

# 模式识别

复习资料 2012

உருவகம்



# 中国科学技术大学

## 2006—2007 学年第一学期考试试题

课程名称：模式识别

姓名：郝 学号： 得分：

### 1. 是非判断题 (每小问 2.0 分, 共 20 分)

- ✓ (1) 感知器算法在训练样本为线性可分的情况下必定收敛。 (✓)  
 X (2) 势函数法仅仅适用于线性可分的情况。 (X)  
 X (3) 一般来说, 模式空间的维数都是无限的。 (有限 (只是维数非常多而已))  
 ✓ (4) 最小风险判决与模式发生的后验概率有关。 (X) ✓  
 X (5) 文法  $G = (N, \Sigma, P, S)$  是一个正则文法, 其中:  
 $N = \{S, A, B\}$ ,  $\Sigma = \{a, b\}$ , 以及  
 $P: (1) S \rightarrow Aa, (2) A \rightarrow Aa, (2) A \rightarrow Bb, (3) A \rightarrow a, (4) B \rightarrow b.$   
 (X)  $a \neq b$   
 X (6) 除非特殊情况, 多元正态随机向量经线性变换后一般不再服从正态分布。 (X)  
 ✓ (7) 设  $T$  是由一个非确定的有限状态自动机接受的链集, 则必存在一个能接受  $T$  的确定的有限状态自动机。 (✓)  
 X (8) 多元正态分布的条件分布一般不是正态分布。 (X)  
 X (9) 多元正态分布的边缘分布一般不是正态分布。 (X)  
 ✓ (10) 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。 ( )

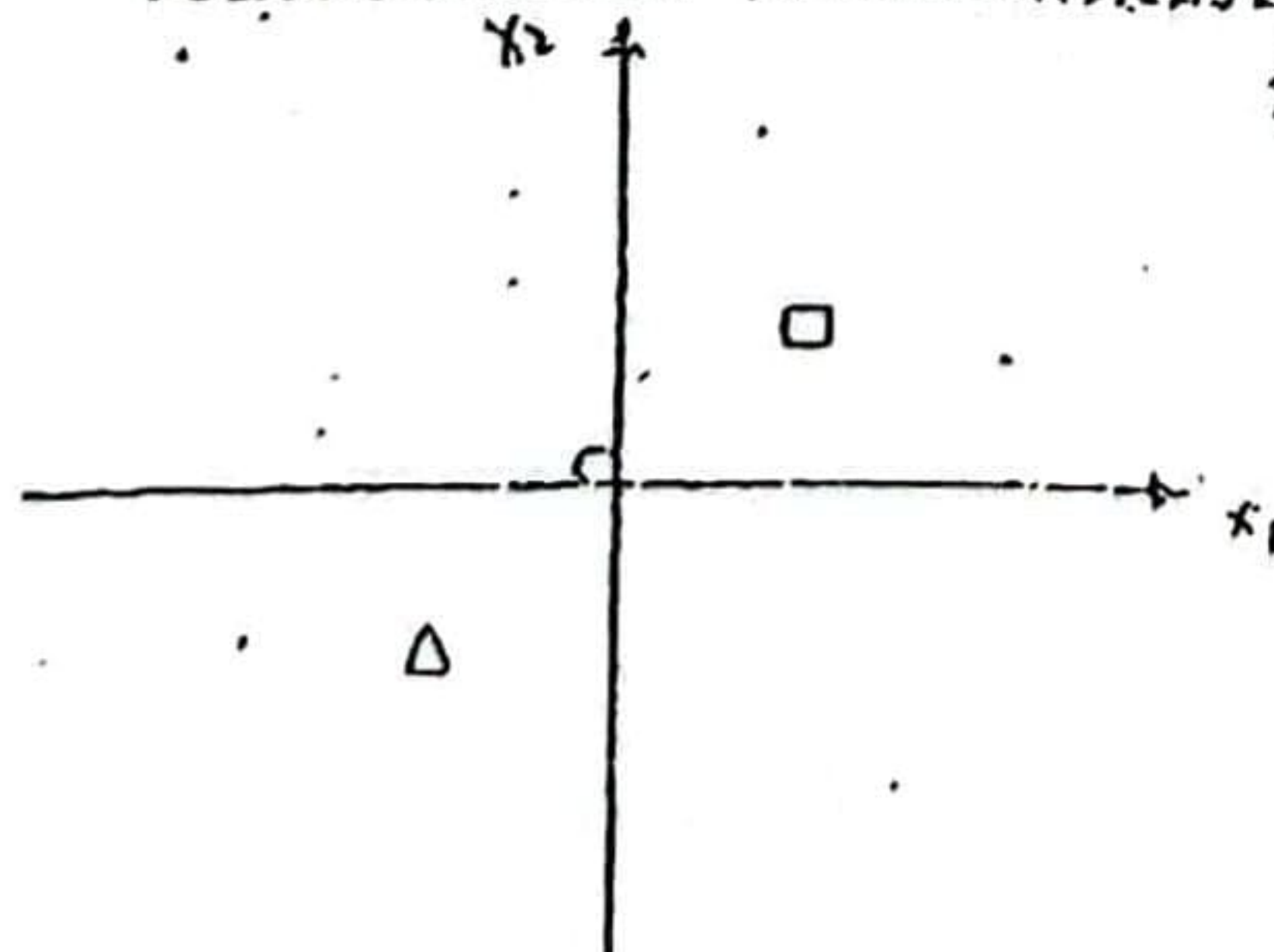
### 2. 试判断下列文法的类型:

- 正则文法: (1)  $G_1: S \rightarrow 0 | 1 | 00 | 11$   
 $G_3: S \rightarrow 0 | 1 | 0A | 1B, A \rightarrow 0, B \rightarrow 0$  正则  
 正则文法: (2)  $G_5: S \rightarrow A | B | BB, A \rightarrow 0, B \rightarrow 0$  上下文无关  
 上下文无关文法: (3)  $G_4: S \rightarrow aBC | aSBC$   
 $aB \rightarrow ab, bB \rightarrow bb, bC \rightarrow bc, cC \rightarrow cc$  上下文无关文法  
 正则文法: (4)  $G_7: S \rightarrow \varepsilon | 0S$  (10 分)

### 3. 考虑下述三类问题, 如图所示, 已知每类有一个训练样本: {样本 1, 2, 3}

$\omega_1$  类:  $(-1, -1)^T$ ,  $\omega_2$  类:  $(0, 0)^T$ ,  $\omega_3$  类:  $(1, 1)^T$

试用感知器算法求解对应的分类器 (设初始解权向量  $0$ , 步长为  $1$ , 要求给出相应解权向量的表达式, 并给出结果的图示)。 (20 分)



解: (1) 训练样本向量增广化, 将样本向量增广训练模式

$$x_1 = (-1, -1)^T, x_2 = (0, 0, 1)^T, x_3 = (1, 1, 1)^T$$

(2) 采用感知器训练算法, 设迭代步数  $k=1$ , 增量  $\rho=1$ , 设初值

$$w_1(1) = (0, 0, 0)^T, w_2(1) = (0, 0, 0)^T, w_3(1) = (0, 0, 0)^T$$

$k=1, x_1 \in \omega_1, d_1(x_1) = d_2(x_1)$  和  $d_1(x_1) = d_3(x_1)$ , 所以

$$w_1(2) = w_1(1) + x_1 = (-1, -1, 1)^T$$

$$w_2(2) = w_2(1) - x_1 = (1, 1, -1)^T$$

$$w_3(2) = w_3(1) - x_1 = (1, 1, -1)^T$$

$k=2, x_2 \in \omega_2, d_1(x_2) = w_1(2)^T x_2 = 1, d_2(x_2) = w_2(2)^T x_2 = -1$

4. 设两类样本的类内离散矩阵分别为  $d_1(x_1) = w_1(2)^T x_1 = -1, d_2(x_2) < d_1(x_2), d_2(x_2) = d_3(x_2)$

$$w_1(3) = w_1(2) - x_2 = (-1, -1, 0)^T$$

$$w_2(3) = w_2(2) + x_2 = (1, 1, 0)^T$$

$$\text{第 3 页 } w_3(3) = w_3(2) - x_2 = (1, 1, -2)^T$$



4. 已知两类的均值  $\mu_1 = (2, 0)^T$ ,  $\mu_2 = (2, 2)^T$  (0 2)

$$\therefore S_w^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$w^* = S_w^{-1}(\mu_1 - \mu_2) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$\therefore$  决策面方程为  $-x_2 = 0$ , 也即  $x_2 = 0$ .

解:  $S_1 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ,  $S_2 = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$  决策面为  $d(x) = -x_2$

均值  $\mu_1 = (2, 0)^T$ ,  $\mu_2 = (2, 2)^T$ , 试用 Fisher 准则求其决策面方程。(10分)

5. 根据下列所给文法, 构造相应的 FA, 并画出状态转移图:

G1:  $S \rightarrow a|aA$

$A \rightarrow a|aA|cA|bB$

$B \rightarrow a|b|c|aB|bB|cB$ . (10分)

6. 假设已经获得两类二维的模式样本:

$w_1: \{(0,0)^T, (2,0)^T, (2,2)^T, (0,2)^T\}$   $\mu_1 = (1, 1)^T$   $\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$w_2: \{(4,4)^T, (6,4)^T, (6,6)^T, (4,6)^T\}$   $\mu_2 = (5, 5)^T$   $\Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

两类均服从正态分布, 且先验概率相等。(1) 求两类之间的决策面方程; (2) 绘出决策面。(20分)

7. 试画出句法模式识别的系统框图并简述每部分功能。(10分)

解: 构造如下:

设  $T = \{a, b, c\}$ ,  $\Sigma = \{a, b, c\}$

由题可知所给文法是正则文法

设  $G = (N, \Sigma, P, S)$

则必存在一个 FA  $A = (Z, \Sigma, q_0, F)$

满足  $T(A) = T(G)$

构造如下:  $\Sigma = \{a, b, c\}$

$q_0 = S$

$F = \{T\}$ ,  $T$  为附加终止态

$N = N \cup \{T\} = \{S, A, B, T\}$

$N = \{S, A, B\}$ ,  $\Sigma = \{a, b, c\}$

系统框图:

由于  $S \rightarrow aA$  在  $P$  中得  $\delta(S, a) = A$

$S \rightarrow a$   $\dots$   $\delta(S, a) = \{T\}$

$A \rightarrow a$   $\dots$   $\delta(A, a) = \{T\}$

$A \rightarrow aA$   $\dots$   $\delta(A, a) = A$

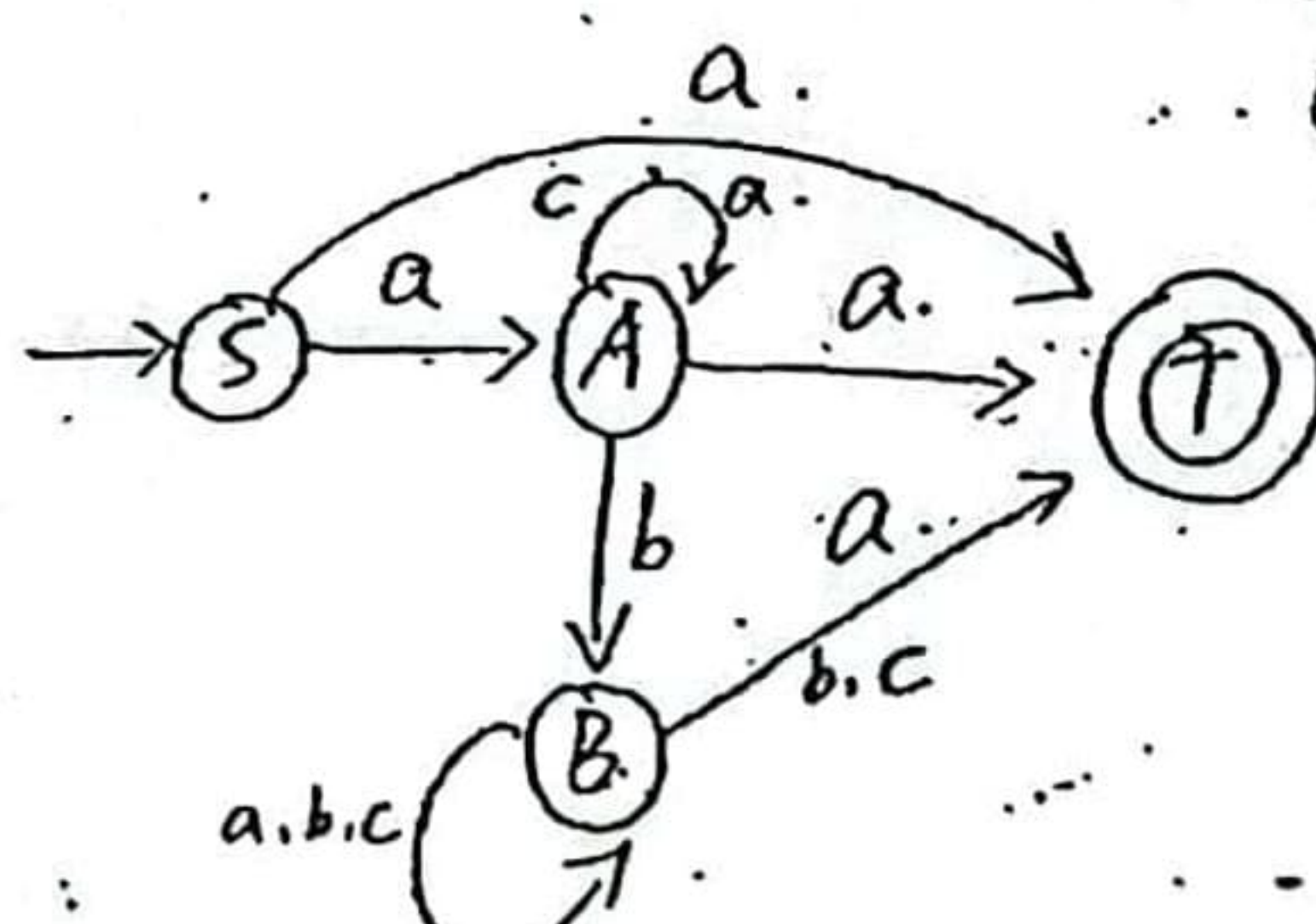
$A \rightarrow cA$   $\dots$   $\delta(A, c) = A$

$A \rightarrow bB$   $\dots$   $\delta(A, b) = B$

$B \rightarrow a, B \rightarrow aB$   $\dots$   $\delta(B, a) = \{B, T\}$

$B \rightarrow b, B \rightarrow bB$   $\dots$   $\delta(B, b) = \{B, T\}$

$B \rightarrow c, B \rightarrow cB$   $\dots$   $\delta(B, c) = \{B, T\}$



$k=5$   $x_1 = x_2 \in w_1$   $d_1(x_1) = w_1(5)x_1 = 0$

$d_2(x_2) = w_2(5)x_2 = -1$   $d_3(x_2) = w_3(5)x_2 = -1$

$w_1(6) = w_1(5) - x_2 = 1$

$w_2(6) = w_2(5) + x_2 = 0$

$w_3(6) = w_3(5) - x_2 = 0$



2004 年度秋学期研究生期末考试试题

课程名称: 模式识别

姓名: 刘原 学号: S404137019 得分: \_\_\_\_\_

闭卷部分:

是非判断题 (将答案写在圆括弧内, 每小题 1.0 分, 共 30 分)

- (1) 设  $X$  是  $d$  维正态随机向量,  $\alpha$  是与  $X$  同维的常向量, 则线性组合  $\alpha^T X$  是一个  $d$  维的正态随机向量. ☒ (X) 多元 分布的随机变量  $X$  的任何线性组合都是正态的
- (2) 多元正态随机向量经线性变换后一般不再服从正态分布. ☒ (X)
- (3) 在训练样本线性可分的情况下, 感知器算法必定收敛. ☒ (X) 当  $A=A^T$  时
- (4) 设  $X$  为  $d$  维向量,  $A$  为  $d \times d$  的非奇异的实矩阵, 则  $\frac{\partial}{\partial X} (X^T A X) = 2AX$  成立. ☒ (X)  $= (A + A^T)X$
- (5) 势函数法被用于求解线性可分的两类问题时, 和感知器算法一样, 给出能对已知训练样本进行正确分类的区分超平面. ☒ (X)
- (6) 两点间的欧氏距离是平移和旋转变换下的不变量. ☒ (X) 对
- (7) 最小错误概率判决准则是最小风险判决准则在 0-1 损失定义下的一个特例. ☒ (X)
- (8) 最小风险判决与模式发生的后验概率无关. ☒ (X)
- (9) 文法  $G = (N, \Sigma, P, S)$  是一个正则文法, 其中:  
 $N = \{S, A, B\}$ ,  $\Sigma = \{a, b\}$ , 以及  
 $P: (1) S \rightarrow aA, (2) S \rightarrow Bb, (3) A \rightarrow Ab, (4) B \rightarrow Ab$ . ☒ (X)
- (10) 当且仅当一个语言是由有限状态文法生成时, 它能被一个有限状态自动机所识别. ☒ (X)
- (11) 设  $T$  是由一个非确定的有限状态自动机接受的链集, 则必存在一个能接受  $T$  的确定的有限状态自动机. ☒ (X)
- (12) 一个由有限状态文法所生成的语言, 仅能被一个有限状态自动机所识别. ☒ (X)
- (13) 所谓留一法是从  $N$  个可用样本中选取  $(N-1)$  个样本用来设计分类器, 然后, 用剩余的一个样本对所设计的分类器进行检验, 以确定该分类器是否可被接受的一种方法. ☒ (X)
- (14) 多元正态分布等密度点的轨迹是一个包括超球面、超椭圆面以及超双曲面等在内的超二次曲面. ☒ (X)
- (15) 多元正态分布的条件分布不一定是正态分布. ☒ (X)
- (16) 多元正态分布的边缘分布仍然是正态分布. ☒ (X)
- (17) 一个上下文无关语言是非正则的, 当且仅当该文法是自嵌入的. ☒ (X)



... 设  $G = (N, \Sigma, P, S)$ , 其中:

$N = \{S, A, B\}$ ,  $\Sigma = \{a, b\}$ , 以及  $U(1) = a^2 A$

$P: (1) S \rightarrow aA, (2) A \rightarrow aA, (3) A \rightarrow \bar{a}B, (4) B \rightarrow b$

则  $L(G) = \{a^n b | n = 2, 4, \dots\}$  (X)

(19) 设  $X$  为  $d$  维正态随机向量, 对应的协方差矩阵为实对称阵, 则  $X$  的各分量是相互独立的正态随机变量. (X)

(20) 设  $T$  是由一个非确定的下推自动机接受的语汇, 则已证明必存在一个能接受  $T$  的确定的下推自动机. (X)

(21) 三角形集合上的相似关系是等价关系. (X)

(22) 下述文法  $G = (N, \Sigma, P, S)$  中的产生式是 Chomsky 范式, 其中:

$N = \{S, A, B, C\}$ ,  $\Sigma = \{a, b\}$ , 以及

$P: (1) S \rightarrow B, (2) S \rightarrow BC, (3) A \rightarrow \varepsilon, (4) B \rightarrow a, (5) C \rightarrow SA$

(23) 在一个一维特征空间中, 用于区分两个类别的分界面退化为一个点. (X)

(24) 设  $a, b$  属于字母表, 则根据定义, 串  $abba$  的长度  $|abba| = 2$ . (X)

(25) 有限状态文法是上下文有关文法的一个子集. (X)

(26) 在训练样本线性可分的情况下, 决策分界面的法线方向和权向量的方向一致. (X)

(27) 在多类问题中, 若能用一个超平面将任意一类样本同其它所有类别的样本分开, 则称该组样本是总体线性可分的; 又若能用一个超平面将所有类别中任意两个类别的样本分开, 则称该组样本是成对线性可分的. 总体线性可分的样本组一定是成对线性可分的; 反之, 成对线性可分的样本组不一定是总体线性可分的. (X)

(28) 设  $A$  为实对称矩阵,  $U$  为正交矩阵, 则  $U^T A U$  仍为实对称矩阵. (X)

(29) 设  $D$  是余码算子, 即对于码集合  $A$  和  $A$  中的符号  $a$ , 有  $D_a A = \{x | ax \in A\}$  成立. 据此, 若  $a_1, a_2$  也是  $A$  中的符号, 则应有  $D_{a_1 a_2} A = D_{a_1}(D_{a_2} A)$  成立. (X)

(30) 若句子  $w$  由串  $x, y$  和  $z$  链接而成, 即  $w = xyz$ , 则称  $x$  为  $w$  的前缀,  $y$  为  $w$  的子串,  $z$  为  $w$  的后缀. 根据此定义, 若  $a, b$  属于字母表, 则句子  $aab$  的所有前缀是  $a, aa, aab$ . (X)

$$\begin{aligned} (U^T A U)^T &= (U^T A U)^T U \\ &= U^T A^T U \\ &= U^T A U \end{aligned}$$

$$U^T A U$$

$$U^T U = E$$

$$\begin{aligned} (U^T A U)^T &= U^T A^T U \\ &= U^T A U = U^T A U \end{aligned}$$

同法

$$D_{a_1 a_2} A = \{x | a_1 a_2 x \in A\}$$

$$D_{a_2} A = \{x | a_2 x \in A\}$$

$$D_{a_1} \{x | a_2 x \in A\}$$



1.  $\mu_1, \mu_2$  是分布分别为  $N(\mu_1, \Sigma), N(\mu_2, \Sigma)$  的两个类, 依用

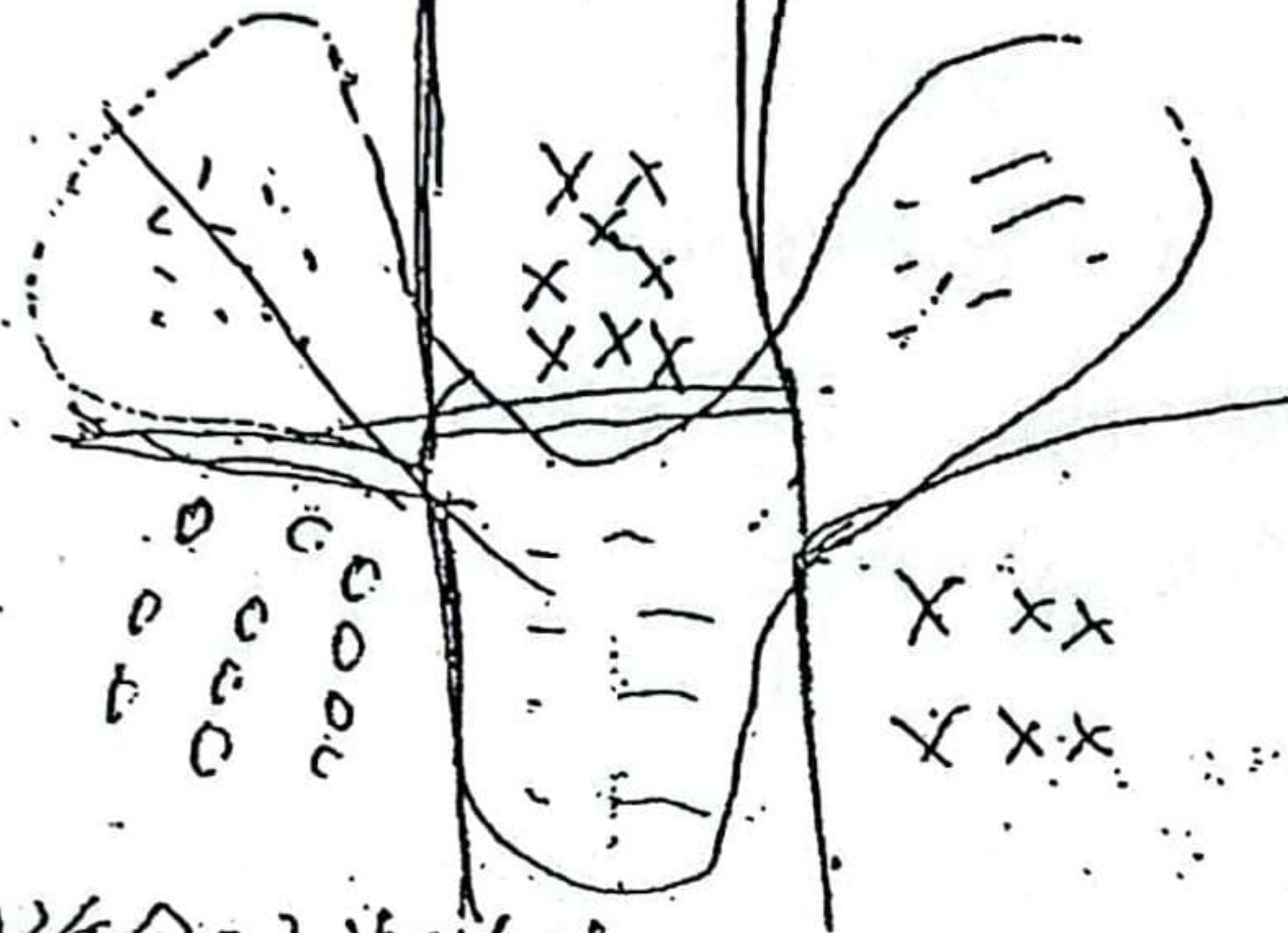
贝叶斯分类画出其分界面

$\mu_1$

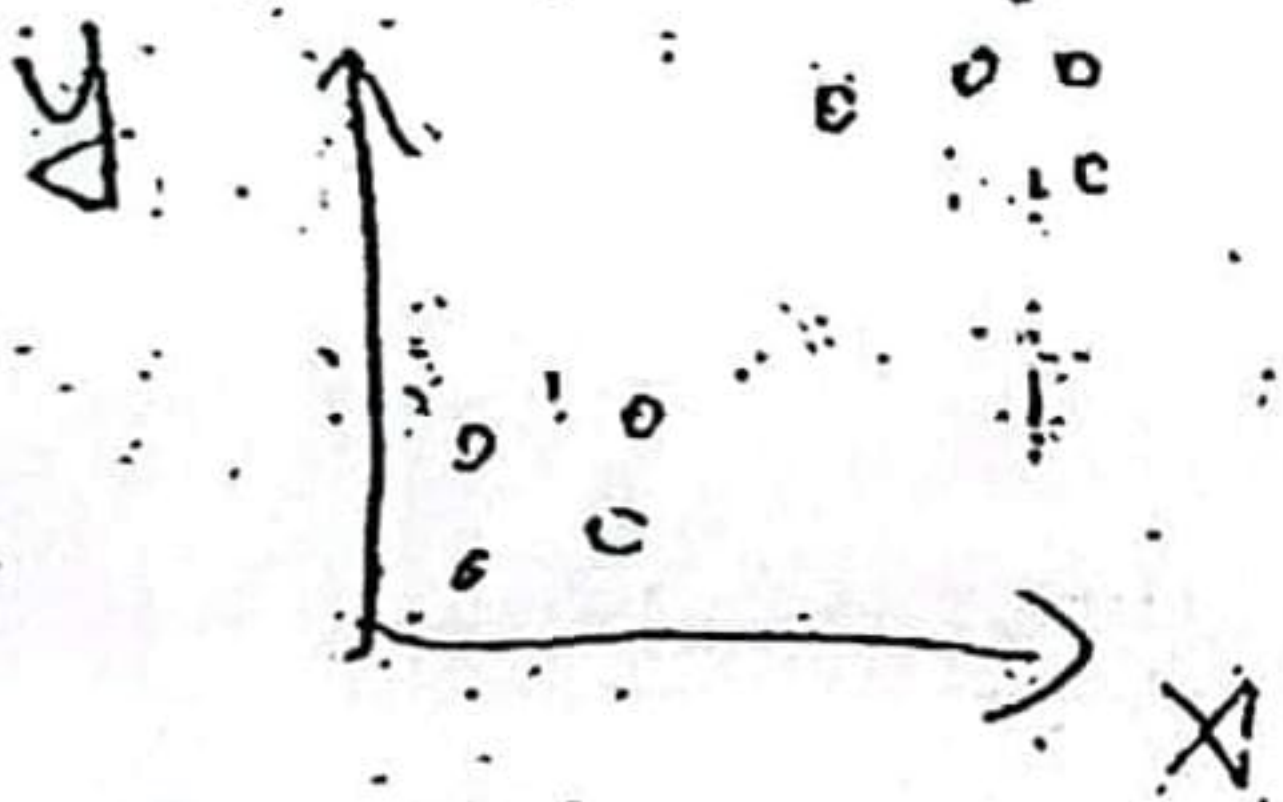
$$\begin{aligned} p(\omega_1|x) &= ? \\ p(\omega_2|x) &= ? \end{aligned} \quad \begin{aligned} p(\mu_1, \mu_2) &= ? \\ p(\mu_1, \mu_2) &= ? \end{aligned}$$

$\mu_1$

2. 试用三类分线或线性分类器画出分界面, 说明分类步骤



三. 用近邻法分类



四. 确定基元, 写出下面的文法表达式



五.  $X, Y$  满足  $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)$  的分布 求  $X+Y$  的分布



一、 填空与选择填空 (本题答案写在此试卷上, 30 分)

1、模式识别系统的基本构成单元包括: 模式采集、特征提取与选择和 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、聚类分析算法属于 (1) ; 判别域代数界面方程法属于 (3) 。

(1) 无监督分类 (2) 有监督分类 (3) 统计模式识别方法 (4) 句法模式识别方法

4、若描述模式的特征量为 0-1 二值特征量, 则一般采用 (4) 进行相似性度量。

(1) 距离测度 (2) 模糊测度 (3) 相似测度 (4) 匹配测度

5、下列函数可以作为聚类分析中的准则函数的有 (1) (3) (4) 。

$$(1) J = \text{Tr}[S_W^{-1} S_B] \quad (2) J = |S_W S_B^{-1}| \quad (3) J = \sum_{j=1}^c \sum_{i=1}^{n_j} \|\bar{x}_i^{(j)} - \bar{m}_j\|^2$$
$$(4) J = \sum_{j=1}^c (\bar{m}_j - \bar{m})' (\bar{m}_j - \bar{m})$$

6、Fisher 线性判别函数的求解过程是将 N 维特征矢量投影在 (2) 中进行。

(1) 二维空间 (2) 一维空间 (3) N-1 维空间

7、下列判别域界面方程法中只适用于线性可分情况的算法有 (1) ; 线性可分 不可分都适用的有 (3) 。

(1) 感知器算法 (2) H-K 算法 (3) 积累位势函数法

8、下列四元组中满足文法定义的有 (1) (2) (4) 。

(1)  $(\{A, B\}, \{0, 1\}, \{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0, B \rightarrow BA, B \rightarrow 0\}, A)$

(2)  $(\{A\}, \{0, 1\}, \{A \rightarrow 0, A \rightarrow 0A\}, A)$

(3)  $(\{S\}, \{a, b\}, \{S \rightarrow 00S, S \rightarrow 11S, S \rightarrow 00, S \rightarrow 11\}, S)$



(4) ( $\{A\}$ ,  $\{0, 1\}$ ,  $\{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0\}$ ,  $A$ )

9、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))。

10、欧式距离具有 (1、2)；马式距离具有 (1、2、3、4)。

(1) 平移不变性 (2) 旋转不变性 (3) 尺度缩放不变性 (4) 不受量纲影响的特性

11、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。)。

12、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。

13、积累势函数法较之于 H-K 算法的优点是 (该方法可用于非线性可分情况(也可用于线性可分情况))；位势函数  $K(x, x_k)$  与积累位势函数  $K(x)$  的关系为

$$K(\bar{x}) = \sum_{\bar{x}_k \in \bar{X}} \alpha_k K(\bar{x}, \bar{x}_k)$$

14、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；最小最大判决准则主要用于 (先验概率未知的) 情况。

15、“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 ( $C_n^m \gg n$ ) 的条件下，可以使用分支定界法以减少计算量。



16、散度  $J_{ij}$  越大, 说明  $\omega_i$  类模式与  $\omega_j$  类模式的分布 (差别越大); 当  $\omega_i$  类模式与  $\omega_j$  类模式的分布相同时,  $J_{ij} = (0)$ 。

17、已知有限状态自动机  $A_f = (\Sigma, Q, \delta, q_0, F)$ ,  $\Sigma = \{0, 1\}$ ;  $Q = \{q_0, q_1\}$ ;  $\delta$ :  $\delta(q_0, 0) = q_1, \delta(q_0, 1) = q_1, \delta(q_1, 0) = q_0, \delta(q_1, 1) = q_0$ ;  $q_0 = q_0$ ;  $F = \{q_0\}$ 。现有输入字符串: (a) 00011101011, (b) 1100110011, (c) 101100111000, (d) 0010011, 试问, 用  $A_f$  对上述字符串进行分类的结果为 ( $\omega_1: \{a, d\}; \omega_2: \{b, c\}$ )。

18、影响聚类算法结果的主要因素有 (②③④)。

①已知类别的样本质量; ②分类准则; ③特征选取; ④模式相似性测度。

19、模式识别中, 马式距离较之于欧式距离的优点是 (③④)。

①平移不变性; ②旋转不变性; ③尺度不变性; ④考虑了模式的分布。

20、基于二次准则函数的 H-K 算法较之于感知器算法的优点是 (①③)。

①可以判别问题是否线性可分; ②其解完全适用于非线性可分的情况;

③其解的适应性更好; ④计算量小。

21、影响基本 C 均值算法的主要因素有 (④①②)。

①样本输入顺序; ②模式相似性测度; ③聚类准则; ④初始类心的选取。

22、位势函数法的积累势函数  $K(x)$  的作用相当于 Bayes 判决中的 (②④)。

①先验概率; ②后验概率; ③类概率密度; ④类概率密度与先验概率的乘积。

23、在统计模式分类问题中, 当先验概率未知时, 可以使用 (②④)。

①最小损失准则; ②最小最大损失准则; ③最小误判概率准则; ④N-P 判决。

24、在 (①③) 情况下, 用分支定界法做特征选择计算量相对较少。



- ①  $C_n^d \gg n$ , ( $n$  为原特征个数,  $d$  为要选出的特征个数); ② 样本较多; ③ 选用的可分性判据  $J$  对特征数目单调不减; ④ 选用的可分性判据  $J$  具有可加性。
- 25、散度  $J_0$  是根据 (\_\_\_③\_\_\_) 构造的可分性判据。
- ① 先验概率; ② 后验概率; ③ 类概率密度; ④ 信息熵; ⑤ 几何距离。
- 26、似然函数的模型已知且为单峰, 则可用 (\_\_\_①②③④⑤\_\_\_) 估计该似然函数。
- ① 矩估计; ② 最大似然估计; ③ Bayes 估计; ④ Bayes 学习; ⑤ Parzen 窗法。
- 27、 $Kn$  近邻元法较之 Parzen 窗法的优点是 (\_\_\_②\_\_\_)。
- ① 所需样本数较少; ② 稳定性较好; ③ 分辨率较高; ④ 连续性较好。
- 28、从分类的角度讲, 用 DKLT 做特征提取主要利用了 DKLT 的性质: (\_\_\_①③\_\_\_)。
- ① 变换产生的新分量正交或不相关; ② 以部分新的分量表示原矢量均方误差最小; ③ 使变换后的矢量能量更趋集中;
- 29、一般, 剪辑  $k$ -NN 最近邻方法在 (\_\_\_①\_\_\_) 的情况下效果较好。
- ① 样本数较大; ② 样本数较小; ③ 样本呈团状分布; ④ 样本呈链状分布。
- 30、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 (\_\_\_②③\_\_\_)。
- ① 已知类别样本质量; ② 分类准则; ③ 特征选取; ④ 量纲。

## 二、(15 分) 简答及证明题

(1) 影响聚类结果的主要因素有那些?

(2) 证明马氏距离是平移不变的、非奇异线性变换不变的。

答: (1) 分类准则, 模式相似性测度, 特征量的选择, 量纲。

(2) 证明:

$$d^2(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)' V^{-1} (\vec{x}_i - \vec{x}_j) \quad (2 \text{ 分})$$



$$V = \frac{1}{m-1} \sum_{i=1}^m (\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})' \quad (2 \text{ 分})$$

$$\bar{\vec{x}} = \frac{1}{m} \sum_{i=1}^m \vec{x}_i \quad (1 \text{ 分})$$

设, 有非奇异线性变换: (1 分)

$$\vec{y} = \frac{1}{m} \sum_{i=1}^m \vec{y}_i = \frac{1}{m} \sum_{i=1}^m A\vec{x}_i = A \frac{1}{m} \sum_{i=1}^m \vec{x}_i = A\bar{\vec{x}}$$

$$\begin{aligned} V_y &= \frac{1}{m-1} \sum_{i=1}^m (\vec{y}_i - \vec{y})(\vec{y}_i - \vec{y})' \\ &= \frac{1}{m-1} \sum_{i=1}^m (A\vec{x}_i - A\bar{\vec{x}})(A\vec{x}_i - A\bar{\vec{x}})' \\ &= \frac{1}{m-1} \sum_{i=1}^m A(\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})'A' \\ &= A \left[ \frac{1}{m-1} \sum_{i=1}^m (\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})' \right] A' = AV_x A' \end{aligned}$$

$$\begin{aligned} d_y^2(\vec{y}_i, \vec{y}_j) &= (\vec{y}_i - \vec{y}_j)' V_y^{-1} (\vec{y}_i - \vec{y}_j) \\ &= (A\vec{x}_i - A\vec{x}_j)' V_y^{-1} (A\vec{x}_i - A\vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' V_y^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' (AV_x A')^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' A^{-1} V_x^{-1} A^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' V_x^{-1} (\vec{x}_i - \vec{x}_j) \\ &= d_x^2(\vec{x}_i, \vec{x}_j) \end{aligned} \quad (4 \text{ 分})$$

三、(8 分) 说明线性判别函数的正负和数值大小在分类中的意义并证明之。

答: (1) (4 分)  $d(\vec{x})$  的绝对值  $|d(\vec{x})|$  正比于  $\vec{x}$  到超平面  $d(\vec{x})=0$  的距离  $d$ .

平面  $\pi$  的方程可以写成  $\frac{\vec{w}_0'}{\|\vec{w}_0\|} \vec{x} = \frac{-w_{n+1}}{\|\vec{w}_0\|}$

式中  $\|\vec{w}_0\| = (w_1^2 + w_2^2 + \dots + w_n^2)^{1/2}$ 。于是  $\vec{n} = \frac{\vec{w}_0'}{\|\vec{w}_0\|}$  是平面  $\pi$  的单位法矢量, 上式可写成

$$\vec{n}' \vec{x} = \frac{-w_{n+1}}{\|\vec{w}_0\|}$$

设  $\vec{p}$  是平面  $\pi$  中的任一点,  $\vec{x}$  是特征空间  $X^n$  中任一点, 点  $\vec{x}$  到平面  $\pi$  的距离为



差矢量  $(\bar{x} - \bar{p})$  在  $\bar{n}$  上的投影的绝对值, 即

$$\begin{aligned} d_x &= |\bar{n}(\bar{x} - \bar{p})| = |\bar{n}\bar{x} - \bar{n}\bar{p}| = \left| \frac{\bar{w}_0'}{\|\bar{w}_0\|} \bar{x} - \frac{\bar{w}_0'}{\|\bar{w}_0\|} \bar{p} \right| \\ &= \left| \frac{\bar{w}_0'}{\|\bar{w}_0\|} \bar{x} + \frac{w_{n+1}}{\|\bar{w}_0\|} \right| \\ &= \frac{|\bar{w}_0' \bar{x} + w_{n+1}|}{\|\bar{w}_0\|} = \frac{1}{\|\bar{w}_0\|} |d(\bar{x})| \end{aligned} \quad (1-1)$$

上式中利用了  $\bar{p}$  在平面  $\pi$  中, 故满足方程

$$\frac{\bar{w}_0'}{\|\bar{w}_0\|} \bar{p} = \frac{-w_{n+1}}{\|\bar{w}_0\|}$$

式(1-1)的分子为判别函数绝对值, 上式表明,  $d(\bar{x})$  的值  $|d(\bar{x})|$  正比于  $\bar{x}$  到超平面  $d(\bar{x})=0$  的距离  $d_x$ , 一个特征矢量代入判别函数后所得值的绝对值越大表明该特征点距判别界面越远。

(2) (4分)  $d(\bar{x})$  的正(负)反映  $\bar{x}$  在超平面  $d(\bar{x})=0$  的正(负)侧

两矢量  $\bar{n}$  和  $(\bar{x} - \bar{p})$  的数积为

$$\bar{n}'(\bar{x} - \bar{p}) = \|\bar{n}\| \|\bar{x} - \bar{p}\| \cos(\bar{n}, (\bar{x} - \bar{p})) \quad (2 \text{ 分})$$

显然, 当  $\bar{n}$  和  $(\bar{x} - \bar{p})$  夹角小于  $90^\circ$  时, 即  $\bar{x}$  在  $\bar{n}$  指向的那个半空间中,  $\cos(\bar{n}, (\bar{x} - \bar{p})) > 0$ ;

反之, 当  $\bar{n}$  和  $(\bar{x} - \bar{p})$  夹角大于  $90^\circ$  时, 即  $\bar{x}$  在  $\bar{n}$  背向的那个半空间中,  $\cos(\bar{n}, (\bar{x} - \bar{p})) < 0$ 。

由于  $\|\bar{w}_0\| > 0$ , 故  $\bar{n}'(\bar{x} - \bar{p})$  和  $\bar{w}_0' \bar{x} + w_{n+1}$  同号。所以, 当  $\bar{x}$  在  $\bar{n}$  指向的半空间中时,

$\bar{w}_0' \bar{x} + w_{n+1} > 0$ ; 当  $\bar{x}$  在  $\bar{n}$  背向的半空间中,  $\bar{w}_0' \bar{x} + w_{n+1} < 0$ 。判别函数值的正负表示出特征点位于哪个半空间中, 或者换句话说, 表示特征点位于界面的哪一侧。

五、(12分, 每问4分) 在目标识别中, 假定有农田和装甲车两种类型, 类型  $\omega_1$  和类型  $\omega_2$  分别代表农田和装甲车, 它们的先验概率分别为 0.8 和 0.2, 损失函数如表 1 所示。现在做了三次试验, 获得三个样本的类概率密度如下:



$$p(x/\omega_1): 0.3, 0.1, 0.6$$

$$p(x/\omega_2): 0.7, 0.8, 0.3$$

- (1) 试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型;
- (2) 假定只考虑前两种判决, 试用贝叶斯最小风险准则判决三个样本各属于哪一类;
- (3) 把拒绝判决考虑在内, 重新考核三次试验的结果。

表 1

型 判决 损失	$\omega_1$	$\omega_2$
$\alpha_1$	1	4
$\alpha_2$	5	1
$\alpha_3$	1	1

解: 由题可知:  $P(\omega_1)=0.7, P(\omega_2)=0.3$ ,  $\frac{P(\omega_2)}{P(\omega_1)}=\frac{3}{7}$ ,  $\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)}=\frac{3}{7}$ ,

$$\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)}=\frac{1}{8}, \quad \frac{P(x_3|\omega_1)}{P(x_3|\omega_2)}=2$$

(1) (4 分) 根据贝叶斯最小误判概率准则知:

$\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} = \frac{P(\omega_2)}{P(\omega_1)}$ , 则可以任判;  $\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{P(\omega_2)}{P(\omega_1)}$ , 则判为  $\omega_2$ ;  $\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$ , 则

判为  $\omega_1$ ;



(2) (4分) 由题可知:  $\frac{P(\omega_2)(\lambda_{21}-\lambda_{22})}{P(\omega_1)(\lambda_{12}-\lambda_{11})} = \frac{0.3(5-1)}{0.7(4-1)} = \frac{4}{7}$

则  $\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} < \frac{4}{7}$ , 判为  $\omega_2$ ;

$\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{4}{7}$ , 判为  $\omega_2$ ;

$\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{4}{7}$ , 判为  $\omega_1$ ;

(3) (4分) 对于两类问题, 对于样本  $x$ , 假设  $P(x)$  已知, 有

$$\begin{aligned} R(\alpha_j|x) &= \lambda(\alpha_j|\omega_1)P(\omega_1|x) + \lambda(\alpha_j|\omega_2)P(\omega_2|x) = \\ &= \frac{\lambda(\alpha_j|\omega_1)P(x|\omega_1)P(\omega_1) + \lambda(\alpha_j|\omega_2)P(x|\omega_2)P(\omega_2)}{P(x)} \end{aligned}$$

则对于第一个样本,

$$R(\alpha_1|x) = \frac{5 \times 0.21}{P(x)}, R(\alpha_2|x) = \frac{4 \times 0.21}{P(x)}, R(\alpha_3|x) = \frac{2 \times 0.21}{P(x)}, \text{ 则拒判};$$

$$R(\alpha_1|x) = \frac{1.03}{P(x)}, R(\alpha_2|x) = \frac{0.59}{P(x)}, R(\alpha_3|x) = \frac{0.24}{P(x)}, \text{ 则拒判};$$

$$R(\alpha_1|x) = \frac{0.78}{P(x)}, R(\alpha_2|x) = \frac{2.19}{P(x)}, R(\alpha_3|x) = \frac{0.51}{P(x)}, \text{ 拒判}.$$

#### 1. 监督学习与非监督学习的区别:

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。

(实例: 道路图) 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，则依据道路路面像素与非道路像素之间的聚类分析进行聚类运算，以实现道路图像的分割。

2. 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。

3. 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。

该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。

其优点是通过对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。

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,  $\Sigma^{-1}$  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球, 如果是单位矩阵  $\Sigma$ , 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, 则依据道路路面像素与非道路像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、 如果观察一个时序信号时在离散时刻序列得到的观察量序列表示为

$O = \{o_1, \dots, o_L\}$ , 而该时序信号的内在状态序列表示成  $S = \{s_1, \dots, s_L\}$ 。如果计算在给定  $O$  条件下出现  $S$  的概率, 试问此概率是何种概率。如果从观察



序列来估计状态序列的最大似然估计, 这与 Bayes 决策中基于最小错误率的决策有什么关系。

答: 在给定观察序列  $O = (o_1, \dots, o_L)$  条件下分析它由某个状态序列  $S$  产生的概率似后验概率, 写成  $P(S|O)$ , 而通过  $O$  求对状态序列的最大似然估计, 与贝叶斯决策的最小错误率决策相当。

六、 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  $\begin{pmatrix} \lambda - 1 & -1/2 \\ -1/2 & \lambda - 1 \end{pmatrix}$  得

$(\lambda - 1)^2 = \frac{1}{4}$ , 则  $\lambda = \begin{cases} \frac{3}{2} \\ \frac{1}{2} \end{cases}$ , 相应的特征向量为:  $\lambda = \frac{3}{2}$ , 对应特征向量为

$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\lambda = \frac{1}{2}$ , 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消除。

七、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;

八、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。

该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。

其优点是通过对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。

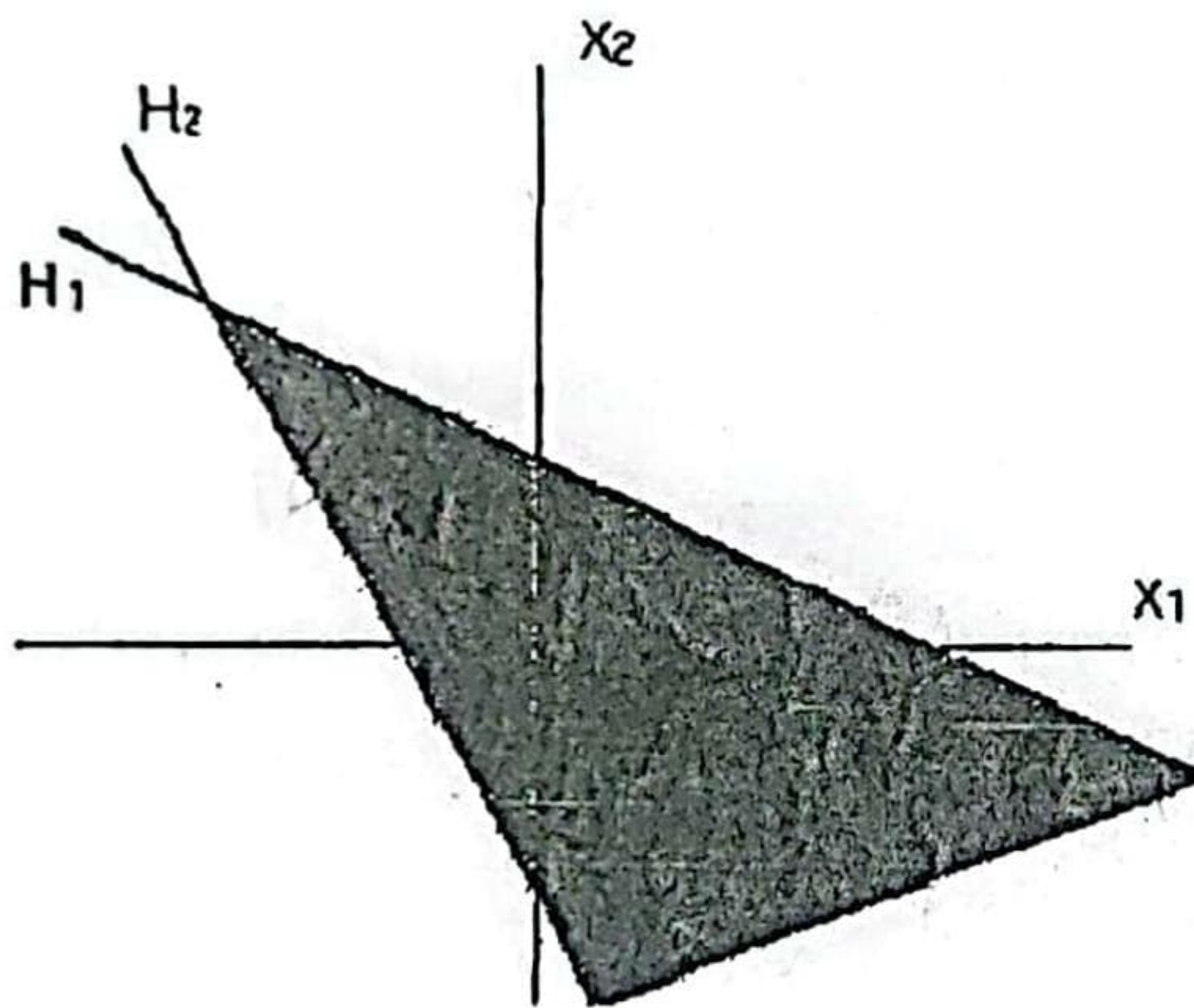
九、 在一两维特征空间,两类决策域由两条直线  $H_1$  和  $H_2$  分界,

$$\text{其中 } H_1: x_1 + 2x_2 - 2 = 0 \quad H_2: 2x_1 + x_2 + 2 = 0$$

而包含  $H_1$  与  $H_2$  的锐角部分为第一类,其余为第二类。

试求:

1. 用一双层感知器构造该分类器
2. 用凹函数的并构造该分类器

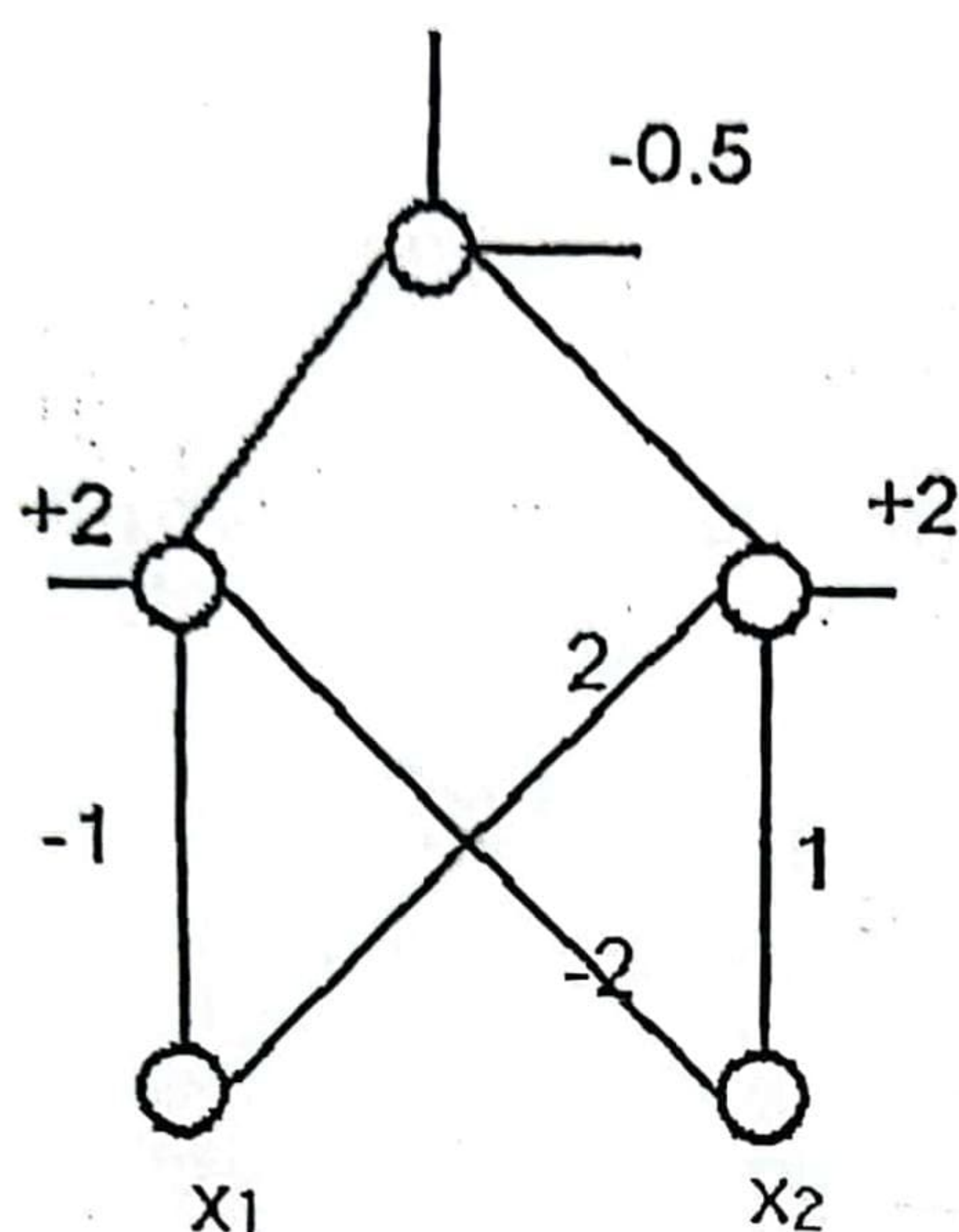


答: 按题意要求

1)  $H_1$  与  $H_2$  将空间划分成四个部分,按使  $H_1$  与  $H_2$  大于零与小于零表示成四个区域,而第一类属于(—+)区域,为方便起见,令

$H_1: -x_1 - 2x_2 + 2 = 0$  则第一类在(++)区域。用双层感知器,神经元用  $\pm 1$  域值,则在第一类样本输入时,两隐层结点的输出均为+1,其余则分别为(+—), (— —), (—+), 故可按图设置域值。





2) 用凹函数的并表示:  $H_1^{(x)} \cap H_2^{(x)}$  或表示成  $P^{(x)} = \min(H_1^{(x)}, H_2^{(x)})$ , 如  $P^{(x)} > 0$ , 则  $x \in \omega_1$ , 否则  $x \in \omega_2$

十、 设有两类正态分布的样本基于最小错误率的贝叶斯决策分界面, 分别为  $X_2=0$ , 以及  $X_1=3$ , 其中两类的协方差矩阵  $|\Sigma_1|=|\Sigma_2|$ , 先验概率相等, 并

且有  $\Sigma_1 = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$ ,  $\mu_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ 。

试求:  $\Sigma_2$  以及  $\mu_2$ 。

答: 设待求  $\Sigma_2 = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ , 待求  $\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{22} \end{pmatrix}$

由于  $|\Sigma_1|=|\Sigma_2|$ , 先验概率相等。

则基于最小错误率的 Bayes 决策规则, 在两类决策面分界面上的样本  $X$  应满足

$$(X - \mu_1)^T \Sigma_1^{-1} (X - \mu_1) = (X - \mu_2)^T \Sigma_2^{-1} (X - \mu_2) \quad (1)$$



其中按题意  $\mu_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ ,  $\Sigma_1^{-1} = \frac{4}{3} \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$  (注: 为方便起见, 在下面计算中先去掉系数  $4/3$ ).

按题意分界面由  $x_1=3$  及  $x_2=0$  两条直线构成, 则分界面方程为

$$x_2(x_1-3)=0 \Rightarrow x_2x_1-3x_2=0 \quad (2)$$

对 (1) 式进行分解有

$$\begin{aligned} X^T \Sigma_1^{-1} X - 2\mu_1^T \Sigma_1^{-1} X + \mu_1^T \Sigma_1^{-1} \mu_1 \\ = X^T \Sigma_2^{-1} X - 2\mu_2^T \Sigma_2^{-1} X + \mu_2^T \Sigma_2^{-1} \mu_2 \end{aligned}$$

$$\begin{aligned} X^T (\Sigma_1^{-1} - \Sigma_2^{-1}) X - 2(\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1}) X + \\ \mu_1^T \Sigma_1^{-1} \mu_1 - \mu_2^T \Sigma_2^{-1} \mu_2 = 0 \quad (3) \end{aligned}$$

由 (3) 式第一项得

$$\begin{aligned} X^T (\Sigma_1^{-1} - \Sigma_2^{-1}) X &= X^T \left[ \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} - \begin{pmatrix} a & b \\ b & c \end{pmatrix} \right] X \\ &= x_1(1-a) + 2x_1x_2(1/2-b) + x_2^2(1-c) \quad (4) \end{aligned}$$

将 (4) 式与 (2) 式对比可知

$$a=1, c=1$$

又由  $c=1$  与  $|\Sigma_1| \neq |\Sigma_2|$ , 得  $b \neq 1/2$ ,  $b$  有两种可能, 即  $b=1/2$  或  $b=-1/2$ ,

如果  $b=1/2$ , 则表明  $|\Sigma_1| = |\Sigma_2|$ , 此时分界面方程应为线性, 与题意不符,

只有  $b=-1/2$

$$\text{则 (4) 式为: } 2X_1X_2 \quad (5)$$

将相应结果代入 (3) 式第二项有



$$\begin{aligned}
& 2[\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1}]x \\
&= 2\left[(-1, 2)\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} - \mu_2^T \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}\right]x \\
&= 2\left[(0, \frac{3}{2}) - (\mu_{21} - \frac{1}{2}\mu_{22}, -\frac{1}{2}\mu_{21} + \mu_{22})\right]\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= 2\left[(\mu_{21} + \frac{1}{2}\mu_{22})x_1 + (\frac{1}{2}\mu_{21} - \mu_{22} + \frac{3}{2})x_2\right] \quad (6)
\end{aligned}$$

则结合 (5) (2) 应有

$$-\mu_{21} + \frac{1}{2}\mu_{22} = 0, \text{ 则 } 2\mu_{21} = \mu_{22} \quad (7)$$

$$\frac{1}{2}\mu_{21} - \mu_{22} + \frac{3}{2} = 3$$

解得  $\mu_{21} = -1, \mu_{22} = -2$ ,

$$\text{由 } \Sigma_2^{-1} = \frac{4}{3}\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix} \text{ 得 } \Sigma_2 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

九、证明在  $\Sigma$  正定或半正定时, Mahalanobis 距离  $r$  符合距离定义的三个条件, 即

$$(1) r(a, b) = r(b, a)$$

$$(2) \text{ 当且仅当 } a = b \text{ 时, 有 } r(a, b) = 0$$

$$(3) r(a, c) \leq r(a, b) + r(b, c)$$

证明:

$$(1) \text{ 根据定义 } r(a, b) = (a - b)^T \Sigma^{-1} (a - b) = (b - a)^T \Sigma^{-1} (b - a)$$



(2) 由于  $\Sigma$  为对称阵, 故  $\Sigma$  可以分解为  $\Sigma = P^T D^T D P$ , 其中所有特征值大于等于零。

$$D = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\lambda_n} \end{pmatrix}, \text{ 且}$$

可以认为

$$\begin{aligned} r(a, b) &= (a - b)^T \Sigma^{-1} (a - b) = \\ &= (a - b)^T P^T D^T D P (a - b) = (D P (a - b))^T D P (a - b) \end{aligned}$$

这就变为了传统意义上的欧氏距离, 可以由欧氏距离满足的性质直接证明本命题。

十、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。



十一、已知有两类数据,分别为

$$\omega_1: (1, 0), (2, 0), (1, 1)$$

$$\omega_2: (-1, 0), (0, 1), (-1, 1)$$

试求: 该组数据的类内及类间离散矩阵  $S_w$  及  $S_b$ 。

答: 第一类的均值向量为

$$m_1 = \left(\frac{4}{3}, \frac{1}{3}\right), \quad m_2 = \left(-\frac{2}{3}, \frac{2}{3}\right)$$

$$\therefore S_1 = \frac{1}{9} \begin{pmatrix} 6 & -5 \\ -5 & 6 \end{pmatrix}, \quad S_2 = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$S_w = S_1 + S_2 = \frac{1}{9} \begin{pmatrix} 12 & -2 \\ -2 & 12 \end{pmatrix}$$

$$S_b = \begin{pmatrix} \frac{6}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{6}{3} & -\frac{1}{3} \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 36 & -6 \\ -6 & 1 \end{pmatrix};$$

十二、设一个二维空间中的两类样本服从正态分布, 其参数分别为:

$$\mu_1 = (1, 0)', \quad \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mu_2 = (-1, 0)', \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

先验概率  $P(\omega_1) = P(\omega_2)$ ,

试证明: 其基于最小错误率的贝叶斯决策分界面方程为一圆, 并求其方程。

证明: 先验概率相等条件下, 基于最小错误率贝叶斯决策的分界面上两类条件概



率密度函数相等。

因此有：

$$-\frac{1}{2}(X-u_1)^T \Sigma_1^{-1}(X-u_1) - \frac{1}{2} \ln |\Sigma_1| =$$
$$-\frac{1}{2}(X-u_2)^T \Sigma_2^{-1}(X-u_2) - \frac{1}{2} \ln |\Sigma_2|$$

$$(x_1 - 1)^2 + x_2^2 = \frac{1}{2}(x_1 + 1)^2 + \frac{1}{2}x_2^2 + \ln 4$$

化简为  $(x_1 - 3)^2 + x_2^2 = 8 + 2 \ln 4$ ，是一个圆的方程。

十三、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。

十四、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: 1.

解: 利用贝叶斯公式, 分别计算出  $\omega_1$  及  $\omega_2$  的后验概率。

$$P(\omega_1|x) = \frac{p(x|\omega_1)P(\omega_1)}{\sum_{j=1}^2 p(x|\omega_j)P(\omega_j)} = \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.818$$

$$P(\omega_2|x) = 1 - p(\omega_1|x) = 0.182$$

根据贝叶斯决策规则式(2-2), 有

$$P(\omega_1|x) = 0.818 > P(\omega_2|x) = 0.182$$

所以合理的决策是把  $x$  归类于正常状态。

2.



解:已知条件为

$$\begin{array}{ll} P(\omega_1) = 0.9, & P(\omega_2) = 0.1 \\ p(x|\omega_1) = 0.2, & p(x|\omega_2) = 0.4 \\ \lambda_{11} = 0, & \lambda_{12} = 6 \\ \lambda_{21} = 1, & \lambda_{22} = 0 \end{array}$$

根据 1 的计算结果可知后验概率为

$$P(\omega_1|x) = 0.818, \quad P(\omega_2|x) = 0.182$$

再计算出条件风险

$$R(a_1|x) = \sum_{j=1}^2 \lambda_{1j} P(\omega_j|x) = \lambda_{12} P(\omega_2|x) = 1.092$$

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

由于  $R(a_1|x) > R(a_2|x)$

即决策为  $\omega_2$  的条件风险小于决策为  $\omega_1$  的条件风险,因此我们采取决策行动  $a_2$ ,即判断待识别的细胞  $x$  为  $\omega_2$  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”就起了主导作用。

十五、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。



十六、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。

十七、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。

两类别问题: 判别函数

$$g_1(x) = \lambda_{11}p(\omega_1|x) + \lambda_{12}p(\omega_2|x)$$

$$g_2(x) = \lambda_{21}p(\omega_1|x) + \lambda_{22}p(\omega_2|x)$$

决策面方程:  $g_1(x) = g_2(x)$

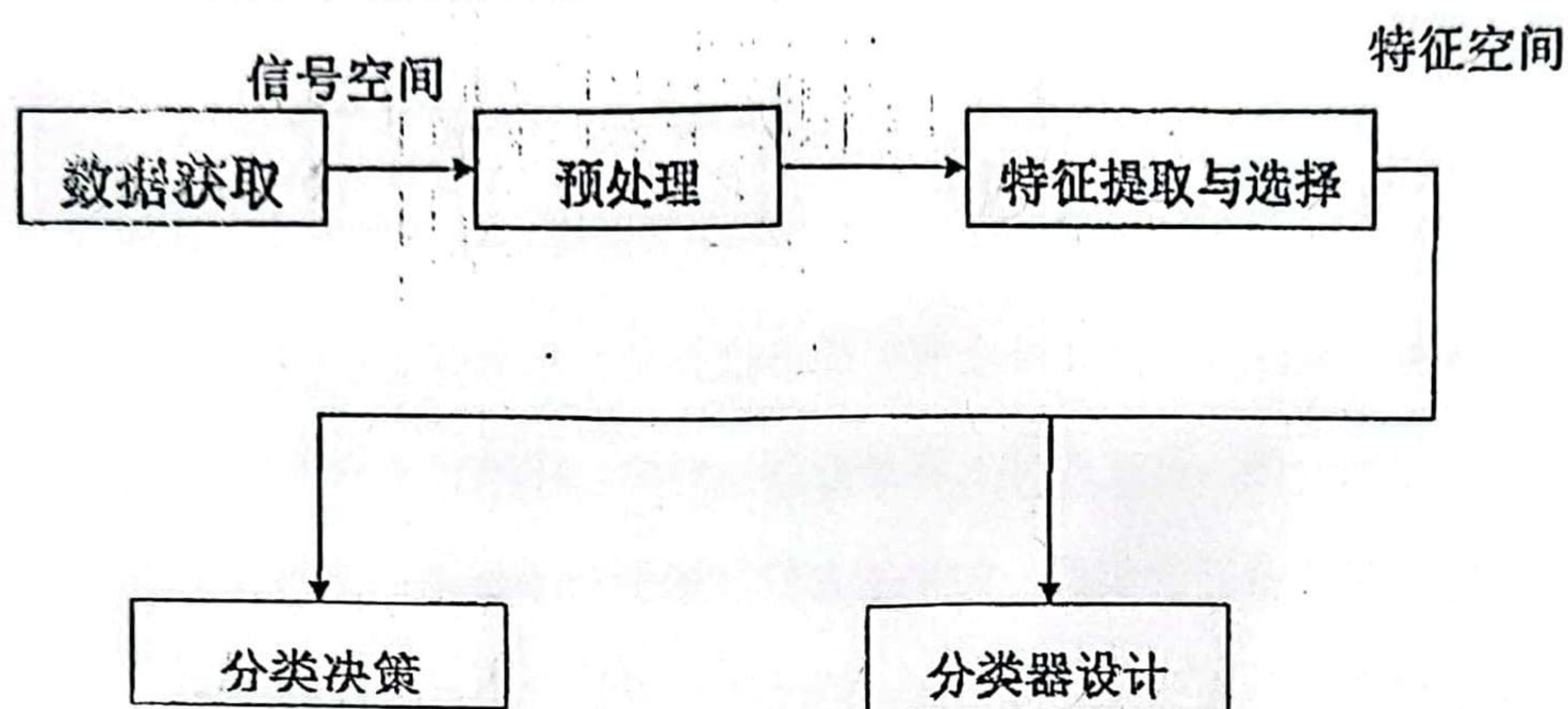
$c$  类别问题: 判别函数

$$g_i(x) = \sum_{j=1}^c \lambda_{ji}p(\omega_j|x), \quad i = 1, \dots, c$$

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



十八、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以为波形。

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。

十九、有两类样本集

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



$$x_1^1 = [0, 0, 1]^T, \quad x_2^1 = [0, 1, 0]^T, \quad x_1^2 = [0, 1, 1]^T, \quad x_2^2 = [1, 1, 1]^T$$

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。

$$(1) \quad E[xx^T] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}, \text{ 其对应的本征值和本征向量为:}$$

$$\Lambda = \begin{bmatrix} \frac{1}{4} & & \\ & \frac{1}{4} & \\ & & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$\text{降到二维空间取 } U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$\text{对应的坐标 } \omega_1: \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix} \quad \omega_2: \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{3\sqrt{2}}{\sqrt{6}} \\ 0 \end{bmatrix}$$

$$(2) \quad m_1 = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}, \quad m_2 = \begin{bmatrix} \frac{1}{4} \\ \frac{3}{4} \\ \frac{3}{4} \\ \frac{1}{4} \end{bmatrix}, \quad S_w = \frac{1}{12} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

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

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

二十、定性说明基于参数方法和非参数方法的概率密度估计有什么区别？



答：基于参数方法：是由已知类别的样本集对总体分布的某些参数进行统计推断

非参数方法：已知样本所属类别，但未知总体概率密度函数形式

假设二类模式服从如下的正态分布：

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

请分别求出使  $\text{tr}(S_w^{-1}S_b)$  和  $\text{tr}(S_t)$  ( $S_t = S_w + S_b$ ) 最大化的 一维特征空间的变换  
二十一、矢量。

答：

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

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

$$S_w = \frac{1}{2} (S_1 + S_2) = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}, S_w^{-1} = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

$$S_w^{-1}S_b = \begin{bmatrix} 5 & 5 \\ 2 & 2 \end{bmatrix}$$

因为  $S_w^{-1}S_b$  的秩为 1, 所以  $S_w^{-1}S_b$  只有一个非零本征值,  $W$  是  $D \times 1$  矩阵, 即  $W=w$ . 为求  $S_w^{-1}S_b$  的本征值应解方程：

$$S_w^{-1}S_b w = \lambda_1 w$$

或

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

$\frac{1}{4} (\mu_1 - \mu_2)^T w$  是标量, 所以

$$w = S_w^{-1} (\mu_1 - \mu_2) = \begin{bmatrix} 10 \\ 4 \end{bmatrix}$$

二十二、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。

二十三、对于两类问题, 假定  $P(x|w_i) \sim N(\mu_i, \Sigma_i)$ ,  $i=1,2$ ,  $x$  为  $d$  维特征向量

请给出以下三种情况下的贝叶斯判别函数, 并说明各有什么特点:

1.  $\Sigma_1 \neq \Sigma_2$
2.  $\Sigma_1 = \Sigma_2$
3.  $\Sigma_1 = \Sigma_2 = \sigma^2 I$ ,  $I$  为单位矩阵

答:

1.  $\Sigma_1 \neq \Sigma_2$

各类协方差不同

$$\begin{aligned} g_i(x) &= -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i) \\ &= x^T W_i x + w_i^T x + w_{i0} \end{aligned}$$

其中  $W_i = -\frac{1}{2} \Sigma_i^{-1}$



$$w_i = \Sigma_i^{-1} \mu_i$$

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

决策面

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

$$x^T (W_i - W_j)x + (w_i - w_j)^T x + w_{i0} - w_{j0} = 0$$

为超二次曲面

2.  $\Sigma_1 = \Sigma_2$

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

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

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

$$w^T (x - x_0) = 0 \quad w = \Sigma^{-1} (\mu_i - \mu_j)$$

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

判别边界仍是一条直线，但不垂直于均值的连线。

3.  $\Sigma_1 = \Sigma_2 = \sigma^2 I$



$$g_i(x) = -\frac{(x - \mu_i)'(x - \mu_i)}{2\sigma^2} + \ln P(\omega_i)$$

$$g_i(x) = \frac{(x'x - 2\mu_i'x + \mu_i'\mu_i)}{2\sigma^2} + \ln P(\omega_i)$$

$$g_i(x) = w_i'x + w_{i0}$$

$$w_i = \frac{\mu_i}{\sigma^2}; \quad w_{i0} = -\frac{1}{2\sigma^2} \mu_i'\mu_i + \ln P(\omega_i)$$



判决平面:  $g_i(x) = g_j(x)$

$$g_i(x) = w_i'x + w_{i0}$$

$$w_i = \frac{\mu_i}{\sigma^2}; \quad w_{i0} = -\frac{1}{2\sigma^2} \mu_i' \mu_i + \ln P(\omega_i)$$

$$\frac{\mu_i'}{\sigma^2} x - \frac{1}{2\sigma^2} \mu_i' \mu_i + \ln P(\omega_i) = \frac{\mu_j'}{\sigma^2} x - \frac{1}{2\sigma^2} \mu_j' \mu_j + \ln P(\omega_j)$$

$$w'(x - x_0) = 0 \quad w = \mu_i - \mu_j$$

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

判别边界是一条直线，且垂直于均值的连线。

二十四、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







## 《模式识别》试题答案及评分标准 (B 卷)

姓名: \_\_\_\_\_ 学号: \_\_\_\_\_ 成绩: \_\_\_\_\_

### 一、填空与选择填空 (本题答案写在此试卷上, 30 分)

- 1、模式识别系统的基本构成单元包括: 模式采集、特征提取与选择和 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、聚类分析算法属于 (1); 判别域代数界面方程法属于 (3)。  
(1) 无监督分类 (2) 有监督分类 (3) 统计模式识别方法 (4) 句法模式识别方法
- 4、若描述模式的特征量为 0-1 二值特征量, 则一般采用 (4) 进行相似性度量。  
(1) 距离测度 (2) 模糊测度 (3) 相似测度 (4) 匹配测度
- 5、下列函数可以作为聚类分析中的准则函数的有 (1) (3) (4)。

$$\begin{aligned} (1) \quad J &= \text{Tr}[S_W^{-1} S_B] & (2) \quad J &= |S_W S_B^{-1}| & (3) \quad J &= \sum_{j=1}^c \sum_{i=1}^{n_j} |\bar{x}_i^{(j)} - \bar{m}_j|^2 \\ (4) \quad J &= \sum_{j=1}^c (\bar{m}_j - \bar{m})' (\bar{m}_j - \bar{m}) \end{aligned}$$

- 6、Fisher 线性判别函数的求解过程是将 N 维特征矢量投影在 (2) 中进行。  
(1) 二维空间 (2) 一维空间 (3) N-1 维空间
- 7、下列判别域界面方程法中只适用于线性可分情况的算法有 (1); 线性可分、不可分都适用的有 (3)。  
(1) 感知器算法 (2) H-K 算法 (3) 积累位势函数法
- 8、下列四元组中满足文法定义的有 (1) (2) (4)。  
(1)  $(\{A, B\}, \{0, 1\}, \{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0, B \rightarrow BA, B \rightarrow 0\}, A)$   
(2)  $(\{A\}, \{0, 1\}, \{A \rightarrow 0, A \rightarrow 0A\}, A)$   
(3)  $(\{S\}, \{a, b\}, \{S \rightarrow 00S, S \rightarrow 11S, S \rightarrow 00, S \rightarrow 11\}, S)$   
(4)  $(\{A\}, \{0, 1\}, \{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0\}, A)$

### 二、(15 分) 简答及证明题

- (1) 影响聚类结果的主要因素有那些?
- (2) 证明马氏距离是平移不变的、非奇异线性变换不变的。

(1) (5 分)

答: 分类准则, 模式相似性测度, 特征量的选择, 量纲。

(2) (10 分)



证明:

$$d^2(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)' V^{-1} (\vec{x}_i - \vec{x}_j) \quad (2 \text{ 分})$$

$$V = \frac{1}{m-1} \sum_{i=1}^m (\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})' \quad (2 \text{ 分})$$

$$\bar{\vec{x}} = \frac{1}{m} \sum_{i=1}^m \vec{x}_i \quad (1 \text{ 分})$$

设, 有非奇异线性变换: (1 分)

$$\vec{y} = \frac{1}{m} \sum_{i=1}^m \vec{y}_i = \frac{1}{m} \sum_{i=1}^m A \vec{x}_i = A \frac{1}{m} \sum_{i=1}^m \vec{x}_i = A \bar{\vec{x}}$$

$$\begin{aligned} V_y &= \frac{1}{m-1} \sum_{i=1}^m (\vec{y}_i - \vec{y})(\vec{y}_i - \vec{y})' \\ &= \frac{1}{m-1} \sum_{i=1}^m (A \vec{x}_i - A \bar{\vec{x}})(A \vec{x}_i - A \bar{\vec{x}})' \\ &= \frac{1}{m-1} \sum_{i=1}^m A (\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})' A' \\ &= A \left[ \frac{1}{m-1} \sum_{i=1}^m (\vec{x}_i - \bar{\vec{x}})(\vec{x}_i - \bar{\vec{x}})' \right] A' = A V_x A' \end{aligned}$$

$$\begin{aligned} d_y^2(\vec{y}_i, \vec{y}_j) &= (\vec{y}_i - \vec{y}_j)' V_y^{-1} (\vec{y}_i - \vec{y}_j) \\ &= (A \vec{x}_i - A \vec{x}_j)' V_y^{-1} (A \vec{x}_i - A \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' V_y^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' (A V_x A')^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' A' A^{-1} V_x^{-1} A^{-1} A (\vec{x}_i - \vec{x}_j) \\ &= (\vec{x}_i - \vec{x}_j)' V_x^{-1} (\vec{x}_i - \vec{x}_j) \\ &= d_x^2(\vec{x}_i, \vec{x}_j) \end{aligned} \quad (4 \text{ 分})$$

三、(8 分)说明线性判别函数的正负和数值大小在分类中的意义并证明之。

答:

(1) (4 分)  $d(\vec{x})$  的绝对值  $|d(\vec{x})|$  正比于  $\vec{x}$  到超平面  $d(\vec{x})=0$  的距离  $d_r$ .

平面  $\pi$  的方程可以写成  $\frac{\vec{w}_0'}{\|\vec{w}_0'\|} \vec{x} = \frac{-w_{n+1}}{\|\vec{w}_0'\|}$

式中  $\|\vec{w}_0'\| = (w_1^2 + w_2^2 + \dots + w_n^2)^{1/2}$ , 于是  $\vec{n} = \frac{\vec{w}_0'}{\|\vec{w}_0'\|}$  是平面  $\pi$  的单位法矢量, 上式可写成

$$\vec{n}' \vec{x} = \frac{-w_{n+1}}{\|\vec{w}_0'\|}$$



设  $\vec{p}$  是平面  $\pi$  中的任一点,  $\vec{x}$  是特征空间  $X^n$  中任一点, 点  $\vec{x}$  到平面  $\pi$  的距离为差矢量  $(\vec{x} - \vec{p})$  在  $\vec{n}$  上的投影的绝对值, 即

$$\begin{aligned} d_x &= |\vec{n}'(\vec{x} - \vec{p})| = |\vec{n}'\vec{x} - \vec{n}'\vec{p}| = \left| \frac{\vec{w}_0'}{\|\vec{w}_0\|} \vec{x} - \frac{\vec{w}_0'}{\|\vec{w}_0\|} \vec{p} \right| \\ &= \left| \frac{\vec{w}_0'}{\|\vec{w}_0\|} \vec{x} + \frac{w_{n+1}}{\|\vec{w}_0\|} \right| \\ &= \frac{|\vec{w}_0' \vec{x} + w_{n+1}|}{\|\vec{w}_0\|} = \frac{1}{\|\vec{w}_0\|} |d(\vec{x})| \end{aligned} \quad (1-1)$$

上式中利用了  $\vec{p}$  在平面  $\pi$  中, 故满足方程

$$\frac{\vec{w}_0'}{\|\vec{w}_0\|} \vec{p} = \frac{-w_{n+1}}{\|\vec{w}_0\|}$$

式(1-1)的分子为判别函数绝对值, 上式表明,  $d(\vec{x})$  的值  $|d(\vec{x})|$  正比于  $\vec{x}$  到超平面  $d(\vec{x}) = 0$  的距离  $d_x$ , 一个特征矢量代入判别函数后所得值的绝对值越大表明该特征点距判别界面越远。

(2) (4分)  $d(\vec{x})$  的正(负)反映  $\vec{x}$  在超平面  $d(\vec{x}) = 0$  的正(负)侧

两矢量  $\vec{n}$  和  $(\vec{x} - \vec{p})$  的数积为

$$\vec{n}'(\vec{x} - \vec{p}) = \|\vec{n}\| \|\vec{x} - \vec{p}\| \cos(\vec{n}, (\vec{x} - \vec{p})) \quad (2分)$$

显然, 当  $\vec{n}$  和  $(\vec{x} - \vec{p})$  夹角小于  $90^\circ$  时, 即  $\vec{x}$  在  $\vec{n}$  指向的那个半空间中,  $\cos(\vec{n}, (\vec{x} - \vec{p})) > 0$ ; 反之, 当  $\vec{n}$  和  $(\vec{x} - \vec{p})$  夹角大于  $90^\circ$  时, 即  $\vec{x}$  在  $\vec{n}$  背向的那个半空间中,  $\cos(\vec{n}, (\vec{x} - \vec{p})) < 0$ 。由于  $\|\vec{w}_0\| > 0$ , 故  $\vec{n}'(\vec{x} - \vec{p})$  和  $\vec{w}_0' \vec{x} + w_{n+1}$  同号。所以, 当  $\vec{x}$  在  $\vec{n}$  指向的半空间中时,  $\vec{w}_0' \vec{x} + w_{n+1} > 0$ ; 当  $\vec{x}$  在  $\vec{n}$  背向的半空间中,  $\vec{w}_0' \vec{x} + w_{n+1} < 0$ 。判别函数值的正负表示出特征点位于哪个半空间中, 或者换句话说, 表示特征点位于界面的哪一侧。

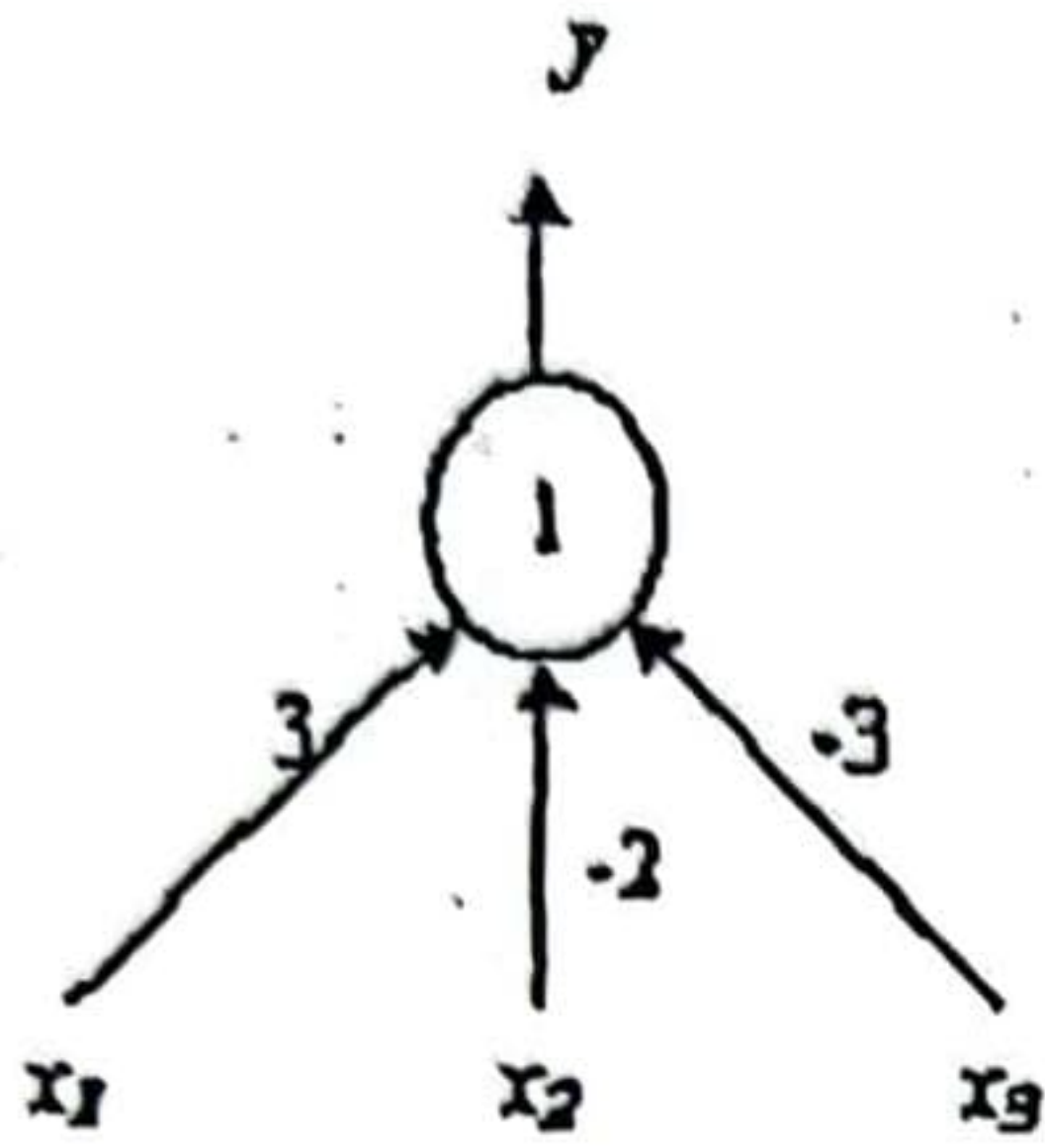
四、(10分) 已知样本集:

$\omega_1: \{(0, 0, 0)^T, (1, 0, 0)^T, (1, 0, 1)^T, (1, 1, 0)^T, \}$

$\omega_2: \{(0, 0, 1)^T, (0, 1, 1)^T, (0, 1, 0)^T, (1, 1, 1)^T, \}$

试利用感知器算法设计一个能对该样本集正确分类的人工神经网络。





解：（画出结构图 2 分，计算流程 4 分，结果正确 2 分）

$W=[3, -2, -3, 1]'$

算法：

```
%--- Step1. Input Pattern Vectors X.
X_input = [ 0 0 0; 1 0 0; 1 0 1; 1 1 0;...
            0 0 1; 0 1 1; 0 1 0; 1 1 1];

% Span and Normalize Vectors
[m,n] = size(X_input);
X = [X_input, ones(m, 1)];
[m,n] = size(X);
for i = 5:m
    X(i,1:n) = -X(i,1:n);
end

%--- Step2. Input Initial Weight Vectors W0
W = [ -1 -2 -2 0]; (知道初值可以任意设定, 1 分)

%--- Step3. The weight vector is corrected according to the preceding rule
k = 0;
j = 0;
while ((j < m) && (k < 1000)) % repeat until the algorithm converges to a solution
    for i = 1:m
        d = W * X(i,1:n)';
        if (d <= 0)
            j = 0;
            W = W + X(i,1:n);
        else
            j = j+1;
        end
    end
```



```

end
k = k+1;
end
%— Step4. Display the result
if (k >= 1000)
    fprintf('Not linearly separable! k=%i',k);
else
    fprintf('Weight Vector W = ');
    disp(W);
end

```

五、(12分, 每问4分) 在目标识别中, 假定有农田和装甲车两种类型, 类型  $\omega_1$  和类型  $\omega_2$  分别代表农田和装甲车, 它们的先验概率分别为 0.8 和 0.2, 损失函数如表 1 所示。现在做了三次试验, 获得三个样本的概率密度如下:

$$p(x/\omega_1): 0.3, 0.1, 0.6$$

$$p(x/\omega_2): 0.7, 0.8, 0.3$$

- (1) 试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型;
- (2) 假定只考虑前两种判决, 试用贝叶斯最小风险准则判决三个样本各属于哪一类;
- (3) 把拒绝判决考虑在内, 重新考核三次试验的结果。

表 1

损失 判决	类型	
	$\omega_1$	$\omega_2$
$\alpha_1$	1	4
$\alpha_2$	5	1
$\alpha_3$	1	1

解: 由题可知:  $P(\omega_1) = 0.7, P(\omega_2) = 0.3, \frac{P(\omega_2)}{P(\omega_1)} = \frac{3}{7}, \frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} = \frac{3}{7},$



$$\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} = \frac{1}{8}, \quad \frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} = 2$$

(1) (4分) 根据贝叶斯最小误判概率准则知:

$$\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} = \frac{P(\omega_1)}{P(\omega_2)}; \text{ 则可以任判,}$$

$$\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{P(\omega_1)}{P(\omega_2)}, \text{ 则判为 } \omega_2,$$

$$\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{P(\omega_1)}{P(\omega_2)}, \text{ 则判为 } \omega_1;$$

$$(2) (4分) \text{ 由题可知: } \frac{P(\omega_2)(\lambda_{21}-\lambda_{22})}{P(\omega_1)(\lambda_{12}-\lambda_{11})} = \frac{0.3(5-1)}{0.7(4-1)} = \frac{4}{7}$$

$$\text{则 } \frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} < \frac{4}{7}, \text{ 判为 } \omega_2;$$

$$\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{4}{7}, \text{ 判为 } \omega_2;$$

$$\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{4}{7}, \text{ 判为 } \omega_1;$$

(3) (4分) 对于两类问题, 对于样本  $x$ , 假设  $P(x)$  已知, 有

$$\begin{aligned} R(\alpha_j|x) &= \lambda(\alpha_j|\omega_1)P(\omega_1|x) + \lambda(\alpha_j|\omega_2)P(\omega_2|x) = \\ &= \frac{\lambda(\alpha_j|\omega_1)P(x|\omega_1)P(\omega_1) + \lambda(\alpha_j|\omega_2)P(x|\omega_2)P(\omega_2)}{P(x)} \end{aligned}$$

则对于第一个样本,

$$R(\alpha_1|x) = \frac{5 \times 0.21}{P(x)}, R(\alpha_2|x) = \frac{4 \times 0.21}{P(x)}, R(\alpha_3|x) = \frac{2 \times 0.21}{P(x)}, \text{ 则拒判;}$$

$$R(\alpha_1|x) = \frac{1.03}{P(x)}, R(\alpha_2|x) = \frac{0.59}{P(x)}, R(\alpha_3|x) = \frac{0.24}{P(x)}, \text{ 则拒判;}$$

$$R(\alpha_1|x) = \frac{0.78}{P(x)}, R(\alpha_2|x) = \frac{2.19}{P(x)}, R(\alpha_3|x) = \frac{0.51}{P(x)}, \text{ 拒判.}$$

六、(30分) 上机实验及作业 (时间另计)。

(评分标准: 四个上机实验各5分 (不能处理任意维数样本扣1分), 平时作业10分。)



# 中国科学技术大学模式识别试题

(2012 年春季学期)

姓名: \_\_\_\_\_ 学号: \_\_\_\_\_ 成绩: \_\_\_\_\_

## 一、填空与选择填空(本题答案写在此试卷上, 30 分)

- 1、模式识别系统的基本构成单元包括: \_\_\_\_\_ 和 \_\_\_\_\_。
- 2、统计模式识别中描述模式的方法一般使用 \_\_\_\_\_; 句法模式识别中模式描述方法一般有 \_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_。
- 3、聚类分析算法属于 \_\_\_\_\_, 判别域代数界面方程法属于 \_\_\_\_\_。  
(1) 无监督分类 (2) 有监督分类 (3) 统计模式识别方法 (4) 句法模式识别方法
- 4、若描述模式的特征量为 0-1 二值特征量, 则一般采用 \_\_\_\_\_ 进行相似性度量。  
(1) 距离测度 (2) 模糊测度 (3) 相似测度 (4) 匹配测度
- 5、下列函数可以作为聚类分析中的准则函数的有 \_\_\_\_\_。

$$(1) J = \text{Tr}[S_W^{-1} S_B] \quad (2) J = |S_W S_B^{-1}| \quad (3) J = \sum_{j=1}^c \sum_{i=1}^{n_j} \|\bar{x}_i^{(j)} - \bar{m}_j\|^2$$
$$(4) J = \sum_{j=1}^c (\bar{m}_j - \bar{m})' (\bar{m}_j - \bar{m})$$

- 6、Fisher 线性判别函数的求解过程是将 N 维特征矢量投影在 \_\_\_\_\_ 中进行。  
(1) 二维空间 (2) 一维空间 (3) N-1 维空间
- 7、下列判别域界面方程法中只适用于线性可分情况的算法有 \_\_\_\_\_; 线性可分、不可分都适用的有 \_\_\_\_\_。  
(1) 感知器算法 (2) H-K 算法 (3) 积累位势函数法
- 8、下列四元组中满足文法定义的有 \_\_\_\_\_。  
(1)  $(\{A, B\}, \{0, 1\}, \{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0, B \rightarrow BA, B \rightarrow 0\}, A)$   
(2)  $(\{A\}, \{0, 1\}, \{A \rightarrow 0, A \rightarrow 0A\}, A)$   
(3)  $(\{S\}, \{a, b\}, \{S \rightarrow 00S, S \rightarrow 11S, S \rightarrow 00, S \rightarrow 11\}, S)$   
(4)  $(\{A\}, \{0, 1\}, \{A \rightarrow 01, A \rightarrow 0A1, A \rightarrow 1A0\}, A)$

## 二、(15 分) 简答及证明题

- (1) 影响聚类结果的主要因素有那些?
- (2) 证明马氏距离是平移不变的、非奇异线性变换不变的。
- (3) 画出对样本集

$$\omega: \{(0, 0, 0)^T, (1, 0, 0)^T, (1, 0, 1)^T, (1, 1, 0)^T\}$$



$$\omega_2: \{(0,0,1)^T, (0,1,1)^T, (0,1,0)^T, (1,1,1)^T\}$$

实现正确分类的神经网络模型。

四、(10 分) 说明线性判别函数的正负和数值大小在分类中的意义并证明之。

三、(15 分) 在目标识别中, 假定有农田和装甲车两种类型, 类型  $\omega_1$  和类型  $\omega_2$  分别代表农田和装甲车, 它们的先验概率分别为 0.8 和 0.2, 损失函数如表 1 所示。现在做了三次试验, 获得三个样本的类概率密度如下:

$$p(x/\omega_1): 0.3, 0.1, 0.6$$

$$p(x/\omega_2): 0.7, 0.8, 0.3$$

- (1) 试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型;
- (2) 假定只考虑前两种判决, 试用贝叶斯最小风险准则判决三个样本各属于哪一类;
- (3) 把拒绝判决考虑在内, 重新考核三次试验的结果。

表 1

类型 损失	$\omega_1$	$\omega_2$
判决		
$\alpha_1$	0.5	3
$\alpha_2$	5	1
$\alpha_3$	2	2



# 中国科学技术大学

2013-2014 学年第二学期研究生考试试卷

考试科目: 模式识别

得分: \_\_\_\_\_

学生所在系: 10系 姓名: 徐子伟 学号: 517010071

## 1. 填空 (每小题2分, 共10分)

(1) 设  $X$  为  $d$  维的正态随机向量, 其对应的协方差矩阵为  $\Sigma$ , 则  $X$  的各个分量是彼此相互独立的正态随机变量。

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix}$$

(2) 设  $G = (N, \Sigma, P, S)$ , 其中:

$N = \{S\}$ ,  $\Sigma = \{a, b\}$ , 以及

$P: (1) S \rightarrow aSb, (2) S \rightarrow bSa, (3) S \rightarrow a, (4) S \rightarrow b$ .

则  $L(G) = \{a^n b^n \mid n \geq 1\}$  或  $\{a^n b^n \mid n \geq 0\}$  或  $\{a^n b^n \mid n \geq 1\}$  或  $\{a^n b^n \mid n \geq 0\}$

(3) 设  $a, b, c$  属于字母表, 则句子  $abcba$  的所有子串是:

$a, b, c, ab, bc, ca, abc, bcb, cba, abcba$

(4) 两点间的欧氏距离是  $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

(5) 设  $a, b, c$  属于字母表, 则根据定义,  $a^*b^*c^*$  表示  $\{a^n b^m c^k \mid n, m, k \geq 0\}$

## 2. 是非判断题 (每小题2分, 共20分)

(1) 对于一个模式识别问题而言, 如果训练样本是正交的, 则训练样本是正交的。

(2) 设  $T$  是由一个非确定性的有限状态自动机。

(3) 设  $T$  是由一个非确定性的有限状态自动机。

(4) 设  $T$  是由一个非确定性的有限状态自动机。

(5) 设  $T$  是由一个非确定性的有限状态自动机。



- (7) 服从正态分布的若干分布不一定是正态分布。 (X)
- (8) 所谓第一类是从  $n$  个可用样本中任意选取  $(n-1)$  个样本进行分类型设计，并用剩下的一个样本对所设计的分类型进行检验的一种方法。 (X)
- (9) 三角形集合上的相似关系是等价关系。 (X)
- (10) 设  $X$  是  $d$  维的正态随机向量， $\alpha$  是与  $X$  同维的正态随机向量，则  $\alpha^T X$  是一个  $d$  维的正态随机向量。 (✓)

3. 设  $X$  是  $d$  维随机向量，其均值向量和协方差矩阵分别为  $\mu$  和  $\Sigma$ ，证明： $X^T \Sigma^{-1} \mu = \mu^T \Sigma^{-1} X$ 。  
(共 10 分)

证明 对于多元正态分布，均值为  $\mu$ ，协方差为  $\Sigma$   
对任意  $X$  和  $\mu$  有

$$d^2(X, \mu) = (X - \mu)^T \Sigma^{-1} (X - \mu) \text{ 且在 } E(X) = \mu$$

$$\therefore d^2(X, \mu) = (X^T \Sigma^{-1} X) - X^T \Sigma^{-1} \mu - \mu^T \Sigma^{-1} X + \mu^T \Sigma^{-1} \mu$$

$$\therefore X \text{ 与 } \mu \text{ 同属一个样本} \therefore d^2(X, \mu) = 0$$

$$\therefore \text{有 } X^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} X = X^T \Sigma^{-1} X + \mu^T \Sigma^{-1} \mu$$

$$\text{又 } E(X^T \Sigma^{-1} \mu) = E(X^T) \Sigma^{-1} \mu = \mu^T \Sigma^{-1} \mu$$

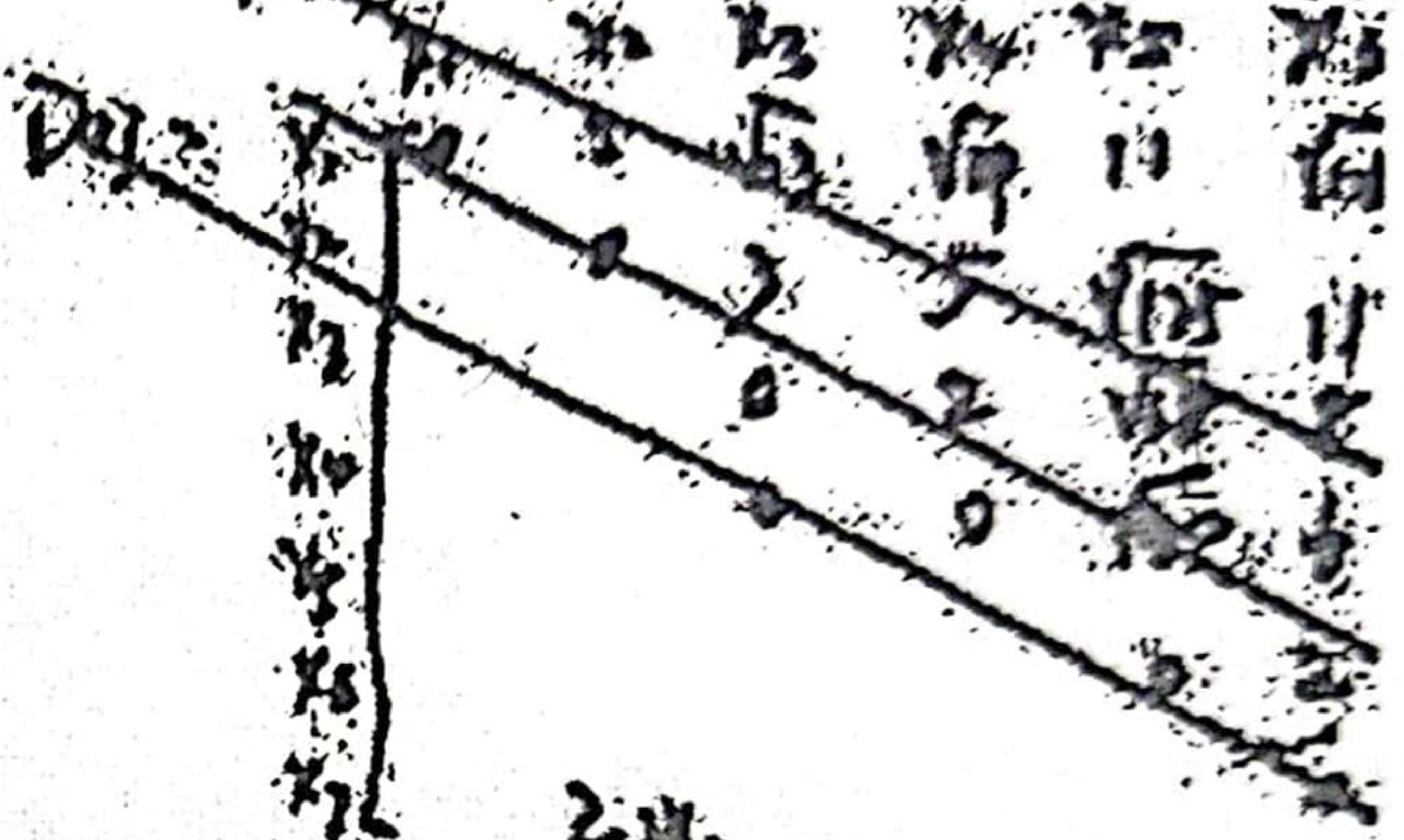
$$\text{从而有 } E(\mu^T \Sigma^{-1} X) = \mu^T \Sigma^{-1} \mu$$

$$\therefore \mu^T \Sigma^{-1} X = \mu^T \Sigma^{-1} \mu$$



聚苯異氰酸酯成膜分析，給出聚異結果。〔第20卷〕

心欲養而無以事



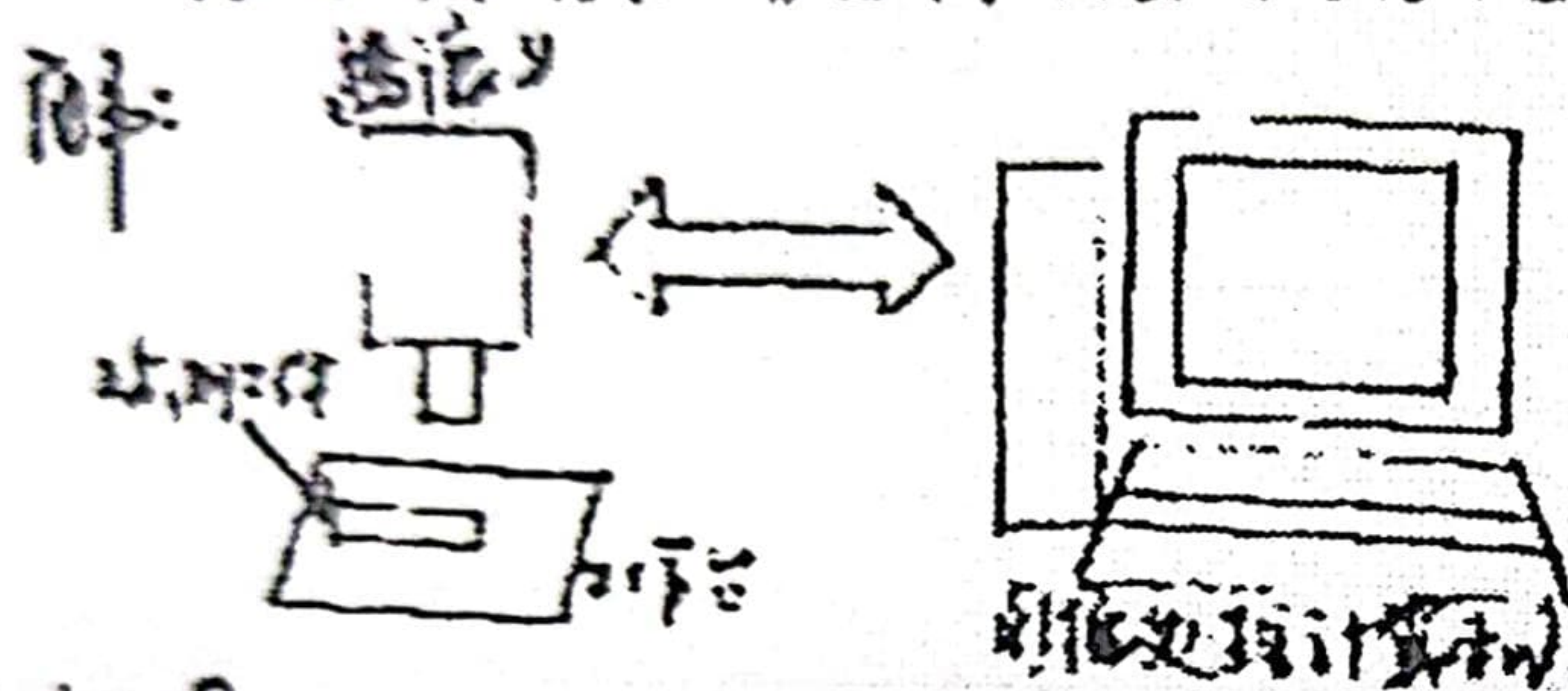
1945年10月1日

$$Y_1, Y_2, Y_3, Y_4, \dots, Y_n$$

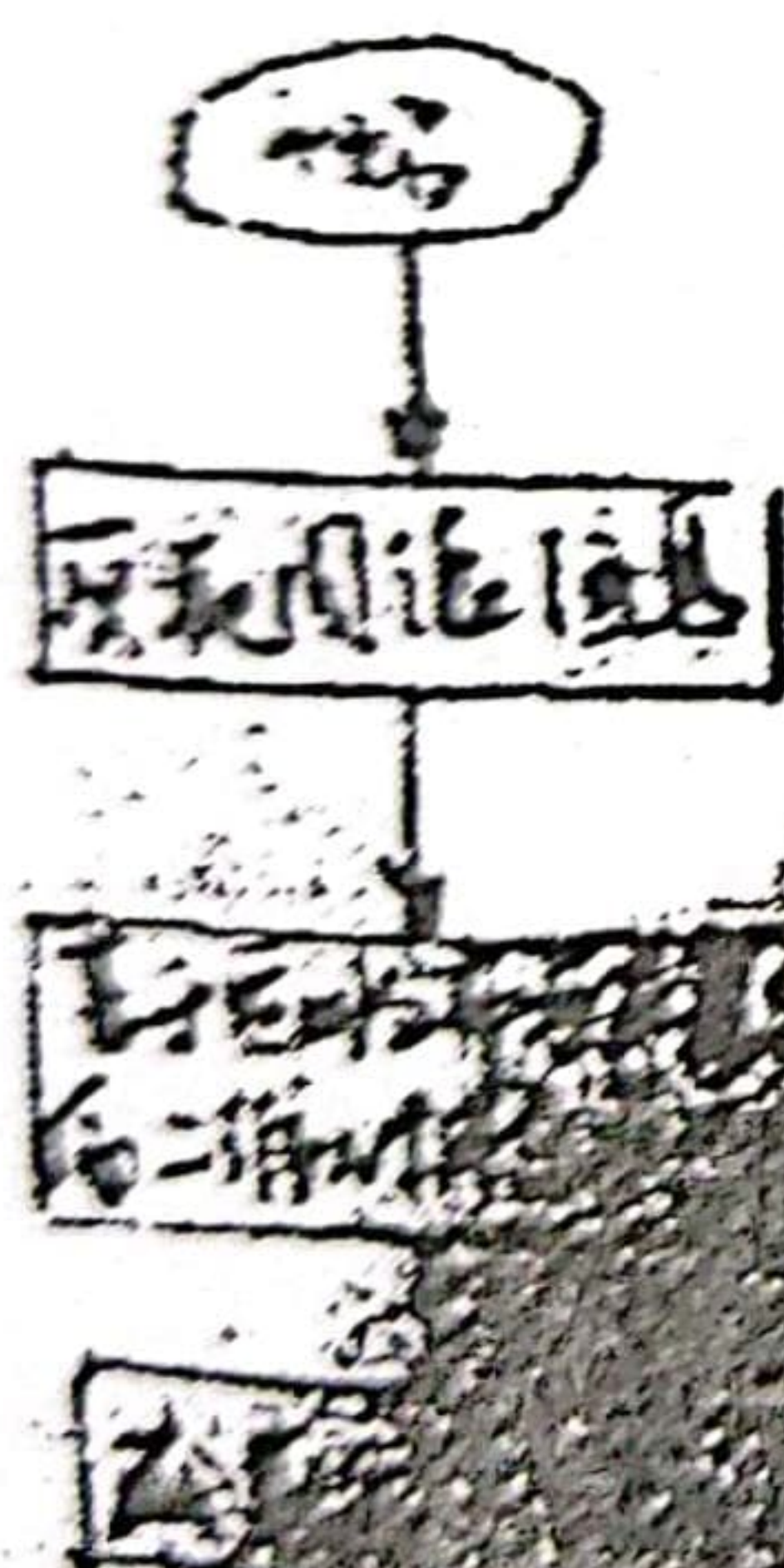
12	7	3	1
0	10	5	4
3	2	6	8
9	11	4	7



7. 在一个工件的图像检测应用中，需从背景中正确检测出待测工件。已知待测工件在图像中表现为一个连续区域，其所占据的面积和整个图像的面积之比  $\theta$ ，待测工件的灰度概率密度函数近似呈正态分布，均值未知，但和背景均值有极大差别。另外，背景的灰度概率密度函数也近似呈正态分布，均值已知。现要求充分利用关于待测工件的空间分布和灰度分布的先验信息，设计一个算法将待测工件从图像中正确地提取出来。要求结合图示，简要说明设计思想，并画出程序框图。（共 20 分）



程序框图：



该系统由工作台、摄像头以及图像采集卡、计算机等组成。其中，工作台和摄像头分别对工件和背景进行成像，摄像头采集到的图像数据通过图像采集卡输入到计算机中，计算机对图像数据进行图像处理，提取出工件的轮廓信息。



# 《模式识别》试题库

## 一、基本概念题

1.1 模式识别的三大核心问题是：\_\_\_\_\_。

1.2、模式分布为团状时，选用\_\_\_\_\_聚类算法较好。

1.3 欧式距离具有\_\_\_\_\_。 马式距离具有\_\_\_\_\_。

(1) 平移不变性 (2) 旋转不变性 (3) 尺度缩放不变性 (4) 不受量纲影响的特性

1.4 描述模式相似的测度有：\_\_\_\_\_。

(1) 距离测度 (2) 模糊测度 (3) 相似测度 (4) 匹配测度

1.5 利用两类方法处理多类问题的技术途径有：(1) \_\_\_\_\_；(2) \_\_\_\_\_；

(3) \_\_\_\_\_。其中最常用的是第\_\_\_\_\_个技术途径。

1.6 判别函数的正负和数值大小在分类中的意义

是：\_\_\_\_\_。

1.7 感知器算法\_\_\_\_\_。

(1) 只适用于线性可分的情况；(2) 线性可分、不可分都适用。

1.8 积累位势函数法的判别界面一般为\_\_\_\_\_。

(1) 线性界面；(2) 非线性界面。

1.9 基于距离的类别可分性判据有：\_\_\_\_\_。

$$(1) \text{Tr}[S_w^{-1}S_B] \quad (2) \frac{S_B}{S_W} \quad (3) \left| \frac{S_B}{S_W + S_B} \right|$$

1.10 作为统计判别问题的模式分类，在( )情况下，可使用聂曼-皮尔逊判决准则。

1.11 确定性模式非线性分类的势函数法中，位势函数  $K(x, x_1)$  与积累位势函数  $K(x)$  的关系为

( )。

1.12 用作确定性模式非线性分类的势函数法，通常，两个  $n$  维向量  $x$  和  $x_1$  的函数  $K(x, x_1)$  若同时满足下列三个条件，都可作为势函数。

① ( )；

② ( )；

③  $K(x, x_1)$  是光滑函数，且是  $x$  和  $x_1$  之间距离的单调下降函数。

1.13 散度  $J_{ij}$  越大，说明  $\omega_i$  类模式与  $\omega_j$  类模式的分布( )。当  $\omega_i$  类模式与  $\omega_j$  类模式的分布相同时， $J_{ij} = ( )$ 。



1.14 若用 Parzen 窗法估计模式的类概率密度函数, 窗口尺寸  $h_1$  过小可能产生的问题是

( ),  $h_1$  过大可能产生的问题是 ( )。

1.15 信息熵可以作为一种可分性判据的原因

是: \_\_\_\_\_。

1.16 作为统计判别问题的模式分类, 在 ( ) 条件下, 最小损失判决规则与最小错误判决规则是等价的。

1.17 随机变量  $l(\bar{x}) = p(\bar{x}|\omega_1)/p(\bar{x}|\omega_2)$ ,  $l(\bar{x})$  又称似然比, 则  $E\{l(\bar{x})|\omega_2\} = ( )$ 。在最小误判概率准则下, 对数似然比 Bayes 判决规则为 ( )。

1.18 影响类概率密度估计质量的最重要因素是

( )。

1.19 基于熵的可分性判据定义为  $J_H = E_x[-\sum_{i=1}^c P(\omega_i|\bar{x}) \log P(\omega_i|\bar{x})]$ ,  $J_H$  越 ( ), 说明模式的可分性越强。

当  $P(\omega_i|\bar{x}) = ( )$  ( $i=1, 2, \dots, c$ ) 时,  $J_H$  取极大值。

1.20  $K_n$  近邻元法较之于 Parzen 窗法的优势在于

( )。

上述两种算法的共同弱点主要是 ( )。

1.21 已知有限状态自动机  $A_f = (\Sigma, Q, \delta, q_0, F)$ ,  $\Sigma = \{0, 1\}$ ;  $Q = \{q_0, q_1\}$ ;

$\delta: \delta(q_0, 0) = q_1, \delta(q_0, 1) = q_1, \delta(q_1, 0) = q_0, \delta(q_1, 1) = q_0; q_0 = q_0; F = \{q_0\}$ 。现有输入字符串: (a) 00011101011,

(b) 1100110011, (c) 101100111000, (d) 0010011, 试问, 用  $A_f$  对上述字符串进行分类的结果为

( )。

1.22 句法模式识别中模式描述方法有: \_\_\_\_\_。

(1) 符号串 (2) 树 (3) 图 (4) 特征向量

1.23 设集合  $X = \{a, b, c, d\}$  上的关系,  $R = \{(a, a), (a, b), (a, d), (b, b), (b, a), (b, d), (c, c), (d, d), (d, a), (d, b)\}$ , 则

$a, b, c, d$  生成的  $R$  等价类分别为 (  $[a]_R =$  ,  $[b]_R =$  ,  $[c]_R =$  ,

$[d]_R =$  )。

1.24 如果集合  $X$  上的关系  $R$  是传递的, ( ) 和 ( ) 的, 则称  $R$  是一个等价关系。

1.25 一个模式识别系统由那几部分组成? 画出其原理框图。

1.26 统计模式识别中, 模式是如何描述的。

1.27 简述随机矢量之间的统计关系: 不相关, 正交, 独立的定义及它们之间的关系。

1.28 试证明, 对于正态分布, 不相关与独立是等价的。

1.29 试证明, 多元正态随机矢量的线性变换仍为多元正态随机矢量。

1.30 试证明, 多元正态随机矢量  $\bar{X}$  的分量的线性组合是一正态随机变量。



## 第二部分 分析、证明、计算题

### 第二章 聚类分析

2.1 影响聚类结果的主要因素有哪些?

2.2 马氏距离有哪些优点?

2.3 如果各模式类呈现链状分布, 衡量其类间距离用最小距离还是用最大距离? 为什么?

2.4 动态聚类算法较之于简单聚类算法的改进之处何在? 层次聚类算法是动态聚类算法吗? 比较层次聚类算法与c-均值算法的优劣。

2.5 ISODATA 算法较之于 c-均值算法的优势何在?

2.6 简述最小张树算法的优点。

2.7 证明马氏距离是平移不变的、非奇异线性变换不变的。

2.8 设, 类  $\omega_p$ 、 $\omega_q$  的重心分别为  $\bar{x}_p$ 、 $\bar{x}_q$ , 它们分别有样本  $n_p$ 、 $n_q$  个。将和  $\omega_q$  合并为  $\omega_l$ , 则  $\omega_l$  有  $n_l = n_p + n_q$  个样本。另一类  $\omega_k$  的重心为  $\bar{x}_k$ 。试证明  $\omega_k$  与  $\omega_l$  的距离平方是

$$D_{kl}^2 = \frac{n_p}{n_k + n_l} D_{kp}^2 + \frac{n_q}{n_k + n_l} D_{kq}^2 - \frac{n_p n_q}{n_k + n_l} D_{pq}^2$$

2.9 (1) 设有 M 类模式  $\omega_i$ ,  $i=1, 2, \dots, M$ , 试证明总体散布矩阵  $S_T$  是总类内散布矩阵  $S_w$  与类间散布矩阵  $S_b$  之和, 即  $S_T = S_w + S_b$ 。

(2) 设有二维样本:  $x_1=(-1, 0)^T$ ,  $x_2=(0, -1)^T$ ,  $x_3=(0, 0)^T$ ,  $x_4=(2, 0)^T$  和  $x_5=(0, 2)^T$ 。试选用一种合适的方法进行一维特征提取  $y_i = W^T x_i$ 。要求求出变换矩阵  $W$ , 并求出变换结果  $y_i$ , ( $i=1, 2, 3, 4, 5$ )。

(3) 根据 (2) 特征提取后的一维特征, 选用一种合适的聚类算法将这些样本分为两类, 要求每类样本个数不少于两个, 并写出聚类过程。

2.10 (1) 试给出 c-均值算法的算法流程图;

(2) 试证明 c-均值算法可使误差平方和准则

$$J^{(k)} = \sum_{j=1}^c \sum_{\bar{x}_i \in \omega_j^{(k)}} (\bar{x}_i - \bar{z}_j^{(k)})^T (\bar{x}_i - \bar{z}_j^{(k)})$$

最小。

其中,  $k$  是迭代次数;  $\bar{z}_j^{(k)}$  是  $\omega_j^{(k)}$  的样本均值。

2.11 现有  $2k+1$  个一维样本, 其中  $k$  个样本在  $x=-2$  处重合, 另  $k$  个样本在  $x=0$  处重合, 只有 1 个在  $x=a>0$  处。若  $a=2(k+1)$ , 证明, 使误差平方和准则  $J_c$  最小的两类划分是  $x=0$  处的  $k$  个样本与  $x=a$  处的 1 个样本为一类, 其余为另一类。这里,

$$c = N_j$$

$$J_c = \sum_{j=1}^c \sum_{i=1}^N (x_i - m_j)^2$$

$$j=1 \quad i=1$$



其中,  $c$  为类别数,  $N_j$  是第  $j$  类的样本个数,  $x_i \in \omega_j, i=1, 2, \dots, N_j$ ,  $m_j$  是第  $j$  类的样本均值。

2.12 有样本集  $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ , 试用谱系聚类算法对其分类。

2.13 设有样本集  $S = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ , 证明类心  $\bar{z}$  到  $S$  中各样本点距离平方和  $\sum_{i=1}^n (\bar{x}_i - \bar{z})^T (\bar{x}_i - \bar{z})$  为最小时, 有

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i$$

2.14 假设  $s$  为模式矢量集  $X$  上的距离相似侧度, 有  $\forall x, y > 0, s(x, y) > 0$  且当  $a > 0$  时,  $d(x, y) = a/s(x, y)$ 。证明  $d$  是距离差异性测度。

2.15 证明欧氏距离满足旋转不变性。

提示: 运用 Minkowski 不等式, 对于两矢量  $x = [x_1, \dots, x_l]^T$  和

$d_{\min}^{ss}(s_{\min}^{ss}), d_{\max}^{ss}(s_{\max}^{ss}), d_{\text{avg}}^{ss}(s_{\text{avg}}^{ss}), d_{\text{mean}}^{ss}(s_{\text{mean}}^{ss})d(s)$ , 满足

$$\left( \sum_{i=1}^l |x_i + y_i|^p \right)^{1/p} \leq \left( \sum_{i=1}^l |x_i|^p \right)^{1/p} + \left( \sum_{i=1}^l |y_i|^p \right)^{1/p}$$

2.16 证明:

(a) 如果  $s$  是类  $X$  上的距离相似侧度,  $\forall x, y > 0, s(x, y) > 0$ , 那么对于  $\forall a > 0$ ,  $s(x, y) + a$  也是类  $X$  上的距离测度。

(b) 如果  $d$  是类  $X$  上的距离差异性测度, 那么对于  $\forall a > 0$ ,  $d + a$  也是类  $X$  上的距离差异性测度

2.17 假设  $f: R^+ \rightarrow R^+$  是连续单调递增函数, 满足

$$f(x) + f(y) \geq f(x + y), \forall x, y \in R^+$$

$d$  是类  $X$  上的距离差异性测度且  $d_0 \geq 0$ 。证明  $f(d)$  也是类  $X$  上的距离差异性测度。

2.18 假设  $s$  为类  $X$  上的距离相似侧度, 有  $\forall x, y > 0, s(x, y) > 0$ ,  $f: R^+ \rightarrow R^+$  是连续单调递增函数, 满足

$$f(x) + f(y) \geq f\left(\frac{1}{\frac{1}{x} + \frac{1}{y}}\right), \forall x, y \in R^+$$

证明  $f(x)$  是  $X$  上的距离相似侧度。

2.19 证明: 对于模式矢量集  $X$  上任意两个矢量  $\bar{x}$  和  $\bar{y}$  有

$$d_{\infty}(\bar{x}, \bar{y}) \leq d_2(\bar{x}, \bar{y}) \leq d_1(\bar{x}, \bar{y})$$



- 2.20 (a) 证明公式  $s_F^q(\bar{x}, \bar{y}) = \left( \sum_{i=1}^l s(\bar{x}_i, \bar{y}_i)^q \right)^{1/q}$  中  $s_F(\bar{x}, \bar{y})$  的最大最小值分别是  $0.5l^{1/q}$ .

- (b) 证明当  $q \rightarrow +\infty$  时, 公式  $s_F^q(\bar{x}, \bar{y}) = \left( \sum_{i=1}^l s(\bar{x}_i, \bar{y}_i)^q \right)^{1/q}$  中  $s_F(\bar{x}, \bar{y}) = \max_{1 \leq i \leq l} s(\bar{x}_i, \bar{y}_i)$

- 2.21 假设  $d$  是模式矢量集  $X$  上的差异性测度,  $s = d_{\max} - d$  是相应相似测度.

证明  $S_{avg}^{ps}(x, C) = d_{\max} - d_{avg}^{ps}(x, C), \forall x \in X, C \subset X$

其中  $S_{avg}^{ps}$  和  $d_{avg}^{ps}$  是分别根据  $s$  和  $d$  所定义的.  $\Psi_{avg}^{ps}$  的定义来自于下面公式, 其中第一个集合只含有一个矢量.

提示: 平均亲近函数

$$\Psi_{avg}^{ps}(D_i, D_j) = \frac{1}{n_{D_i} n_{D_j}} \sum_{x \in D_i} \sum_{y \in D_j} \Psi(x, y)$$

, 其中  $n_{D_i}$  和  $n_{D_j}$  分别是集合  $D_i$  和  $D_j$  的势. 即使  $\Psi$

是测度, 显然  $\Psi_{avg}^{ps}$  不是测度. 在公式中,  $D_i$  和  $D_j$  中的所有矢量都参与计算.

- 2.22 假设  $x, y \in \{0, 1\}^l$ . 证明  $d_2(x, y) = \sqrt{d_{Hamming}(x, y)}$ .

- 2.23 考虑一维空间的两矢量,  $x = [x_1, \dots, x_l]^T$  和  $y = [y_1, \dots, y_l]^T$ ,  $|x_i - y_i| = \max_{j=1, \dots, l} \{|x_j - y_j|\}$ , 定义距离  $d_n(x, y)$  为

$$d_n(x, y) = |x_i - y_i| + \frac{1}{l - [(l-2)/2]} \sum_{j=1, j \neq i}^l |x_j - y_j|$$

这个距离曾被提议作为欧氏距离的近似值.

- (a) 证明  $d_n$  是距离.

- (b) 比较  $d_n$  和  $d_2$  的计算复杂度.

- 2.24 若定义下列准则函数

$$J_T = \sum_{i=1}^c \sum_{x \in X_i} (x - m_i)^T S_T^{-1} (x - m_i)$$

其中  $m_i$  是  $X_i$  中  $N_i$  个样本的均值向量,  $S_T$  是总散布矩阵,

- (1) 证明  $J_T$  对数据的非奇异线性变换具有不变性.

- (2) 证明把  $X_i$  中的样本  $\hat{x}$  转移到  $X_j$  中去, 则使  $J_T$  改变为



$$J_T^* = J_T - \left[ \frac{N_j}{N_j + 1} (\hat{x} - m_j)^T S_T^{-1} (\hat{x} - m_j) - \frac{N_j}{N_j - 1} (\hat{x} - m_j)^T S_T^{-1} (\hat{x} - m_j) \right]$$

(3) 写出使  $J_T$  最小化的迭代程序。

2. 25 证明对于 C-均值算法, 聚类准则函数满足使算法收敛的条件。(即若  $J(\Gamma, \tilde{K}) \leq J(\Gamma, K)$ , 则有

$$J(\tilde{\Gamma}, \tilde{K}) \leq J(\Gamma, \tilde{K})$$

2. 26 令  $\Delta(y, K_j) = \frac{1}{2} (y - m_j)^T \Sigma_j^{-1} (y - m_j) + \frac{1}{2} \log |\Sigma_j|$  是点到聚类的相似性度量, 式中  $m_j$  和  $\Sigma_j$  是聚类  $\Gamma_j$  的

均值和协方差矩阵, 若把一点从  $\Gamma_i$  转移到  $\Gamma_j$  中去, 计算由公式

$$J_K = \sum_{i=1}^c \sum_{y \in \Gamma_i} \Delta(y, K_j)$$

所示  $J_K$  的变化值。

### 第三章 判别域代数界面方程法

3.1 证明感知器算法在训练模式是线性可分的情况下, 经过有限次迭代后可以收敛到正确的解向量  $w^*$ 。

3.2

(1) 试给出 LMSE 算法 (H-K 算法) 的算法流程图;

(2) 试证明  $X^* e(k) = 0$ , 这里,  $X^*$  是伪逆矩阵;  $e(k)$  为第  $k$  次迭代的误差向量;

(3) 已知两类模式样本  $\omega_1$ :  $x_1 = (-1, 0)^T$ ,  $x_2 = (1, 0)^T$ ;  $\omega_2$ :  $x_3 = (0, 0)^T$ ,  $x_4 = (0, -1)^T$ 。

试用 LMSE 算法判断其线性可分性。

3.3 设等式方程组  $Xw = \bar{b}$ , 其中: 属于  $\omega_1$  的样本作为  $X$  的前  $N_1$  行, 属于  $\omega_2$  的样本作为  $X$  的后  $N_2$  行。证明:

$$b = \left( \underbrace{\frac{N}{N_1}, \dots, \frac{N}{N_1}}_{N_1}, \underbrace{\frac{N}{N_2}, \dots, \frac{N}{N_2}}_{N_2} \right)$$

当余量矢量

时, MSE 解等价于 Fisher 解。

3.4 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  $\bar{x}_5 = (0, 2)^T$ ,  $\{\bar{x}_1, \bar{x}_2, \bar{x}_3\} \in \omega_1$ ,  $\{\bar{x}_4, \bar{x}_5\} \in \omega_2$ 。

试用感知器算法求出分类决策函数, 并判断  $\bar{x}_6 = (1, 1)^T$  属于哪一类?

3.4. 已知模式样本  $x_1 = (0, 0)^T$ ,  $x_2 = (1, 0)^T$ ,  $x_3 = (-1, 1)^T$  分别属于三个模式类别, 即,  $x_1 \in \omega_1$ ,  $x_2 \in \omega_2$ ,  $x_3 \in \omega_3$ ,

(1) 试用感知器算法求判别函数  $g_i(x)$ , 使之满足, 若  $x_i \in \omega_i$  则  $g_i(x) > 0$ ,  $i = 1, 2, 3$ ;

(2) 求出相应的判决界面方程, 并画出解区域的示意图。

给定校正增量因子  $C=1$ , 初始值可以取:



$w_1(1)=(4, -9, -4)^T$ ,  $w_2(1)=(4, 1, -4)^T$ ,  $w_3(1)=(-4, -1, -6)^T$ .

3.5 已知  $\omega_1: \{(0, 0)^T\}$ ,  $\omega_2: \{(1, 1)^T\}$ ,  $\omega_3: \{(-1, 1)^T\}$ . 用感知器算法求该三类问题的判别函数, 并画出解区域.

3.6 试证明:

(1) 从  $\bar{x}$  到超平面  $g(\bar{x}) = \bar{w}^T \bar{x} + w_0 = 0$  的距离  $r = \frac{|g(\bar{x})|}{\|\bar{w}\|}$  是在  $g(\bar{x}_q) = 0$  的约束条件下, 使  $\|\bar{x} - \bar{x}_q\|^2$  达到极小的解.

(2)  $\bar{x}$  在超平面上的投影是  $\bar{x}_p = \bar{x} - \frac{g(\bar{x})}{\|\bar{w}\|^2} \bar{w}$ .

3.7 设有一维空间二次判别函数  $g(x) = 5 + 7x + 9x^2$ , 试将其映射成广义齐次线性判别函数  $g(\bar{x}) = \bar{a}^T \bar{y}$ .

3.8 对二维线性判别函数  $g(\bar{x}) = x_1 + 2x_2 - 2$

(1) 将判别函数写成  $g(\bar{x}) = \bar{w}^T \bar{x} + w_0$  的形式, 并画出  $g(\bar{x}) = 0$  的几何图形;

(2) 将其映射成广义齐次线性判别函数  $g(\bar{x}) = \bar{a}^T \bar{y}$ ;

(3) 指出上述  $X$  空间实际是  $Y$  空间的一个子空间, 且  $\bar{a}^T \bar{y} = 0$  对  $X$  子空间的划分与原空间中  $\bar{w}^T \bar{x} + w_0 = 0$  对原  $X$  空间的划分相同, 并在图上表示出来.

3.9 指出在 Fisher 线性判别中,  $\bar{w}$  的比例因子对 Fisher 判别结果无影响的原因.

3.10 证明两向量外积组成的矩阵一般是奇异的.

3.11 证明, 在几何上, 感知器准则函数值正比于被错分类样本到决策面的距离之和.

3.12 解释为什么感知器函数是一个连续分段的线性分类器.

3.13 如果在感知器算法中  $\rho_i = \rho$ , 那么在  $k_0 = \frac{\|w(0) - \alpha w^*\|}{\beta^2 \rho (2 - \rho)}$  步之后, 这个算法收敛, 其中  $\alpha = \frac{\beta^2}{|\gamma|}$ ,  $\rho < 2$ .

3.14 证明感知器算法的正确分类和错误分类在有限个反复的运算以后是收敛的



3.15 考虑一种情况, 在类  $\omega_1$  中包含两个特征向量,  $[0,1]^T$ , 类  $\omega_2$  中包含  $[1,0]^T$  和  $[1,1]^T$  两个向量。根据

感知器算法, 其中  $\rho=1$ ,  $\omega(0)=[0.5,0.5]^T$ , 设计一个线性分离器来区分这两类

3.16 在上一章 2.12 问题中两分类问题中, 取  $\mu_1^T=[1,1]$ ,  $\mu_2^T=[0,0]$ ,  $\sigma_1^2=\sigma_2^2=0.2$ , 对于每一类产生 50 个向量。为了确保对于这两类的线性分离, 对于向量  $[1,1]$  类确保  $x_1+x_2 < 1$ ,

对于  $[0,0]$  向量类  $x_1+x_2 > 1$ 。下面的步骤就是使用这些向量去设计一个线性分类器使用 (3.21) 中的感知器算法。在收敛以后, 画出相关的判定线

3.17 假如 2.12 问题中是多类分类问题, 每一类有 100 个样本点。根据 LMS 算法使用这些数据去设计一个线性分类器。当所有的点被带入这个算法中进行计算的时候, 画出这个算法收敛的相关超平面。其中  $\rho_k = \rho = 0.01$ , 然后使用  $\rho = 0.01$ 。观察这个结果

3.18 证明, 使用 KESLER 构造器, 经过前面 3.21 感知器算法的有限步正确与错误分类计算后, 对于一个  $x_{(t)} \in \omega_i$ , 变为

$$\begin{aligned} \omega_i(t+1) &= \omega_i(t) + \rho x_{(t)} \quad \text{if } \omega_i^T x_{(t)} < \omega_j^T(t) x_{(t)}, \quad j \neq i \\ \omega_i(t+1) &= \omega_i(t) - \rho x_{(t)} \quad \text{if } \omega_i^T x_{(t)} < \omega_j^T(t) x_{(t)}, \quad j \neq i \\ \omega_k(t+1) &= \omega_k(t) \quad \forall k \neq j \quad \text{and } k \neq i \end{aligned}$$

3.19 证明理想权重向量的误差平方和趋近于 MSE 的解。

3.20 使用均方误差和的原则解决问题 3.6 并设计一个线性分类器。

3.21 证明设计一个 M 类的线性分类器, 有最佳误差平方和。分类器减少到 M 等价个有相应的效果。

3.22 证明, 假如  $x, y$  服从联合高斯分布, 对于  $x$  条件下  $y$  的分布是

$$\begin{aligned} E[y|x] &= \frac{\alpha \sigma_x x}{\sigma_x} + \mu_y - \frac{\alpha \sigma_x \mu_y}{\sigma_x}, \\ \Sigma &= \begin{bmatrix} \sigma_x^2 & \alpha \sigma_x \sigma_y \\ \alpha \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix} \end{aligned}$$

3.23 取 M 类分类器按照参数函数  $g(x; \omega_k)$  的形式存在, 目的是估计参数  $\omega_k$ , 使得分类器根据输入向量  $x$  能够产生期望的响应输出值。假设在每一类中  $x$  是随机分布, 分类器的输出根据相关期望响应值的不同而不同。按照高斯已知变量的一个高斯分布, 假设所有的输出都是相同的。证明按照误差平方和的原则, ML 估计是产生一个等价的估计值。



提示：在已知的类别当中取出  $N$  个训练样本值，对于他们中的每一个形成  $y_i = g(x; \omega_i) - d_i$ ， $d_i$  是第  $k$  类中

第  $i$  个样本点的期望响应值， $y_i$  服从正态 0 均值，方差为  $\sigma^2$  的分布。这个似然函数使用  $y_i$

3.24 在二类分类问题中，贝叶斯最佳判定截面是通过  $g(x) = P(\omega_1|x) - P(\omega_2|x) = 0$  给出，证明 MSE 中训练一个判定界面  $f(x; \omega)$ ，目的是对两类进行有效判别，相关的，它等价于在 MSE 最优感知中，它等价于  $f(x; \omega)$  的渐近函数形式  $g(\cdot)$ 。

3.25 假设在两类分类问题中有服从联合分布的特征向量，他们在有共同的方差  $\Sigma$ 。设计一个线性 MSE 分类器，证明在 2.11 问题中的贝叶斯分类器和这个结果的 MSE 分类器仅仅通过一个阈值就可以区分。简化起见，仅仅考虑等概率的类的情况。

提示：计算 MSE 超平面  $\omega^T x + \omega_0 = 0$ ，增加  $x$  的维数，它的解按照下列方式提供，

$$\begin{bmatrix} R & E[x] \\ E[x]^T & 1 \end{bmatrix} \begin{bmatrix} w \\ w_0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\mu_1 - \mu_2) \\ 0 \end{bmatrix}$$

相关的  $R$  和  $\Sigma$  在 MSE 分类器中按照下列的形式给出

$$(\mu_1 - \mu_2)^T \Sigma^{-1} \left( x - \frac{1}{2}(\mu_1 + \mu_2) \right) > (<) 0$$

#### 第四章 统计判决

4.1 使用最小最大损失判决规则的错分概率是最小吗？为什么？

4.2 当  $\Sigma_1 = \Sigma_2$  时，先验概率对决策超平面的位置影响如何？

4.3 假设在某个地区的细胞识别中正常  $\omega_1$  和异常  $\omega_2$  两类的先验概率分别为

正常状态， $P(\omega_1) = 0.9$

异常状态， $P(\omega_2) = 0.1$

现有一待识的细胞，其观测值为  $x$ ，从类条件概率密度分布曲线上查得  $p(x/\omega_1) = 0.2$ ， $p(x/\omega_2) = 0.4$

并且已知损失系数为  $\lambda_{11}=0$ ， $\lambda_{12}=1$ ， $\lambda_{21}=6$ ， $\lambda_{22}=0$ ，



试对该细胞以以下两种方法进行分类：①基于最小错误概率准则的贝叶斯判决；②基于最小损失准则的贝叶斯判决。请分析两种分类结果的异同及原因。

4.4 试用最大似然估计的方法估计单变量正态分布的均值  $\mu$  和方差  $\sigma^2$ 。

4.5 已知两个一维模式类别的类概率密度函数为

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

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

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。

4.6 在目标识别中，假定有农田和装甲车两种类型，类型  $\omega_1$  和类型  $\omega_2$  分别代表农田和装甲车，它们的先验概率分别为 0.8 和 0.2，损失函数如表 1 所示。现在做了三次试验，获得三个样本的类概率密度如下：

$$p(x/\omega_1): 0.3, 0.1, 0.6$$

$$p(x/\omega_2): 0.7, 0.8, 0.3$$

- (1) 试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型；
- (2) 假定只考虑前两种判决，试用贝叶斯最小风险准则判决三个样本各属于哪一个类型；
- (3) 把拒绝判决考虑在内，重新考核三次试验的结果。

表 1

类型	$\omega_1$	$\omega_2$
损失		
判决 $\alpha_1$	0.5	3
$\alpha_2$	5	1
$\alpha_3$	2	2



4.7 已知两个一维模式类别的类概率密度函数为

$$p(x|\omega_1) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{其它} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 2-2x, & 0 \leq x \leq 1 \\ 0, & \text{其它} \end{cases}$$

先验概率  $P(\omega_1)=P(\omega_2)$ , 损失函数,  $\lambda_{11}=\lambda_{22}=0$ ,  $\lambda_{12}=0.6$ ,  $\lambda_{21}=0.4$ .

(1) 求最小平均损失 Bayes 判决函数;

(2) 求总的误判概率  $P(e)$ ;

(3) 对于一个两类一维问题, 若这两类的类概率密度分别服从正态分布  $N(0, \sigma^2)$  和  $N(1, \sigma^2)$ , 证明使平均决策风险

最小的决策门限为 
$$x_0 = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{21}P(\omega_2)}{\lambda_{12}P(\omega_1)}$$

这里, 假设风险函数  $\lambda_{11}=\lambda_{22}=0$ . 一维正态分布: 
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

4.8 设 
$$\hat{C}(N) = \frac{1}{N} \sum_{j=1}^N (\bar{x}_j - \hat{\bar{m}}(N))(\bar{x}_j - \hat{\bar{m}}(N))^T$$
 是基于样本集  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$  对总体  $\bar{x} \sim N(\bar{m}, C)$  的协方差矩

阵的最大似然估计. 试推导由  $\hat{C}(N)$  求增加一个样本  $\bar{x}_{N+1}$  后协方差矩阵的估计  $\hat{C}(N+1)$  的递推公式. 其中,

$\hat{\bar{m}}(N)$  是基于样本集  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$  对总体  $\bar{x}$  的均值向量  $\bar{m}$  的最大似然估计 
$$\hat{\bar{m}}(N) = \frac{1}{N} \sum_{j=1}^N \bar{x}_j$$

4.9 设以下两类模式均为正态分布

$\omega_1: \{(0,0)^T, (2,0)^T, (2,2)^T, (0,2)^T\}$

$\omega_2: \{(4,4)^T, (6,4)^T, (6,6)^T, (4,6)^T\}$

(1) 设  $P(\omega_1)=P(\omega_2)=1/2$ , 求该两类模式之间的 Bayes 判别界面的方程.

(2) 绘出判别界面.

4.10 设以下两类模式均为正态分布

$\omega_1: \{(-5,-5)^T, (-5,-4)^T, (-4,-5)^T, (-6,-5)^T, (-5,-6)^T\}$

$\omega_2: \{(5,5)^T, (5,6)^T, (6,5)^T, (5,4)^T, (4,5)^T\}$

(1) 试用正交函数逼近法求类概率密度的估计  $\hat{p}(\bar{x}|\omega_1)$  和  $\hat{p}(\bar{x}|\omega_2)$ , 可选用 Hermite 正交多项式前四项低阶基函

数:  $H_0(x)=1$ ,  $H_1(x)=2x$ ,  $H_2(x)=4x^2-2$ ,  $H_3(x)=8x^3-12x$ ;



(2) 设  $P(\omega_1) = P(\omega_2) = 1/2$ , 求 Bayes 判决函数;

(3) 给出判别界面方程和图示。

4.11 证明在多元问题中, 贝叶斯决策准则使错误分类概率最小。

提示: 使用正确分类概率来证明要方便一些。

4.12 在一个两类一维问题中, 两类的概率分布密度函数分别为高斯分布  $N(0, \sigma^2)$  和  $N(1, \sigma^2)$ , 证明使平均风险最小的门限  $x_0$  为:

$$x_0 = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{21} P(\omega_2)}{\lambda_{12} P(\omega_1)} \quad \text{其中 } \lambda_{11} = \lambda_{22} = 0.$$

4.13 假设两类问题中损失矩阵为  $L = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$ ,  $\epsilon_1$  是将本来属于  $\omega_1$  类的样本错分为  $\omega_2$  的概率,  $\epsilon_2$  是将本来属于  $\omega_2$  类的样本错分为  $\omega_1$  的概率。试证明平均风险为

4.14 证明在多元分类问题中,  $M$  类的分类错误概率上限为  $P_e = (M-1)/M$ 。

提示: 对于每一个向量  $x$  最大后验概率密度函数  $P(\omega_i | \bar{x})$ ,  $i=1, 2, \dots, M$ , 大于或等于  $1/M$ 。这等价于每一个  $P(\omega_i | \bar{x})$  都是相等的。

4.15 假设在一维两类分类当中样本点符合 Rayleigh 概率密度函数分布:

$$p(x | \omega_i) = \begin{cases} \frac{x}{\sigma_i^2} \exp\left(-\frac{x^2}{2\sigma_i^2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{试求判决边界 } g(x) = 0.$$

4.16 在两类分类问题中, 限定其中一类的错分误概率为  $\epsilon_1 = \epsilon$ , 证明, 使另一类的错分概率  $\epsilon_2$  最小等价于似然比判

决: 如果  $P(\omega_1)/P(\omega_2) > \theta$ , 则判  $x \in \omega_1$ , 这里,  $\theta$  是使  $\epsilon_1 = \epsilon$  成立的似然比判决门限。

注: 这就是 Neyman-Pearson 判决准则, 它类似于贝叶斯最小风险准则。

提示: 该问题等价于用 Lagrange 乘子法, 使  $q = \theta(\epsilon_1 - \epsilon) + \epsilon_2$  最小化。

4.17. 二维三类问题, 假设每一类都服从同一正态分布, 且特征向量的协方差矩阵为

$$\Sigma = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.8 \end{bmatrix}$$

各类的均值向量分别是  $[0.1, 0.1]^T$ ,  $[2.1, 1.9]^T$ ,  $[-1.5, 2.0]^T$ 。



(1) 用贝叶斯最小错误概率分类器将向量  $[1.6, 1.5]^T$  分类。

(2) 画出距离向量  $[2.1, 1.9]^T$  的等马氏距离曲线图 (略图)。

4.18. 在两类三维空间分类问题中, 每一类中的特征向量都服从正态分布, 协方差矩阵为

$$\Sigma = \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.3 & -0.1 \\ 0.1 & -0.1 & 0.3 \end{bmatrix}$$

这两类的各自的均值向量分别为  $[0, 0, 0]^T$  和  $[0.5, 0.5, 0.5]^T$ 。试推导相应的线性决策函数和决策界面方程。

4.19. 在两类等概率分类问题中, 每一类中的特征向量的协方差矩阵均为  $\Sigma$ , 相关的均值向量为  $\mu_1, \mu_2$ , 证明对于贝叶斯最小错误概率分类器, 错误概率分布是

$$P_B = \int_{(1/2)d_m}^{+\infty} \frac{1}{2\pi} \exp(-z^2/2) dz$$

其中,  $d_m$  是这两个均值向量之间的马氏距离。该函数是  $d_m$  的增函数。

提示: 对数似然比  $u = \ln p(x | \omega_1) - \ln p(x | \omega_2)$  是一个随机变量, 且服从高斯分布:

$$N\left(\frac{1}{2}d_m^2, d_m^2\right), \forall x \in \omega_1; \text{ 和 } N\left(-\frac{1}{2}d_m^2, d_m^2\right), \forall x \in \omega_2. \text{ 据此计算错误概率。}$$

4.20. 证明假设每个向量都遵循高斯概率密度函数分布, 在 (2.19) 的最大似然概率检测

$$x \in \omega_1(\omega_2) \text{ if } l_{12} = \frac{p(x|\omega_1)}{p(x|\omega_2)} > (<) \theta$$

等价于

$$d_m^2(\mu_1, x | \Sigma_1) - d_m^2(\mu_2, x | \Sigma_2) + \ln \frac{|\Sigma_1|}{|\Sigma_2|} < (>) -2 \ln \theta$$

这里  $d_m^2(\mu_i, x | \Sigma_i)$  是  $\mu_i$  和  $x$  之间关于  $\Sigma_i$  矩阵的马氏距离。

4.21. 如果  $\Sigma_1 = \Sigma_2 = \Sigma$ , 证明上个问题成为  $(\mu_2 - \mu_1)^T \Sigma^{-1} x > (<) \theta$ , 这里



$$\Theta = \ln \theta + 1/2 (\|\mu_1\| \Sigma - \|\mu_2\| \Sigma)$$

4.22. 在二维两类问题中, 每一类  $\omega_1, \omega_2$  都服从以下分布:

$$p(\vec{x}|\omega_1) = \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{1}{2\sigma_1^2} (\vec{x} - \vec{\mu}_1)^T (\vec{x} - \vec{\mu}_1)\right)$$

$$p(\vec{x}|\omega_2) = \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{1}{2\sigma_2^2} (\vec{x} - \vec{\mu}_2)^T (\vec{x} - \vec{\mu}_2)\right)$$

其中  $\vec{\mu}_1 = (1, 1)^T$ ,  $\vec{\mu}_2 = (1.5, 1.5)^T$ ,  $\sigma_1^2 = \sigma_2^2 = 0.2$  假设  $P(\omega_1) = P(\omega_2)$ , 设计一个贝叶斯分类器, 满足

(a) 错误分类概率最小

$$\Lambda = \begin{bmatrix} 0 & 1 \\ 0.5 & 0 \end{bmatrix}$$

(b) 具有损失矩阵  $\Lambda$  的平均风险最小

使用一个伪随机的数值产生器, 从每一个类中得到 100 个特征向量。按照上面的概率密度函数。使用这个分类器去

分类已经产生的向量。对于每个事例中的错误概率是多少? 用  $\vec{\mu}_2 = (3.0, 3.0)^T$  重复这个实验。

4.23. 重复上面的实验, 特征向量服从以下分布:

$$p(\vec{x}|\omega_i) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\vec{x} - \vec{\mu}_i)^T \Sigma^{-1} (\vec{x} - \vec{\mu}_i)\right)$$

而且

$$\Sigma = \begin{bmatrix} 1.01 & 0.2 \\ 0.2 & 1.01 \end{bmatrix}$$

并且  $\mu_1 = [1, 1]^T$ ,  $\mu_2 = [1.5, 1.5]^T$

提示: 一个高斯随机向量的线性变换仍然是一个高斯随机向量。注意

$$\begin{bmatrix} 1.01 & 0.2 \\ 0.2 & 1.01 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}$$

4.24. 二维两类问题, 假设两类服从同一正态分布, 其协方差矩阵为  $\Sigma = \begin{pmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{pmatrix}$ , 均值向量分别为

$\vec{\mu}_1 = (0, 0)^T$ ,  $\vec{\mu}_2 = (3, 3)^T$ 。试用贝叶斯分类器对向量  $\vec{x} = (1.0, 2.2)^T$  进行分类。



4.25. 假设在两类一维问题中  $p(x|\omega_1)$  服从高斯分布  $N(\mu, \sigma^2)$ ,  $p(x|\omega_2)$  服从  $a$  到  $b$  之间的均匀分布. 证明贝

叶斯错误概率的上限为  $G\left(\frac{b-\mu}{\sigma}\right) - G\left(\frac{a-\mu}{\sigma}\right)$ , 其中  $G(x) \equiv P(y \leq x)$ , 并且  $y$  服从高斯  $N(0,1)$  分布.

4.26. 证明随机向量  $\frac{\partial \ln(p(x_i; \vec{\theta}))}{\partial \vec{\theta}}$  的均值是  $\vec{0}$ .

4.27. 在掷硬币的游戏实验中, 正面 (1) 出现的概率是  $q$ , 反面出现的概率是  $(1-q)$ . 设  $x_i, i=1, 2, \dots, N$  是这个

实验的结果,  $x_i \in \{0, 1\}$ , 证明  $q$  的最大似然估计是  $q_{ML} = \frac{1}{N} \sum_{i=1}^N x_i$  提示:

似然函数是:  $P(X; q) = \prod_{i=1}^N q^{x_i} (1-q)^{(1-x_i)}$

证明 ML 结果是下列方程的解  $q^{\sum x_i} (1-q)^{(N-\sum x_i)} \left( \frac{\sum x_i}{q} - \frac{N-\sum x_i}{1-q} \right) = 0$

4.28. 随机变量  $x$  服从高斯  $N(\mu, \sigma^2)$  分布,  $\mu$  未知. 给定该变量的  $N$  个观测值, 设  $L(\mu)$  为  $\mu$  的对数似然函数:

$L(\mu) = \ln(p(x; \mu))$ . 试求该随机变量的 Cramer-Rao 界:  $-E\left[\frac{\partial^2 L(\mu)}{\partial^2 \mu}\right]$ . 将该结果与  $\mu$  的 ML 估计值的方差进行比较, 有何结论? 假如这个未知参数是方差  $\sigma^2$ , 结论又如何? .

4.29. 证明假如似然函数是高斯函数有未知的均值  $\mu$ , 和协方差矩阵  $\Sigma$ , 然后 ML 估计如下给出

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

$$\hat{\Sigma} = \frac{1}{N} \sum_{k=1}^N \left( x_k - \hat{\mu} \right) \left( x_k - \hat{\mu} \right)^T$$

4.30. 随机变量  $x$  服从 Erlang 分布, 概率密度函数为

$$p(x; \theta) = \theta^2 x \exp(-\theta x) u(x)$$

其中  $u(x)$  是一个阶跃函数



$$u(x) = \begin{cases} 1 & , x > 0 \\ 0 & , x < 0 \end{cases}$$

假设  $x$  的  $N$  个观测值  $x_1, \dots, x_N$ , 证明  $\theta$  的最大似然估计为

$$\hat{\theta}_{ML} = \frac{2N}{\sum_{k=1}^N x_k}$$

4.31. 随机变量  $x$  是服从正态分布  $N(\mu, \sigma^2)$ , 其中未知参数  $\mu$  服从 Rayleigh 分布, 其概率密度函数为

$$p(\mu) = \frac{\mu \exp(-\mu^2/2\sigma_\mu^2)}{\sigma_\mu^2}$$

试证明  $\mu$  的最大后验概率估计为

$$\hat{\mu}_{MAP} = \frac{Z}{2R} (1 + \sqrt{1 - 4R/Z^2})$$

其中, 
$$Z = \frac{1}{\sigma^2} \sum_{k=1}^N x_k, \quad R = \frac{N}{\sigma^2} + \frac{1}{\sigma_\mu^2}$$

4.32. 证明对于对数正态分布

$$p(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \theta)^2}{2\sigma^2}\right), x > 0$$

最大似然估计为

$$\hat{\theta}_{ML} = \frac{1}{N} \sum_{k=1}^N \ln x_k$$

4.33. 若已知一个随机变量  $x$  的均值和方差:

$$\mu = \int_{-\infty}^{\infty} x p(x) dx,$$

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

试证明, 该随机变量概率密度函数的最大熵估计服从高斯分布  $N(\mu, \sigma^2)$

4.34.  $P$  为一个随机点  $x$  位于某区间  $h$  的概率。给定  $x$  的  $N$  个观测值, 其中有  $k$  个落入区间  $h$  的概率服从二项式分布:



$$\text{prob}\{k\} = \frac{N!}{k!(N-k)!} P^k (1-P)^{N-k}$$

证明  $E[k/N] = P$ ，并且它的方差为  $\sigma^2 = E[(k/N - P)^2] = P(1-P)/N$ 。

而且这个概率估计  $P=k/N$  是无偏的和渐进一致的。

## 第五章 特征提取与选择

5.1 设有  $M$  类模式  $\omega_i$ ,  $i=1, 2, \dots, M$ , 试证明总体散布矩阵  $S_t$  是总类内散布矩阵  $S_w$  与类间散布矩阵  $S_b$  之和, 即  $S_t = S_w + S_b$ 。

5.2 下面哪个矩阵可以用在二维空间线性变换中, 并保持马氏距离的特性? 请解释原因。

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}; \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}; \quad C = \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}$$

5.3 Bhattacharyya 可分性判据定义为

$$J_B = -\ln \int_{\Omega} [p(\bar{x} | \omega_1) p(\bar{x} | \omega_2)]^{1/2} d\bar{x}$$

式中  $\Omega$  表示特征空间。试证明, 在最小误判概率准则判决下, 最小最小误判概率有

$$P(e) \leq [P(\omega_1) P(\omega_2)]^{1/2} \exp(-J_B)。$$

5.4 令  $x_i$ ,  $i=1, 2, 3$  为独立的二值特征, 且  $p(x_i=1 | \omega_1) = \alpha_i$ ,  $p(x_i=1 | \omega_2) = \beta_i$ , 两类的先验概率相等, 且  $\alpha_i, \beta_i$  满足以下条件:

$$(1) \alpha_1 < \beta_1, \forall i, (2) \beta_1 - \alpha_1 > \beta_2 - \alpha_2 > \beta_3 - \alpha_3。$$

试证明各特征分别使用时之错误概率  $e(x_i)$  满足:  $e(x_1) < e(x_2) < e(x_3)$ 。

5.5 按上题条件, 试证明当两个特征合用时, 其错误概率为

$$e(x_i, x_j) = \frac{1}{2} [e(x_i) + e(x_j) - (\beta_i - \alpha_i) | e(x_j) - \alpha_j | - (\beta_j - \alpha_j) | e(x_i) - \alpha_i |]$$

请找出使  $e(x_1, x_2) < e(x_2, x_3)$  之条件。

5.6 同上题, 如果给定

$$\alpha_1 = 0.10, \alpha_2 = 0.05, \alpha_3 = 0.01,$$

$$\beta_1 = 0.90, \beta_2 = 0.80, \beta_3 = 0.70$$

试计算  $e(x_1), e(x_2), e(x_3), e(x_1, x_2), e(x_1, x_3), e(x_2, x_3)$ 。



5.7 已知以下两类模式

$$\omega_1: \{(0,0,0)^T, (1,0,0)^T, (1,0,1)^T, (1,1,0)^T\}$$

$$\omega_2: \{(0,0,1)^T, (0,1,0)^T, (0,1,1)^T, (1,1,1)^T\}$$

试用 K-L 变换分别把特征空间维数降到  $d=2$  和  $d=1$ , 并作图画出样本在该特征空间中的位置。

5.8 令  $\Sigma_i$  和  $P_i$  分别是  $\omega_i$  类 ( $i=1,2$ ) 的协方差矩阵和先验概率。假设对数据进行了白化变换, 即使  $B^T S_w B = I$ 。

这里,  $S_w = \sum_i P_i \Sigma_i$ ,  $I$  是单位矩阵。试证明矩阵  $P_1 B^T \Sigma_1 B$  和  $P_2 B^T \Sigma_2 B$  所产生的 K-L 坐标轴是相同的。若用  $\Lambda_i$  表示矩阵  $P_i B^T \Sigma_i B$  的本征值矩阵, 求证  $\Lambda_1 = I - \Lambda_2$ 。

8.1 三类  $\omega_1, \omega_2, \omega_3$ , 其中  $\omega_1: \{(1,0)^T, (2,0)^T, (1,1)^T\}$ ,  $\omega_2: \{(-1,0)^T, (-1,1)^T, (0,1)^T\}$ , 求  $S_w$  及  $S_b$ 。

8.3 令  $p(x|\omega_i) \sim N(\mu_i, \Sigma_i)$ ,  $i=1,2$ , 假定各特征分量  $x_j, j=1,2,\dots,D$ , 相互独立, 试证按式

$$J_D = I_{jj} + I_{jj} = \int_X [p(x|\omega_i) - p(x|\omega_j)] \ln \frac{p(x|\omega_i)}{p(x|\omega_j)} dx$$

定义的散度  $J_D$  可写为  $J_D = \sum_{j=1}^D J_{Dj}$

8.6 两个一维正态分布, 其期望与方差如下:

第一组  $\mu_1=0, \mu_2=2, \sigma_1^2=4, \sigma_2^2=0.25$ ; 第二组  $\mu_1=0, \mu_2=2, \sigma_1^2=1, \sigma_2^2=1$ 。

求 Bhattacharyya 距离及散度。

8.10 令  $x_i, i=1,2,3$  为独立的二值特征, 且  $p(x_i=1|\omega_1)=\alpha_i, p(x_i=1|\omega_2)=\beta_i$ , 二类先验概率相等, 且  $\alpha_i, \beta_i$  满足以下条件:

$$\textcircled{1} \alpha_i < \beta_i, \forall i, \textcircled{2} \beta_1 - \alpha_1 > \beta_2 - \alpha_2 > \beta_3 - \alpha_3.$$

试证各特征分别使用时之错误概率  $e(x_i)$  满足:  $e(x_1) < e(x_2) < e(x_3)$ 。

8.11 令  $x_i, i=1,2,3$  为独立的二值特征, 且  $p(x_i=1|\omega_1)=\alpha_i, p(x_i=1|\omega_2)=\beta_i$ , 二类先验概率相等, 且  $\alpha_i, \beta_i$  满足以下条件:

$$\textcircled{1} \alpha_i < \beta_i, \forall i, \textcircled{2} \beta_1 - \alpha_1 > \beta_2 - \alpha_2 > \beta_3 - \alpha_3.$$

试证当两个特征合用时其错误概率为

$$e(x_i, x_j) = \frac{1}{2} [e(x_i) + e(x_j) - (\beta_i - \alpha_i) | e(x_j) - \alpha_j | - (\beta_j - \alpha_j) | e(x_i) - \alpha_i |]$$



找出使  $e(x_1, x_2) < e(x_2, x_3)$  之条件。

8.12 令  $x_i, i=1, 2, 3$  为独立的二值特征, 且  $p(x_i=1|\omega_1)=\alpha_i, p(x_i=1|\omega_2)=\beta_i$ , 二类先验概率相等, 且  $\alpha_i, \beta_i$  满足以下条件:

$$\textcircled{1} \alpha_i < \beta_i, \forall i; \textcircled{2} \beta_1 - \alpha_1 > \beta_2 - \alpha_2 > \beta_3 - \alpha_3.$$

当两个特征合用时其错误概率为  $e(x_i, x_j) = \frac{1}{2}[e(x_i) + e(x_j) - (\beta_i - \alpha_i)|e(x_j) - \alpha_j| - (\beta_j - \alpha_j)|e(x_i) - \alpha_i|]$ , 如果

给 定  $\alpha_1 = 0.10, \alpha_2 = 0.05, \alpha_3 = 0.01, \beta_1 = 0.90, \beta_2 = 0.80, \beta_3 = 0.70$  . 试 计 算

$e(x_1), e(x_2), e(x_3); e(x_1, x_2), e(x_1, x_3), e(x_2, x_3)$  .

## 第六章 句法模式识别

6.1 句法模式识别中, 模式类是如何描述的。

6.2 文法  $G=(V_n, V_t, P, S)$  中起始符  $S$  可以是非终止符吗?

6.3 试给出能产生不含子串 011 的 0, 1 字符串的全体的正则文法  $G_n$ , 画出  $G_n$  对应的有限状态自动机的状态转移图。

设  $x=10010$ , 试根据 CYK 算法导出对  $x$  的最左剖析。

6.4 已知文法  $G=(V_n, V_t, P, S)$ ,  $V_n=\{S, \Delta\}$ ,  $V_t=\{a, d, (, ), +, \times, -, *, \sim\}$ , 其中,  $a, d$  为图形描述语言 (PDL) 的基元:  $a: \rightarrow$ ;  $d: \downarrow$ ; “+”, “ $\times$ ”, “-”, “\*”, “ $\sim$ ” 为基元链接运算: “+” 为头接尾; “ $\times$ ” 为尾尾相接; “-” 为头头相接; “\*” 为头接和尾接; “ $\sim$ ” 为反向; “(, )” 为运算顺序控制符。试写出描述字母 “E” 的 PDL 句子及相应的产生式  $P$ 。

## 第三部分 综合分析题

1、在描述一个完整的模式识别系统的基础上, 给出你所了解的一种模式识别应用实例, 并对其中可能用到的模式识别典型理论和方法进行简要说明。

2、举出日常生活或技术、学术领域中应用模式识别理论解决问题的实例 (包括问题模型, 解决的方法, 体会)

3、若要实现手机上的语音拨号功能, 你认为应该有哪些处理步骤? 分别需要哪些模式识别方法? 试用流程图予以说明。

4、若要实现机器 (计算机、手机等) 的手写字符输入功能, 你认为应该有哪些处理步骤? 分别需要哪些模式识别方法? 试用流程图予以说明。



5、若要实现计算机指纹识别功能，你认为应该有哪些处理步骤？分别需要哪些模式识别方法？试用流程图予以说明。

6、若要实现汽车车牌自动识别，你认为应该有哪些处理步骤？分别需要哪些模式识别方法？试用流程图予以说明。