

中国科学院自动化研究所
2014 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. (16 分) 关于统计学习与支持向量机，请回答如下问题：(1) 给出机器学习问题的形式化表示 (4 分)；(2) 解释学习机器的推广能力 (4 分)；(3) 从几何的角度阐述线性支持向量机的原理 (4 分)；(4) 基于两类支持向量机，设计一个 c 类($c > 2$)分类训练策略 (4 分)。
 2. (10 分) (1) 请描述径向基函数网络的结构和功能 (4 分)；(2) 指出径向基函数网络的参数，分析在训练一个径向基函数网络时如何调节这些参数 (6 分)。
 3. (10 分) (1) 简述 Fisher 线性判别分析的原理 (4 分)；(2) 针对两类分类问题，试证明在正态等方差条件下，Fisher 线性判别等价于贝叶斯判别 (6 分)。
 4. (10 分) 假设在某个局部地区细胞识别中正常 (ω_1)和异常(ω_2)两类的先验分别为 $P(\omega_1)=0.85$ 和 $P(\omega_2)=0.15$ 。现有一待识别细胞，其观察值为 x ，从类条件概率密度分布曲线上查得 $P(x|\omega_1)=0.2$ ， $P(x|\omega_2)=0.4$ ，请对该细胞 x 进行分类，并给出计算过程。
 5. (10 分) 现有七个位于二维空间的样本： $\mathbf{x}_1=(1,0)^T$ ， $\mathbf{x}_2=(0,1)^T$ ， $\mathbf{x}_3=(0,-1)^T$ ， $\mathbf{x}_4=(0,0)^T$ ， $\mathbf{x}_5=(0,2)^T$ ， $\mathbf{x}_6=(0,-2)^T$ ， $\mathbf{x}_7=(-2,0)^T$ ，其中上标 T 表示向量的转置。假定前三个样本属于第一类，后四个样本属于第二类，请画出最近邻法决策面。
 6. (16 分) 在一个模式识别问题中，有下列 8 个样本： $\mathbf{x}_1=(-1,1)^T$ ， $\mathbf{x}_2=(-1,-1)^T$ ， $\mathbf{x}_3=(0,1)^T$ ， $\mathbf{x}_4=(0,-1)^T$ ， $\mathbf{x}_5=(2,1)^T$ ， $\mathbf{x}_6=(2,-1)^T$ ， $\mathbf{x}_7=(3,1)^T$ ， $\mathbf{x}_8=(3,-1)^T$ ，其中上标 T 表示向量的转置。请回答如下问题：(1) 如果不知道这 8 个样本的类别标签，请采用 K-L 变换，计算其特征值和特征向量 (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，请给出一种特征选择方法，并写出相应的计算过程 (6 分)。
 7. (16 分) (1) 给定 m 维空间中的 n 个样本，请给出 C -均值聚类算法的计算步骤(包含算法输入和输出)(8 分)；(2) 针对 C -均值聚类算法，指出影响聚类结果的因素，并给出相应的改进措施 (8 分)。
 8. (12 分) 某单位有 n 位职员，现从每位职员采集到 m ($m > 10$)张正面人脸图像(可能因姿态、表情、光照条件的略微不同而不同)。每张人脸图像为 200(高度) \times 160(宽度)像素大小的灰度图像。现在拟设计一个人脸识别系统，请回答如下问题：(1) 描述拟采用的特征提取方法及计算步骤 (4 分)；(2) 描述拟采用的分类器构造方法及计算步骤 (4 分)；(3) 请从特征提取和分类器构造两方面对你所采用的方法进行评价(即解释采用它们的原因) (4 分)。