

中国科学院自动化研究所

2017 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. (12 分) 对一个  $c$  类分类问题，设各类先验概率为  $P(\omega_i)$ ,  $i=1,2,\dots,c$ , 类条件概率密度为  $P(\mathbf{x}|\omega_i)$ ,  $i=1,2,\dots,c$ 。其中,  $\mathbf{x}$  表示特征向量。设将第  $j$  类模式判定为第  $i$  类的损失为  $\lambda_{ij}$ 。请写出最小错误率贝叶斯决策和最小风险贝叶斯决策的决策规则。

2. (12 分) 关于概率密度函数估计：

- (1) 设一维特征空间中的窗函数为  $\varphi(u) = \begin{cases} 1, & |u| < 1/2 \\ 0, & \text{otherwise} \end{cases}$ 。现有  $n$  个样本  $x_i$ ,  $i=1,\dots,n$ ,

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  $p_n(x)$ ; (6 分)

- (2) 给定一维空间中的三个样本点  $\{-1,0,2\}$ ，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。(6 分)

3. (10 分) 设有  $n$  个  $d$  维空间的训练样本  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ，它们分别属于两个类别，其类别标签分别为  $y_1, y_2, \dots, y_n \in \{+1, -1\}$ 。写出线性不可分情形下的支持向量机学习模型。

4. (10 分) 现有四个来自于两个类别的二维空间中的样本，其中第一类的两个样本为  $(3,2)^T$  和  $(2,2)^T$ ，第二类的两个样本为  $(1,0)^T$  和  $(2,0)^T$ 。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  $\mathbf{a}=(1,0,0)^T$ ，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  $\rho_t$  固定为 1。试利用批处理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 1：“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；注 2：批处理感知器算法是指在权重向量迭代修正的过程中考虑所有错分样本)。

5. (14 分) 关于特征提取：

- (1) 请简述主成分分析方法的原理；(6 分)

- (2) 请简述 Fisher 线性判别分析(Linear Discriminant Analysis, LDA)的主要思想，并给出两类分类问题的 LDA 学习模型（即最优化模型）。(8 分)

(未完待续)

6. (20 分) 在一个模式识别问题中, 有下列 8 个样本, 每个样本为一个二维特征向量:  $(-4, 1)$ ,  $(-2, 1)$ ,  $(-4, -1)$ ,  $(-2, -1)$ ,  $(4, 1)$ ,  $(2, 1)$ ,  $(4, -1)$ ,  $(2, -1)$ 。其中, 括号内的第一个数据表示该样本点的第一个特征, 括号内的第二个数据表示该样本点的第二个特征。
- (1) 现在考虑对这 8 个样本进行聚类。采用 C 均值聚类算法, 并假定聚类中心的个数为 2, 两个初始聚类点分别为  $(-5, 0)$  和  $(5, 0)$ 。根据上述初始条件, 请写出 C 均值聚类算法的计算步骤, 并给出最后的聚类中心; (10 分)
- (2) 对上述 8 个样本, 假设前 4 个样本属于第一类, 后 4 个样本属于第二类, 现在要求只使用一个特征达到这样的分类目的。请给出一种特征选择方法, 并给出计算过程和结果。(10 分)
7. (12 分) 设有  $n$  个  $d$  维空间的训练样本, 这些样本一共来自于  $c$  个类别。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, 给出训练该网络的主要计算步骤; 从网络结构和网络训练的角度指出哪些因素会对分类性能造成影响。
8. (10 分) 请从监督学习(学习过程中利用样本的类别标签)和无监督学习(学习过程中不利用样本的类别标签)的角度, 对如下方法进行归类: Parzen 窗概率密度估计方法、K-近邻分类器、感知器准则方法、主成分分析方法、决策树方法、支持向量机方法、C 均值聚类方法、ISODATA 方法、罗杰斯特回归方法、分级聚类方法、核主成分分析方法、LLE 方法、Adaboost 方法、自组织映射方法、Isomap 方法、Fisher 判别分析方法、谱聚类方法、Recurrent Neural Network 方法、受限玻尔兹曼机方法、卷积神经网络方法。