

中国科学院自动化研究所
2015 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. (5 分) 关于 ROC 曲线。(1) “ROC” 代表三个英文单词的首字母，请顺序写出这三个单词的全称 (3 分)；(2) 针对两类分类问题，请描述 “ROC” 曲线的绘制步骤 (2 分)。
2. (15 分) 关于线性分类器。(1) 请给出 Fisher 线性判别分析的主要计算步骤和分类决策规则 (10 分)；(2) 请给出线性不可分情形下支持向量机学习模型 (5 分)。
3. (8 分) 关于 Boosting 方法。假定有 n 个 m 维空间中的训练样本 $\{x_1, x_2, \dots, x_n\} \subset R^m$ ，进一步假定这些样本属于两个不同的类别。(1) 基于这些样本，请简述采用 Boosting 方法学习一个两类分类器的步骤，并指出其中拟采用的弱分类器 (4 分)；(2) 基于两类 Boosting 方法，请设计一个针对多类分类问题的分类器学习框架 (4 分)。
4. (8 分) 关于聚类。(1) 请给出 ISODATA 方法的主要计算步骤 (4 分)；(2) 请指出 ISODATA 方法和 C 均值算法的不同点 (4 分)。
5. (8 分) 关于主成分分析。(1) 请简述主成分分析方法的原理 (4 分)；(2) 采用核技巧，可以将主成分分析方法发展为核主成分分析方法，请结合核主成分分析方法解释核技巧的工作原理 (4 分)。
6. (14 分) 关于神经网络。(1) 针对多层前馈神经网络，请给出反向传播算法的工作原理和训练步骤 (10 分)；(2) 请分析 “在前馈神经网络中，隐含层数越多对分类预测可能产生的影响” (4 分)。

(未完待续)

7. (10 分) 关于特征选择。给定 n 个 m 维空间中的训练样本 $\{x_1, x_2, \dots, x_n\} \subset R^m$ ，且每个样本的类别标签是已知的。请用迹比值判据 $J_4 = \frac{tr(S_b)}{tr(S_w)}$ 设计一个完整的特征选择算法，其中 $tr(\cdot)$ 表示矩阵求迹运算， S_b 表示类间散度矩阵， S_w 表示总类内散度矩阵。
8. (10 分) 现有一个二维空间中的两类分类问题。记两个类分别为 ω_1 和 ω_2 。假定 ω_1 和 ω_2 出现的先验概率均相等，且 ω_1 和 ω_2 的类条件概率密度均为正态分布。其中， ω_1 的类条件概率密度的均值为 $\mu_1 = [-1, 0]^T$ ，协方差矩阵为 $\Sigma_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ ； ω_2 的类条件概率密度的均值为 $\mu_2 = [1, 0]^T$ ，协方差矩阵为 $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ ，其中上标 T 表示向量转置。请写出负对数似然比决策规则。
9. (14 分) 现有 9 个训练样本，分别属于三个不同的类别。第一类的样本点集为 $\{[1, 0]^T, [2, 0]^T, [1, 1]^T\}$ ，第二类的样本点集为 $\{[0, 1]^T, [-1, 0]^T, [-1, 1]^T\}$ ，第三类的样本点集为 $\{[0, -1]^T, [-1, -1]^T, [0, -2]^T\}$ ，其中上标 T 表示向量转置。假定类先验概率均相等，请计算类间散度矩阵和总类内散度矩阵。
10. (8 分) 关于非线性特征提取。(1) 请描述 IsoMap (isometric feature mapping) 方法的主要计算步骤 (3 分)；(2) 现有一个房间，其四周墙壁均为白色，且室内光照明亮均匀。在该房间的一张白色的桌子上平放着一把彩色茶壶，茶壶所有表面均绘有不同的花纹。另外，室内有一个照相机可沿一个圆移动，该圆所在的平面与桌面平行。茶壶的质心与该圆的圆心重叠。在照相机运动的过程中，照相机的焦距等成像参数保持不变，但其拍摄方向始终对准茶壶，且能将茶壶定位于图像中央，并使整个茶壶大致占满整个图像。假定照相机沿该圆匀速运动一周一共连续拍摄到 400 张不同的图像。现将 IsoMap 方法应用于这些图像数据，并假定最终输出一个二维特征提取结果。如果将这 400 个二维特征点在二维笛卡尔坐标系统中进行绘制，请分析可能呈现出的形状，并给出理由 (5 分)。