

中国科学院自动化研究所
2010 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均无效。
-

1. （16 分）说出四种分类方法：参数统计方法、非参数统计方法、前馈神经网络、支持向量机各有什么优缺点。

2. （14 分）对于 M 类 $(\omega_1, \dots, \omega_M)$ 分类问题，给定每一类的先验概率 $P(\omega_i)$ 和条件概率密度 $p(\mathbf{x} | \omega_i)$ 。请用公式叙述最小错误率（贝叶斯）决策过程并给出相应的最小分类错误率。

3. （18 分）线性判别函数 $f(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$ （ \mathbf{w}_i 和 w_{i0} 为第 i 类的权值矢量和偏移值）是一种常用的分类度量。给定来自 M 类的设计样本 $\mathbf{x}_1, \dots, \mathbf{x}_N$ ，请给出三种估计线性判别函数参数（ \mathbf{w}_i 和 w_{i0} ）的方法，分别写出它们的公式。

4. （15 分）假设一元随机变量的分布 $p(x) \sim N(\mu, \sigma^2)$ ，给定 N 个设计样本 $x_i, i=1, \dots, N$ ，请推导用最大似然法估计均值和方差的计算公式。

5. （12 分）两类概率密度之间的散度 $d_{ij} = \int_{-\infty}^{+\infty} [p(\mathbf{x} | \omega_i) - p(\mathbf{x} | \omega_j)] \ln \frac{p(\mathbf{x} | \omega_i)}{p(\mathbf{x} | \omega_j)} d\mathbf{x}$,

如果有两类模式均服从高斯分布且先验概率和协方差矩阵相同，证明散度等价于 Fisher 准则 $\text{trace}[S_w^{-1} S_b]$

（未完待续）

6. (15 分) 将 n 维特征矢量 \mathbf{x} 近似表示为($m < n$)

$$\hat{\mathbf{x}} = \sum_{i=1}^m y_i \mathbf{e}_i + \sum_{i=m+1}^n c_i \mathbf{e}_i$$

其中 $c_i (i=m+1, \dots, n)$ 为常数, $\mathbf{e}_i (i=1, \dots, n)$ 为正交归一基矢量。证明要使重建均方误差 $E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ 最小化, 需满足以下三个条件: (1) $c_i = E[y_i]$, $i=m+1, \dots, n$; (2) 正交归一基矢量是协方差矩阵 Σ_x 的特征向量; (3) $\mathbf{e}_i, i=m+1, \dots, n$, 是对应最小特征值的特征向量。

7. (10 分) 给定 5 个样本之间相互的距离矩阵

$$P = \begin{bmatrix} 0 & 4 & 9 & 6 & 5 \\ 4 & 0 & 3 & 8 & 7 \\ 9 & 3 & 0 & 3 & 2 \\ 6 & 8 & 3 & 0 & 1 \\ 5 & 7 & 2 & 1 & 0 \end{bmatrix}$$

分别用 Single-link (最近距离) Complete-link (最远距离) 方法对样本进行分级聚类, 画出聚类结果的系统树图(Dendrogram)。