《模式识别》试题答案（A 卷）

（2007 年秋季学期，学历教育合训本科生，理论考核部分，120 分钟）

一、填空与选择填空（本题答案写在此试卷上，30 分）

1、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）**）。

2、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

3、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。**）。

4、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。

5、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（ $K(\vec{x})=\sum_{\vec{x}_k\in\tilde{X}}\alpha_k K(\vec{x},\vec{x}_k)$ ）。

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；最小最大判决准则主要用于（ **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 Jij 越大，说明 ωi 类模式与 ωj 类模式的分布（**差别越大**）；当 ωi 类模式与 ωj 类模式的分布相同时，Jij=（**0**）。

9、 已知有限状态自动机 Af=(∑，Q，δ，q0，F)，∑={0，1}；Q={q0，q1}；δ：δ(q0，0)= q1，δ(q0，1)= q1，δ(q1，0)=q0，δ(q1，1)=q0；q0=q0；F={q0}。现有输入字符串：(a) 00011101011，(b) 1100110011，(c) 101100111000，(d)0010011，试问，用 Af 对上述字符串进行分类的结果为（ **ω1:{a,d};ω2:{b,c}** ）。

二、（15 分）在目标识别中，假定类型 ω1 为敌方目标，类型 ω2 为诱饵（假目标），已知先验概率 P(ω1)=0.2 和 P(ω2)=0.8，类概率密度函数如下：

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

0　　其它

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

（1）求贝叶斯最小误判概率准则下的判决域，并判断样本 x=1.5 属于哪一类；

（2）求总错误概率 P(e)；

（3）假设正确判断的损失 $\lambda 11=\lambda 22=0$，误判损失分别为 $\lambda 12$ 和 $\lambda 21$，若采用最小损失判决准则，$\lambda 12$ 和 $\lambda 21$ 满足怎样的关系时，会使上述对 x=1.5 的判断相反？

解：**（1）应用贝叶斯最小误判概率准则如果　则判**

得 $l_{12}(1.5)=1 <=4$，故 x=1.5 属于 ω2 。

（2）P(e)=

==0.08

(3) 两类问题的最小损失准则的似然比形式的判决规则为：

如果

则判

带入 x=1.5 得到　$\lambda 12 \geqslant 4\lambda 21$

三、（10 分）二维两类问题，已知第一类ω1={三角形 ABC}，三角形 ABC 的顶点坐标分别为 {(1,3),(2,1),(3,2)}；其它区域为第二类ω2。试设计一个能对其正确分类的神经网络。

解：

三角形 ABC 三条边的方程：

(y-3)/(x-1)=(y-1)/(x-2) => d1(x,y)=2x+y-5=0

(y-1)/(x-2)=(y-2)/(x-3) => d2(x,y)=-x+y+1=0

(y-3)/(x-1)=(y-2)/(x-3) => d3(x,y)=-x-2y+7=0

故ω1={(x,y)|(2x+y-5>0)∩(-x+y+1>0) ∩(-x-2y+7>0)}

可取有三个神经元的单隐含层网络，隐含层到输出神经元权值为 1，输出神经元阀值取为 2.5 即可。

四、（15 分）（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最小。

其中，k 是迭代次数；是的样本均值。

解：**（1）框图中给出以下基本步骤：**

1、任选个模式特征矢量作为初始聚类中心。

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

3、 计算重新分类后的各类心。

4、如果任一类的类心改变，则转至(2)；否则结束。

（2）设某样本从聚类移至聚类中，移出后的集合记为，移入后的集合记为。设和所含样本数分别为和，聚类、、和的均矢分别为、、和，显然有

（1） （2）

而这两个新的聚类的类内欧氏距离(平方)和与原来的两个聚类的类内欧氏距离(平方)和的关系是

（3）

（4）

当距比距更近时，使得

（5）

由式(3)、(4 及(5)可知，将分划给类可使 J 变小。这说明在分类问题中不断地计算新分划的各类的类心，并按最小距离原则归类可使 J 值减至极小值。

五、（30 分）上机实验及作业（时间另计）。

略