# 中国科学院自动化研究所
# 2007年招收攻读博士学位研究生入学统一考试试题
# 科目名称：模式识别

**考生须知：**

1. 本试卷满分为100分，全部考试时间总计180分钟。
2. 所有答案必须写在答题纸上，写在试题纸上或草稿纸上一律无效。

---

1. （10分）对于$M$类（$\omega_1,\ldots,\omega_M$）分类问题，给定每一类的先验概率$P(\omega_i)$和条件概率密度$p(\mathbf{x}\mid\omega_i)$。请用公式叙述最小错误率（贝叶斯）决策过程。

2. （10分）对于两类问题，假定$p(\mathbf{x}\mid\omega_i)\sim N(\mu_i,\Sigma_i)$，$i$=1,2，$\mathbf{x}$为$d$维特征矢量。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) $\Sigma_1=\Sigma_2=\sigma^2\mathrm{I}$（I为单位矩阵）。

3. （10分）设$\mathbf{x}$为$d$维二值（0或1）矢量，服从二项（Bernoulli）分布：

$$p(\mathbf{x}\mid\boldsymbol{\theta})=\prod_{i=1}^{d}\theta_i^{x_i}(1-\theta_i)^{(1-x_i)}$$

$\boldsymbol{\theta}=(\theta_1,\ldots,\theta_d)^T$是未知参数矢量。给定$n$个设计样本$\mathbf{x}_i, i=1,\ldots n$，求$\boldsymbol{\theta}$的最大似然估计。

4. （15分）给定$n$个$d$维设计样本$\mathbf{x}_1,\ldots,\mathbf{x}_n$，采用窗函数

$$\varphi(\mathbf{u})=\begin{cases}1 & |u_j|\leq 1/2,\ j=1,\ldots,d\\ 0 & \text{otherwise}\end{cases}$$

和窗宽$h_n$，（1）写出概率密度函数的Parzen窗估计$p_n(\mathbf{x})$；（2）如果设计样本来自$M$个不同的类别，新样本$\mathbf{x}$可根据k-近邻规则进行分类。请从Parzen窗估计推导出k-近邻分类规则。

5. （15 分）设有来自二类的$n$个样本$\mathbf{x}_1,...,\mathbf{x}_n$，类别标号为$b_i \in \{+1,-1\}, i=1,...,n$。采用两种判别函数（1）$f(\mathbf{x})=\mathbf{w}^T\mathbf{x}+w_0$和（2）$f(\mathbf{x})=s(\mathbf{w}^T\mathbf{x}+w_0)=\dfrac{1}{1+\mathrm{e}^{-(\mathbf{w}^T\mathbf{x}+w_0)}}$，用梯度下降法学习参数$\mathbf{w}$和$w_0$，使平方误差准则$\sum_{i=1}^{n}[f(\mathbf{x}_i)-b_i]^2$最小化。分别写出两种判别函数下的学习过程。

6. （20 分）有三类先验概率相等的二维模式服从正态分布$p(\mathbf{x}\mid\omega_i)\sim N(\mu_i,\Sigma_i)$，其中$\mu_1=(0,0)^T$，$\mu_2=(2,1)^T$，$\mu_3=(1,-1)^T$，$\Sigma_1=\Sigma_2=\Sigma_3=\begin{bmatrix}1 & 0\\ 0 & 1/2\end{bmatrix}$。分别求使$\mathbf{w}^T\mathrm{S}_\mathrm{t}\mathbf{w}$和$\mathbf{w}^T\mathrm{S}_\mathrm{w}^{-1}\mathrm{S}_\mathrm{t}\mathbf{w}$（$\mathrm{S}_\mathrm{w}$和$\mathrm{S}_\mathrm{t}$分别为类内离散度和总离散度）最大化的特征矢量$\mathbf{w}$。

7. （10 分）给定一个特征集合$\mathbf{x}=\{x_1,x_2,...,x_D\}$和评价准则（判据）$J(\mathbf{y})$（$\mathbf{y}=\{x_{i_1},...,x_{i_d}\}$是一个特征子集）。描述顺序前向搜索（Sequential forward search）和顺序后向搜索（Sequential backward search）特征选择方法的过程并分析它们各适用于什么场合。

8. （10 分）对于两个样本矢量集$\mathrm{X}_1=\{\mathbf{x}_1^1,...,\mathbf{x}_{n_1}^1\}$和$\mathrm{X}_2=\{\mathbf{x}_1^2,...,\mathbf{x}_{n_2}^2\}$，定义两种距离度量：（1）$d_1(\mathrm{X}_1,\mathrm{X}_2)=d(\mathbf{m}_1,\mathbf{m}_2)$（$d(,)$为欧式距离，$\mathbf{m}_1$和$\mathbf{m}_2$分别为$\mathrm{X}_1$和$\mathrm{X}_2$的均值），（2）$d_2(\mathrm{X}_1,\mathrm{X}_2)=\min_{i,j} d(\mathbf{x}_i^1,\mathbf{x}_j^2)$。在以上两种距离度量下分别用分级聚类法(Hierarchical clustering)将以下 10 个矢量聚成 3 类：(-1,0), (0,0), (0,1), (-0.5,0.5), (0,-1), (1,-1), (1,0), (1,1), (2,0), (3,0).