**四川大学期末考试试题（闭卷）**

**（2021——2022学年第 2 学期） A卷**

课程号：304072020 课序号： 课程名称：模式识别引论 任课教师：卢晓春、赵启军 成绩：

适用专业年级：计算机学院2019级 学生人数：81 印题份数： 学号： 姓名：

|  |
| --- |
| **考 生 承 诺**  我已认真阅读并知晓《四川大学考场规则》和《四川大学本科学生考试违纪作弊处分规定（修订）》，郑重承诺：  1、已按要求将考试禁止携带的文具用品或与考试有关的物品放置在指定地点；  2、不带手机进入考场；  3、考试期间遵守以上两项规定，若有违规行为，同意按照有关条款接受处理。  **考生签名：** |
| **一、填空题（本大题共2小题，每空2分，共6分）**  1. 特征是 从模式中测量得到的值，也是用于分类的依据 。  2. 监督学习中的训练样本类别事先未给定，可通过数据聚类根据样本之间的相似度来划分，这个说法 错误 。（正确、错误）  3. 线性分类器的分界面方程是  。  **二、问答题（本大题共5小题，每题10分，共50分）**  1. 特征个数越多越有利于分类，这个说法是否正确？  答：错误。特征选择的主要目的就是从特征当中选出最有利于分类的特征，以降低特征维数。  2. 给出C均值算法的步骤描述  答：   * 1. 给定允许误差ℇ，令t=1   2. 初始化聚类中心mi(t)，i=1,2,…,c   3. 修正dij，   4. 修正聚类中心mi(t+1)   5. 计算误差E或者Je   6. 如果E< ℇ ，则算法结束；否则t=t+1，转步骤3   3. 以两分类问题为例，描述感知器算法？  答：其特点是随意确定的判别函数初始值，在对样本分类训练过程中逐步修正直至最终确定。   * 两分类问题，有其增广模式向量规范化后的样本集为y={y1,y2,…,yn}。为了确定加权向量a\*，执行下面的训练算法   + 给定初始值：置k=0，权向量a(k)为任意值，可选常数0＜c≤1   + 输入样本yi={y1,y2,…,yn}，计算判决函数值g(yi)=aT(k)yi   + **按如下规则修改权向量**     - **若g(yi) ≤0，则a(k+1)=a(k)+cyi**   + 令k=k+1，返回第二步，当a对所有样本稳定不变时结束   4. 简述分级聚类算法的基本思想及两种基本途径。  答：按事物的相似性，或内在联系组织起来，组成有层次的结构，使得本质上最接近的划为一类，然后把相近的类再合并，依次类推，这就是分级聚类算法的基本思想。   * 分级聚类算法有两种基本思路   + 聚合法(agglomerative)：把所有样本各自看为一类，逐级聚合成一类。基本思路是根据类间相似性大小逐级聚合，每级只把相似性最大的两类聚合成一类，最终把所有样本聚合为一类。   + 分解法(divisive)：把所有样本看做一类，逐级分解为每个样本一类。   5. 如图所示的两类别样本点，在采用近邻法则和SVM支持向量机两种分类方法时，分界面会不会有区别？试分析原因。    答：有区别。Svm是寻找一个使得分类间隔最大的超平面，而近邻法的分界面，是由两类别样本中，有可能成为测试样本的近邻的两个样本之间的中垂面构成。在所给样本案例中，显然，如图所示的4个关键样本，会使得分界面有所区别。    **三、计算题（本大题共2小题，每题12分，共24分）**  1. 某疾病检查，ω1代表正常人，ω2 代表患病者。假设先验概率 P(ω1)=0.9, P(ω2)=0.1。  现有一被检查者，观察值为 x，查得 p(x|ω1)=0.2，p(x|ω2)=0.8，如按照最小错误率分类，则该检查者会被判为正常还是患病者？如从风险角度出发，希望在保证一定错误率的前提下，通过调整决策方式把该检查者尽可能的判定为病人，该如何解决？  答：1）. P(ω1)\* p(x|ω1)=0.18；P(ω2)\* p(x|ω2)=0.08  0.18>0.07，所以，该检查者会被判定为正常。  2）采用最小风险贝叶斯决策规则，设定风险系数，假设正确判决风险系数都为0，第一类错判为第二类时风险系数为1，第二类错判为第一类时，风险系数为a  则根据最小风险贝叶斯公式有时会被判决为第二类。故令a>9/4即可。（注：风险系数的设定方式不唯一） |

第 1 页，共 2 页

试卷编号：

|  |
| --- |
| 2. 给定样本数据如下： [-5, -6]T, [7, 6]T。  1）对其进行K-L离散变换（8分）  2）用1）的结果对样本数据[5，5]T做一维数据压缩。（4分）  答：1）KL变换  1. 求样本总体均值向量 m=[1, 0]T，平移样本，得到数据[-6, -6] T, [6, 6]T  2. 求协方差矩阵    3. 求特征根      4. 变换结果为    2）数据压缩，选择较大的特征值对应的特征根对数据进行变换  1. [5, 5]T – [1, 0] T =[4, 5] T  2. 投影结果为  **五、综述题（本大题共1小题，每题20分，共20分）**  1.联系实际，任选一个分类任务，对任务进行分析，并设计分类器。要求步骤详细可行，符合模式识别系统设计规则（非标准答案试题） |

第 2 页，共 2 页