1, 0)^T$$

$$k=5, \quad x_k = x_5, \quad d(x_k) = w^T(k)x_k = -2 < 0, \quad w(6) = w(5) + x_5 = (-2, -1, -1)^T$$

$$k=6, \quad x_k = x_1, \quad d(x_k) = w^T(k)x_k = 1 > 0, \quad w(7) = w(6)$$

$$k=7, \quad x_k = x_2, \quad d(x_k) = w^T(k)x_k = 0, \quad w(8) = w(7) + x_2 = (-2, -2, 0)^T$$

$$k=8, \quad x_k = x_3, \quad d(x_k) = w^T(k)x_k = 0, \quad w(9) = w(8) + x_3 = (-2, -2, -1)^T$$

$$k=9, \quad x_k = x_4, \quad d(x_k) = w^T(k)x_k = 5 > 0, \quad w(10) = w(9)$$

$$k=10, \quad x_k = x_5, \quad d(x_k) = w^T(k)x_k = 5 > 0, \quad w(11) = w(10)$$

$$k=11, \quad x_k = x_1, \quad d(x_k) = w^T(k)x_k = 1 > 0, \quad w(12) = w(11)$$

$$k=12, \quad x_k = x_2, \quad d(x_k) = w^T(k)x_k = 5 > 0, \quad w(13) = w(12)$$

$$k=13, \quad x_k = x_3, \quad d(x_k) = w^T(k)x_k = 5 > 0, \quad w(14) = w(13)$$

所以, 最后收敛于 $w = (-2, -2, -1)^T$, 分类决策函数为 $d(x) = -2x_1 - 2x_2 - 1$

将 $x_6 = (2, 1)^T$ 代入决策函数, $d(x) = -7 < 0$, 所以 $\bar{x}_6 \in \omega_2$

8、两类样本的均值矢量分别为 $m_1=(4, 2)^T$ 和 $m_2=(-4, -2)^T$,

$$\text{协方差矩阵分别为: } C_1 = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, C_2 = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}, \text{ 两类的先验概率相等,}$$

试求一维特征提取矩阵。

9、设语言 $L(G)$ 的正样本集 $R^+ = \{101, 111\}$ ，试推断出余码文法 G_c 。句法模式识别 ppt 32 页

填空题

- 1、模式识别系统的基本构成单元包括：模式采集、特征选择与提取 和 模式分类。
- 2、统计模式识别中描述模式的方法一般使用 特征矢量；句法模式识别中模式描述方法一般有 串、树、网。
- 3、影响层次聚类算法结果的主要因素有 计算模式距离的测度、聚类准则、类间距离门限、预定的类别数目。
- 4、线性判别函数的正负和数值大小的几何意义是 正（负）表示样本点位于判别界面法向量指向的正（负）半空间中；绝对值正比于样本点到判别界面的距离。
- 5、感知器算法 1。
(1) 只适用于线性可分的情况；(2) 线性可分、不可分都适用。
- 6、在统计模式分类问题中，聂曼-皮尔逊判决准则主要用于 某一种判决错误较另一种判决错误更为重要 情况；最小最大判别准则主要用于 先验概率未知的 情况。
- 7、“特征个数越多越有利于分类”这种说法正确吗？错误。
特征选择的主要目的是 从 n 个特征中选出最有利于分类的 m 个特征 ($m < n$)，以降低特征维数。一般在 可分性判据对特征个数具有单调性 和 ($C_n^m > n$) 的条件下，可以使用分支定界法以减少计算量。
- 8、散度 J_{ij} 越大，说明 ω_i 类模式与 ω_j 类模式的分布 差别越大；
当 ω_i 类模式与 ω_j 类模式的分布相同时， $J_{ij} = 0$ 。

选择题

- 1、影响聚类算法结果的主要因素有 (**BCD**)。
A. 已知类别的样本质量 B. 分类准则 C. 特征选取 D. 模式相似性测度
- 2、模式识别中，马氏距离较之于欧式距离的优点是 (**CD**)。
A. 平移不变性 B. 旋转不变性 C. 尺度不变性 D. 考虑了模式的分布
- 3、影响基本 K-均值算法的主要因素有 (**DAB**)。
A. 样本输入顺序 B. 模式相似性测度 C. 聚类准则 D. 初始类中心的选取
- 4、在统计模式分类问题中，当先验概率未知时，可以使用 (**BD**)。
A. 最小损失准则 B. 最小最大损失准则 C. 最小误判概率准则 D. N-P 判决

5、散度 J_D 是根据 (C) 构造的可分性判据。

A. 先验概率 B. 后验概率 C. 类概率密度 D. 信息熵 E. 几何距离

6、如果以特征向量的相关系数作为模式相似性测度, 则影响聚类算法结果的主要因素有 (BC)。

A. 已知类别样本质量 B. 分类准则 C. 特征选取 D. 量纲

7、欧式距离具有 (AB); 马式距离具有 (ABCD)。

A. 平移不变性 B. 旋转不变性 C. 尺度缩放不变性 D. 不受量纲影响的特性

8、聚类分析算法属于 (A); 判别域代数界面方程法属于 (C)。

A. 无监督分类 B. 有监督分类 C. 统计模式识别方法 D. 句法模式识别方法

9、下列函数可以作为聚类分析中的准则函数的有 (ACD)。

A. $J = Tr[S_W^{-1} S_B]$ B. $J = |S_W S_B^{-1}|$ C. $J = \sum_{j=1}^c \sum_{i=1}^{n_j} \|\bar{x}_i^{(j)} - \bar{m}_j\|^2$ D. $J = \sum_{j=1}^c (\bar{m}_j - \bar{m})'(\bar{m}_j - \bar{m})$

10、Fisher 线性判别函数的求解过程是将 N 维特征矢量投影在 (B) 中进行。

A. 二维空间 B. 一维空间 C. N-1 维空间

简答题

一、试问“模式”与“模式类”的含义。如果一位姓王的先生是位老年人, 试问“王先生”和“老头”谁是模式, 谁是模式类?

答: 在模式识别学科中, 就“模式”与“模式类”而言, 模式类是一类事物的代表, 概念或典型, 而“模式”则是某一事物的具体体现, 如“老头”是模式类, 而王先生则是“模式”是“老头”的具体化。

二、试说明 Mahalanobis 距离平方的定义, 到某点的 Mahalanobis 距离平方为常数的轨迹的几何意义, 它与欧氏距离的区别与联系。

答: Mahalanobis 距离的平方定义为: $r^2(x, u) = (x - u)^T \Sigma^{-1} (x - u)$

其中 x, u 为两个数据, Σ^{-1} 是一个正定对称矩阵 (一般为协方差矩阵)。根据定义, 距某一点的 Mahalanobis 距离相等点的轨迹是超椭圆, 如果是单位矩阵 Σ , 则 Mahalanobis 距离就是通常的欧氏距离。

三、试说明用监督学习与非监督学习两种方法对道路图像中道路区域的划分的基本做法, 以说明这两种学习方法的定义与它们间的区别。

答: 监督学习方法用来对数据实现分类, 分类规则通过训练获得。该训练集由带分类号的数据集组成, 因此监督学习方法的训练过程是离线的。

非监督学习方法不需要单独的离线训练过程, 也没有带分类号 (标号) 的训练数据集, 一般用来对数据集进行分析, 如聚类, 确定其分布的主分量等。

就道路图像的分割而言，监督学习方法则先在训练用图像中获取道路象素与非道路象素集，进行分类器设计，然后用所设计的分类器对道路图像进行分割。

使用非监督学习方法，则依据道路路面象素与非道路象素之间的聚类分析进行聚类运算，以实现道路图像的分割。

四、试述动态聚类与分级聚类这两种方法的原理与不同。

答：动态聚类是指对当前聚类通过迭代运算改善聚类；

分级聚类则是将样本个体，按相似度标准合并，随着相似度要求的降低实现合并。

五、已知一组数据的协方差矩阵为 $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，试问

1. 协方差矩阵中各元素的含义。
2. 求该数组的两个主分量。
3. 主分量分析或称 K-L 变换，它的最佳准则是什么？
4. 为什么说经主分量分析后，消除了各分量之间的相关性。

答：协方差矩阵为 $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，则

1. 对角元素是各分量的方差，非对角元素是各分量之间的协方差。

2. 主分量，通过求协方差矩阵的特征值，用 $\begin{pmatrix} \lambda - 1 & -\frac{1}{2} \\ -\frac{1}{2} & \lambda - 1 \end{pmatrix} = 0$ 得 $(\lambda - 1)^2 = 1/4$ ，则 $\lambda = \begin{cases} 1/2 \\ 3/2 \end{cases}$ ，

相应的： $\lambda = 3/2$ ，对应特征向量为 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ， $\lambda = 1/2$ ，对应 $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ 。

这两个特征向量，即为主分量。

3. K-L 变换的最佳准则为：对一组数据进行按一组正交基分解，在只取相同数量分量的条件下，以均方误差计算截尾误差最小。

4. 在经主分量分解后，协方差矩阵成为对角矩阵，因而各主分量间相关性消除。

六、试列举线性分类器中最著名的三种最佳准则以及它们各自的原理。

答：线性分类器三种最优准则：

Fisher 准则：根据两类样本一般类内密集，类间分离的特点，寻找线性分类器最佳的法线向量方向，使两类样本在该方向上的投影满足类内尽可能密集，类间尽可能分开。

这种度量通过类内离散矩阵 S_w 和类间离散矩阵 S_b 实现。

感知准则函数：准则函数以使错分类样本到分界面距离之和最小为原则。

其优点是通过错分类样本提供的信息对分类器函数进行修正，这种准则是人工神经网络多层感知器的基础。

支持向量机：基本思想是在两类线性可分条件下，所设计的分类器界面使两类之间的间隔为最大，它的基本出发点是使期望泛化风险尽可能小。

七、对一副道路图像，希望把道路部分划分出来，可以采用以下两种方法：

1. 在该图像中分别在道路部分与非道路部分画出一个窗口，把在这两个窗口中的像素数据作为训练集，用 Fisher 准则方法求得分类器参数，再用该分类器对整幅图进行分类。

2. 将整幅图的每个像素的属性记录在一张数据表中，然后用某种方法将这些数据按它们的自然分布状况划分成两类。因此每个像素就分别得到相应的类别号，从而实现了道路图像的分割。

试问以上两种方法哪一种是监督学习，哪个是非监督学习？

答：

第一种方法中标记了两类样本的标号，需要人手工干预训练过程，属于监督学习方法；

第二种方法只是依照数据的自然分布，把它们划分成两类，属于非监督学习方法。

八、试分析五种常用决策规则思想方法的异同。

答、五种常用决策是：

1. **基于最小错误率的贝叶斯决策**，利用概率论中的贝叶斯公式，得出使得错误率最小的分类规则。

2. **基于最小风险的贝叶斯决策**，引入了损失函数，得出使决策风险最小的分类。当在 0-1 损失函数条件下，基于最小风险的贝叶斯决策变成基于最小错误率的贝叶斯决策。

3. 在限定一类错误率条件下使另一类错误率最小的两类别决策。

4. **最大最小决策**：类先验概率未知，考察先验概率变化对错误率的影响，找出使最小贝叶斯奉献最大的先验概率，以这种最坏情况设计分类器。

5. **序贯分类方法**，除了考虑分类造成的损失外，还考虑特征获取造成的代价，先用一部分特征分类，然后逐步加入性特征以减少分类损失，同时平衡总的损失，以求得最有效益。

九、假设在某个地区细胞识别中正常(w_1)和异常(w_2)两类先验概率分别为 $P(w_1)=0.9$ ， $P(w_2)=0.1$ ，

现有一待识别的细胞，其观察值为 x ，从类条件概率密度分布曲线上查得 $P(x/w_1)=0.2$ ，

$P(x/w_2)=0.4$ ，并且已知 $\lambda_{11}=0$ ， $\lambda_{12}=6$ ， $\lambda_{21}=1$ ， $\lambda_{22}=0$

试对该细胞 x 用一下两种方法进行分类：

1. 基于最小错误率的贝叶斯决策；
2. 基于最小风险的贝叶斯决策；

请分析两种结果的异同及原因。

答：利用贝叶斯公式，分别计算出 w_1 和 w_2 的后验概率：

$$p(w_1|x) = \frac{p(x|w_1)p(w_1)}{\sum_{j=1}^2 p(x|w_j)p(w_j)} = \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.818$$

$$p(w_2|x) = 1 - p(w_1|x) = 0.182$$

根据贝叶斯决策规则： $p(w_1|x) = 0.818 > p(w_2|x) = 0.182$ ，所以把 x 归为正常状态。

(2) 根据条件和上面算出的后验概率，计算出条件风险：

$$R(\alpha_1|x) = \sum_{j=1}^2 \lambda_{1j} p(w_j|x) = \lambda_{12} p(w_2|x) = 1.092$$

$$R(\alpha_2|x) = \sum_{j=1}^2 \lambda_{2j} p(w_j|x) = \lambda_{21} p(w_1|x) = 0.818$$

由于 $R(\alpha_1|x) > R(\alpha_2|x)$ ，即决策为 w_1 的条件风险小于决策为 w_2 的条件风险，因此采取决策行动 α_2 ，即判断待识别的细胞 x 为 w_2 类——异常细胞。

将 1 与 2 相对比，其分类结果正好相反，因为这里影响决策结果的因素多了“损失”；而且两类错误决策所造成的损失相差很悬殊，因此“损失”起了主导作用。

十、既然有线性判别函数，为什么还要引进非线性判别函数？试分析由“线性判别函数”向“非线性判别函数”推广的思想和方法。

答：实际中有很多模式识别问题并不是线性可分的，这时就需要采用非线性分类器，比如当两类样本本不具有多峰性质并互相交错时，简单的线性判别函数往往会带来较大的分类错误。这时，树分类器作为一种分段线性分类器，常常能有效地应用于这种情况。

十一、1. 什么是特征选择？

2. 什么是 Fisher 线性判别?

答: 1. 特征选择就是从一组特征中挑选出一些最有效的特征以达到降低特征空间维数的目的。

2. Fisher 线性判别: 可以考虑把 d 维空间的样本投影到一条直线上, 形成一维空间, 即把维数压缩到一维, 这在数学上容易办到, 然而, 即使样本在 d 维空间里形成若干紧凑的互相分得开的集群, 如果把它们投影到一条任意的直线上, 也可能使得几类样本混在一起而变得无法识别。但是在一般情况下, 总可以找到某个方向, 使得在这个方向的直线上, 样本的投影能分开得最好。问题是如何根据实际情况找到这条最好的、最易于分类的投影线, 这就是 Fisher 算法所要解决的基本问题。

十二、写出两类和多类情况下最小风险贝叶斯决策判别函数和决策面方程。

两类问题: 判别函数 $g_1(x) = \lambda_{11}p(w_1|x) + \lambda_{12}p(w_2|x)$

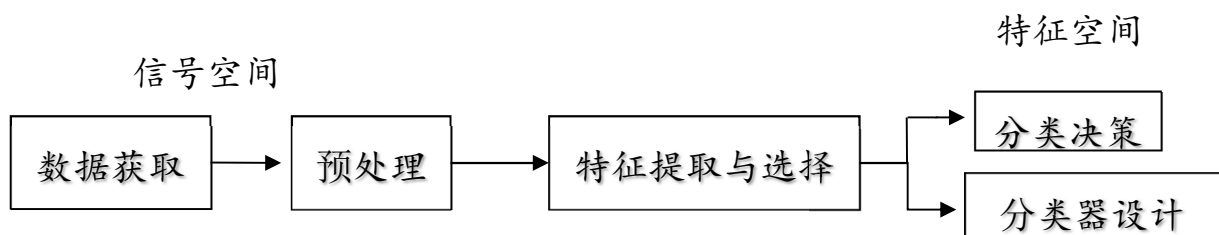
$$g_2(x) = \lambda_{21}p(w_1|x) + \lambda_{22}p(w_2|x)$$

决策面方程: $g_1(x) = g_2(x)$

C 类问题: 判别函数 $g_i(x) = \sum_{j=1}^c \lambda_{ij}p(w_j|x)$, $i = 1, 2, \dots, c$

决策面方程: $g_i(x) = g_j(x)$, $i \neq j$, $i = 1, 2, \dots, c$, $j = 1, 2, \dots, c$

十三、请论述模式识别系统的主要组成部分及其设计流程, 并简述各组成部分中常用方法的主要思想。



信息获取: 通过测量、采样和量化, 可以用矩阵或向量表示二维图像或波形。

预处理: 去除噪声, 加强有用的信息, 并对输入测量仪器或其他因素造成的退化现象进行复原。

特征选择和提取: 为了有效地实现分类识别, 就要对原始数据进行变换, 得到最能反映分类本质的特征。

分类决策: 在特征空间中用统计方法把识别对象归为某一类。

十四、有两类样本集

$$x_1^1 = [0, 0, 0]^T, \quad x_1^2 = [1, 0, 0]^T, \quad x_1^3 = [1, 0, 1]^T, \quad x_1^4 = [1, 1, 0]^T$$

$$x_2^1 = [0, 0, 1]^T, \quad x_2^2 = [0, 1, 0]^T, \quad x_2^3 = [0, 1, 1]^T, \quad x_2^4 = [1, 1, 1]^T$$

1. 用 K-L 变换求其二维特征空间，并求出其特征空间的坐标轴；
2. 使用 Fisher 线性判别方法给出这两类样本的分类面。

$$E[XX^T] = \frac{1}{8} \sum_i x_i x_i^T$$

答： $E[XX^T] = \frac{1}{8} \sum_i x_i x_i^T = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$ ，其对应的特征值为： $\Lambda = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

特征向量为 $U = \begin{bmatrix} 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ ，降到二维空间，取 $U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

对应的坐标： $w_1 : \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix}$ ， $w_2 : \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{3}} \\ 0 \end{bmatrix}$

(2) $m_1 = \begin{bmatrix} 0.75 \\ 0.25 \\ 0.25 \end{bmatrix}$ $m_2 = \begin{bmatrix} 0.25 \\ 0.75 \\ 0.75 \end{bmatrix}$ ， $S_1 = \sum_{i=1}^4 (x_i - m_1)(x_i - m_1)^T = \frac{1}{12} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$ ，

$$S_2 = \sum_{i=1}^4 (x_i - m_2)(x_i - m_2)^T = \frac{1}{12} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}, \quad S_w = \frac{1}{2}(S_1 + S_2) = \frac{1}{12} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

$$w^* = S_w^{-1}(m_1 - m_2) = \begin{bmatrix} 6 & -6 & -6 \end{bmatrix}^T, \quad y_0 = \frac{w^{*T} m_1 + w^{*T} m_2}{2} = -3$$

所以判别函数为： $g(x) = \begin{bmatrix} 6 & -6 & -6 \end{bmatrix}^T x + 3$

十五、假设两类模式服从如下的正态分布：

$$\mu_1 = [1, 3, -1]^T, \quad \mu_2 = [-1, -1, 1]^T, \quad \Sigma_1 = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

分别求出使 $\text{tr}(S_w^{-1} S_b)$ 最大化的一维特征空间的变换矢量。

答: $\mu = [0, 1, 0]^T$, $S_b = \frac{1}{2} \sum_{i=1}^2 (\mu_i - \mu)(\mu_i - \mu)^T = \frac{1}{4} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = \frac{1}{16} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & 2 & 1 \end{bmatrix}$

$$S_w = \frac{1}{2} (\Sigma_1 + \Sigma_2) = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad S_w^{-1} = \frac{1}{8} \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \quad S_w^{-1} S_b = \frac{1}{8 \times 16} \begin{bmatrix} 1 & 2 & -1 \\ 5 & 10 & -5 \\ -8 & 16 & 8 \end{bmatrix}$$

因为 $S_w^{-1} S_b$ 的秩为 1, 所以 $S_w^{-1} S_b$ 只有一个非零特征值, W 是 $D \times 1$ 矩阵, 即 $W = w$ 。

$$s_w = \frac{1}{2} (\Sigma_1 + \Sigma_2) = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, \quad s_w^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0.4 \end{bmatrix}$$

求 $S_w^{-1} S_b$ 的特征值, 解方程 $S_w^{-1} S_b w = \lambda_1 w$, 即 $\frac{1}{4} S_w^{-1} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = \lambda_1 w$

$$\because \frac{1}{4} (\mu_1 - \mu_2)^T w \text{ 为标量}, \therefore w = S_w^{-1} (\mu_1 - \mu_2) = \left[\frac{1}{4}, \frac{5}{4}, -2 \right]^T。$$

十六、设两个家庭, 每家 3-5 人, 选每个人的一张照片, 共 8 张, 混放在一起, 将照片两两对照, 得出描述其“相似程度”的模糊关系矩阵。要求按相似程度聚类, 希望把二个家庭分开。

r_{ij}	1	2	3	4	5	6	7	8
1	1							
2	0	1						
3	0	0	1					
4	0	0.8	0	1				
5	0.5	0	0.2	0	1			
6	0	0.8	0	0.4	0	1		
7	0.4	0.2	0.2	0	0.8	0	1	
8	0	0.5	0.2	0	0	0.8	0	1

十七、设 $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, 标准模型由以下模糊集合表示:

$$\underset{\sim}{A}_1 = \{(1.0, x_1), (0.8, x_2), (0.5, x_3), (0.4, x_4), (0.0, x_5), (0.1, x_6)\}$$

$$\underset{\sim}{A}_2 = \{(0.0, x_1), (1.0, x_2), (0.2, x_3), (0.7, x_4), (0.5, x_5), (0.8, x_6)\}$$

$$\underset{\sim}{A}_3 = \{(0.8, x_1), (0.2, x_2), (0, x_3), (0.5, x_4), (1.0, x_5), (0.7, x_6)\}$$

$$\underline{A}_4 = \{(0.5, x_1), (0.7, x_2), (0.8, x_3), (0, x_4), (0.5, x_5), (1.0, x_6)\}$$

现有一待识别的模型

$$\underline{B} = \{(0.7, x_1), (0.2, x_2), (0.1, x_3), (0.4, x_4), (1.0, x_5), (0.8, x_6)\}$$

采用海明贴近度计算， \underline{B} 与哪个标准模型最相近？

解：海明贴近度： $\sigma_H(\underline{A}, \underline{B}) = 1 - \frac{1}{n} \sum_{i=1}^n |\mu_{\underline{A}}(x_i) - \mu_{\underline{B}}(x_i)|$

$$\sigma_H(\underline{A}_1, \underline{B}) = 1 - \frac{1}{6} (0.3 + 0.6 + 0.4 + 0 + 1 + 0.7) = 1 - \frac{3}{6} = 0.5$$

$$\sigma_H(\underline{A}_2, \underline{B}) = 1 - \frac{1}{6} (0.7 + 0.8 + 0.1 + 0.3 + 0.5 + 0) = 1 - \frac{2.4}{6} = 0.6$$

$$\sigma_H(\underline{A}_3, \underline{B}) = 1 - \frac{1}{6} (0.1 + 0 + 0.1 + 0.1 + 0 + 0.1) = 1 - \frac{0.4}{6} = 0.93$$

$$\sigma_H(\underline{A}_4, \underline{B}) = 1 - \frac{1}{6} (0.2 + 0.5 + 0.7 + 0.4 + 0.5 + 0.2) = 1 - \frac{2.5}{6} = 0.58$$

$$\therefore \max(0.5, 0.6, 0.93, 0.58) = 0.93 = \sigma_H(\underline{A}_3, \underline{B})$$

$\therefore \underline{B}$ 与 \underline{A}_3 最相似。

十八、 设论域 $X = \{x_1, x_2, x_3, x_4, x_5\}$ ，给定模糊关系矩阵：

$$\mathbf{R} = \begin{pmatrix} 1 & 0.48 & 0.62 & 0.41 & 0.47 \\ 0.48 & 1 & 0.48 & 0.41 & 0.47 \\ 0.62 & 0.48 & 1 & 0.41 & 0.47 \\ 0.41 & 0.41 & 0.41 & 1 & 0.41 \\ 0.47 & 0.47 & 0.47 & 0.41 & 1 \end{pmatrix} \text{ 要求按不同 } \lambda \text{ 水平分类。}$$

解：矩阵显然具有自反性、对称性。计算 $\mathbf{R} \circ \mathbf{R}$ ：

$$\mathbf{R} \circ \mathbf{R} = \begin{pmatrix} 1 & 0.48 & 0.62 & 0.41 & 0.47 \\ 0.48 & 1 & 0.48 & 0.41 & 0.47 \\ 0.62 & 0.48 & 1 & 0.41 & 0.47 \\ 0.41 & 0.41 & 0.41 & 1 & 0.41 \\ 0.47 & 0.47 & 0.47 & 0.41 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 0.48 & 0.62 & 0.41 & 0.47 \\ 0.48 & 1 & 0.48 & 0.41 & 0.47 \\ 0.62 & 0.48 & 1 & 0.41 & 0.47 \\ 0.41 & 0.41 & 0.41 & 1 & 0.41 \\ 0.47 & 0.47 & 0.47 & 0.41 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.48 & 0.62 & 0.41 & 0.47 \\ 0.48 & 1 & 0.48 & 0.41 & 0.47 \\ 0.62 & 0.48 & 1 & 0.41 & 0.47 \\ 0.41 & 0.41 & 0.41 & 1 & 0.41 \\ 0.47 & 0.47 & 0.47 & 0.41 & 1 \end{pmatrix} = \mathbf{R}$$

$$\therefore \mathbf{R} \circ \mathbf{R} = \mathbf{R}$$

$\therefore \mathbf{R}$ 为一模糊等价矩阵，可据不同 λ 水平分类。

$$1) \lambda = 1: \quad x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$$

$$R_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix}$$

此时共分五类： $\{x_1\}$ 、 $\{x_2\}$ 、 $\{x_3\}$ 、 $\{x_4\}$ 、 $\{x_5\}$ ，“最细”的分类。

$$2) \lambda = 0.62: R_{0.62} = \begin{matrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} & \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} \end{matrix} \quad \text{此时分为 4 类: } \{x_1, x_3\}, \{x_2\}, \{x_4\}, \{x_5\}。$$

$$3) \lambda = 0.48: R_{0.48} = \begin{matrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} & \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} \end{matrix} \quad \text{此时分为 3 类: } \{x_1, x_2, x_3\}, \{x_4\}, \{x_5\}。$$

$$4) \lambda = 0.47: R_{0.47} = \begin{matrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} & \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} \end{matrix} \quad \text{此时分为 2 类: } \{x_1, x_2, x_3, x_5\}, \{x_4\}。$$

$$5) \lambda = 0.41: R_{0.41} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{此时五个元素合为 1 类, 即最粗的分类。}$$

十九、 现有样本集 $X = \{(0,0)^T, (0,1)^T, (2,1)^T, (2,3)^T, (3,4)^T, (1,0)^T\}$,

试用 C-均值算法进行聚类分析(类数 $C=2$), 初始聚类中心为 $(0,0)^T$ 、 $(0,1)^T$ 。

解:

第一步: 由题意知 $C=2$, 初始聚类中心为 $Z_{10}=(0,0)^T$, $Z_{20}=(0,1)^T$

第二步: $\|x_1 - Z_{10}\| = \|(0,0)^T - (0,0)^T\| = 0$

$$\|x_1 - Z_{20}\| = \|(0,0)^T - (0,1)^T\| = 1$$

因为 $\|x_1 - Z_{10}\| < \|x_1 - Z_{20}\|$, 所以 $x_1 \in \omega_1$

$$\|x_2 - Z_{10}\| = \|(0, 1)^T - (0, 0)^T\| = 1$$

$$\|x_2 - Z_{20}\| = \|(0, 1)^T - (0, 1)^T\| = 0$$

因为 $\|x_2 - Z_{10}\| > \|x_2 - Z_{20}\|$, 所以 $x_2 \in \omega_2$

同理

因为 $\|x_3 - Z_{10}\| = 5/2 > \|x_3 - Z_{20}\| = 2$, 所以 $x_3 \in \omega_2$

因为 $\|x_4 - Z_{10}\| = 8/2 > \|x_4 - Z_{20}\| = 5/2$, 所以 $x_4 \in \omega_2$

$$x_5 \in \omega_2, \quad x_6 \in \omega_1$$

由此得到新的类: $\omega_1 = \{x_1, x_6\}$ $N_1 = 2$, $\omega_2 = \{x_2, x_3, x_4, x_5\}$ $N_2 = 4$

第三步: 根据新分成的两类计算新的聚类中心

$$Z_{11} = (x_1 + x_6)/2 = (0.5, 0)^T$$

$$Z_{21} = (x_2 + x_3 + x_4 + x_5)/4 = (1.75, 2.25)^T$$

第四步: 因为新旧聚类中心不等, 转第二步

第二步: 重新计算 $x_1, x_2, x_3, x_4, x_5, x_6$ 到 Z_{11}, Z_{12} 的距离, 把它们归为最近聚类中心, 重新分

为两类 $\omega_1 = \{x_1, x_2, x_6\}$ $N_1 = 3$, $\omega_2 = \{x_3, x_4, x_5\}$ $N_2 = 3$

第三步: 根据新分成的两类计算新的聚类中心

$$Z_{12} = (x_1 + x_2 + x_6)/3 = (1/3, 1/3)^T$$

$$Z_{22} = (x_3 + x_4 + x_5)/3 = (7/3, 8/3)^T$$

第四步: 因为新旧聚类中心不等, 转第二步

重新计算 $x_1, x_2, x_3, x_4, x_5, x_6$ 到 Z_{12}, Z_{22} 的距离, 把它们归为最近聚类中心, 重新分为两类

$\omega_1 = \{x_1, x_2, x_6\}$ $N_1 = 3$, $\omega_2 = \{x_3, x_4, x_5\}$ $N_2 = 3$

和上次聚类结果一样, 计算结束。

二十. 设语言 $L(G)$ 的正样本集 $R^+ = \{101, 111\}$, 试推断出余码文法 G_c 。

答: **第一步:** 由 R^+ 得 G_c 的终止符集 V_T , 为 $V_T = \{0, 1\}$ 。

第二步: 求出 R^+ 的全部余码, 并组成非终止符集 V_N 。

R^+ 的全部余码为

$$D_\lambda R^+ = \{101, 111\}, \quad D_1 R^+ = \{01, 11\}$$

$$D_{10} R^+ = \{1\}, \quad D_{11} R^+ = \{1\}$$

$$D_{101} R^+ = \{\lambda\}, \quad D_{111} R^+ = \{\lambda\}$$

$$\text{合并:} \quad S = D_\lambda R^+ = \{101, 111\},$$

$$U_1 = D_1 R^+ = \{01, 11\}, \quad U_2 = D_{10} R^+ = \{1\}$$

$$\text{所以 } V_N = \{S, U_1, U_2\}$$

第三步：建立生成式集 P 。

由 $D_1S = U_1$ ，有生成式 $S \rightarrow 1U_1$ ；

由 $D_1U_1 = U_2$ ，有生成式 $U_1 \rightarrow 1U_2$ ；

由 $D_1U_2 = \lambda$ ，有生成式 $U_2 \rightarrow 1$ ；

由 $D_0U_1 = U_2$ ，有生成式 $U_1 \rightarrow 0U_2$ 。

\therefore 余码文法 G_c ：

$$G_c = (V_N, V_T, P, S), \quad V_N = \{S, U_1, U_2\}, \quad V_T = \{0, 1\}$$

$$P: S \rightarrow 1U_1, \quad U_1 \rightarrow 1U_2, \quad U_1 \rightarrow 0U_2, \quad U_2 \rightarrow 1$$