中国科学院自动化研究所

2016年招收攻读博士学位研究生入学统一考试试卷

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). 请写出如下十个在模式识别领域中的英文简写的全称（比如，PR的全称为Pattern Recognition）：SVM、KPCA、MLP、K-NN、LDA、SOM、LLE、ANN、CNN、RBF。

2. (10 分). 有四个二维空间中的样本，它们分别属于两个不同的类别，其中第一类的两个样本为$(1,4)^T$和$(2,6)^T$，第二类的两个样本为$(1,2)^T$和$(2,2)^T$。这里，上标T表示向量转置。若采用规范化增广样本表示形式，并假设初始的权向量$\mathbf{a}=(0,1,0)^T$，其中向量$\mathbf{a}$的第三维对应于样本的齐次坐标。同时，假定梯度更新步长ρ_t固定为1。试利用感知器算法求解线性判别函数$g(\mathbf{y})=\mathbf{a}^T\mathbf{y}$的权向量$\mathbf{a}$。（注：“规范化增广样本表示”是指对齐次坐标表示的样本进行规范化处理）。

3. (15 分). 现有五个四维空间中的样本：$\mathbf{x}_1=(1,3,2,1)^T$、$\mathbf{x}_2=(2,3,1,2)^T$、$\mathbf{x}_3=(2,2,1,2)^T$、$\mathbf{x}_4=(5,5,1,1)^T$和$\mathbf{x}_5=(5,3,2,1)^T$。这里，上标T表示向量转置。请按最小距离准则对上述五个样本进行分级聚类，并画出聚类系统树图。

4. (15 分). 设有n个d维空间的训练样本$\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$，它们分别属于两个不同的类别，其类别标签分别为$y_1,y_2,\ldots,y_n\in\{+1,-1\}$。现有如下线性支持向量机学习模型：

$$
\begin{aligned}
\max_{\alpha_1,\ldots,\alpha_n} \quad & \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j}^{n}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \\
s.t. \quad & \sum_{i=1}^{n} y_i\alpha_i=0, \\
& 0\le\alpha_i\le C,\quad i=1,2,\ldots,n
\end{aligned}
$$

（1）请将上述模型扩展为核支持向量机学习模型 (6 分)；（2）写出核支持向量机分类决策函数 (4 分)；（3）给出两种不同的核函数，并描述一种基于训练数据的核函数参数选择方法 (5 分)。

（未完待续）

5. (12 分). （1）设有一维窗函数 $\varphi(u)=\exp(-|u|)$，并假定有 n 个一维空间中的样本 $x_1, x_2, ..., x_n$。现采用宽度为 h_n 的窗函数，请写出概率密度函数 $p(x)$的 Parzen 窗估计 $p_n(x)$（8 分）；（2）给定一维空间三个样本点{-2，0，2}，请写出概率密度函数 $p(x)$的最近邻（1-NN）估计并画出概率密度函数曲线图（4 分）。

6. (14 分). （1）给出最小误差平方和聚类准则的目标函数（6 分）；（2）给出 C 均值聚类算法的计算步骤，并指出至少两个能够影响 C 均值聚类结果的因素（8 分）。

7. (12 分). （1）设有 n 个 d 维空间的样本 $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$，它们分别来自于 c 个不同的类别。假定要设计一个多层前向神经网络，将该网络训练之后可用于对新样本进行分类。请描述你所设计的网络结构（7 分）；（2）假定一个前向神经网络一共有 150 个隐含层，现在拟采用反向传播算法对其进行训练。在训练过程中，越靠近输出层的连接权重得到更新的机会可能会越来越小或者可能会越来越大（越来越小是指连接权重的梯度更新量均接近于零，越来越大是指连接权重的梯度更新量均远远大于 1）。请问在网络训练过程中为什么会出现上述两种情形，给出防止出现此类问题的具体方法（5 分）。

8. (6 分). 关于主成分分析。设有 n 个 d 维空间的样本 $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$，且样本协方差矩阵为 $\mathbf{\Sigma}$。试证明第一主成分为矩阵 $\mathbf{\Sigma}$ 的最大的特征值对应的特征向量。

9. (6 分). 设有 n 个 d 维空间的样本 $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$，但并不知道它们的具体取值。假定这些样本的均值向量 $\mathbf{x}_c$ 已知，且所有两两样本之间的欧氏距离也已知，试采用古典尺度法求出这些样本的具体取值（注：不考虑正交变换对样本的影响）。