

任课教师:

学号:

姓名:

班级:

订线

装订线

装订线

## 西安电子科技大学

考试时间 120 分钟

## 试 题

题号	一	二	三	四	五	六	总分
分数							

1. 考试形式: 闭卷; 2. 本试卷共 六 大题, 满分 100 分;  
3. 考试日期: 年 月 日; (答题内容请写在装订线外)

## 一、选择题 (15 小题, 每题 2 分, 共 30 分)

选择题答案请填入下方表格

题号	1	2	3	4	5	6	7	8
答案								
题号	9	10	11	12	13	14	15	
答案								

1. 统计学习方法的三要素不包括 ([请将答案写在前面表格中])  
A. 模型 B. 策略  
C. 算法 D. 数据
2. 监督学习分为生成方法与判别方法, 下面属于生成方法的是 ([请将答案写在前面表格中])  
A. 隐马尔科夫模型 B. k 近邻  
C. 决策树 D. 支持向量机
3. 关于 k 近邻方法说法正确的是 ([请将答案写在前面表格中])  
A. k 值越小学习的近似误差越小 B. k 值越大学习的估计误差越大  
C. k 值越小学习的估计误差越小 D. k 值越大学习的近似误差越小
4. 关于朴素贝叶斯估计说法错误的是 ([请将答案写在前面表格中])  
A. 它属于生成学习方法 B. 它的基本假设是条件独立  
C. 它通过后验概率最大化进行分类 D. 它不需要计算联合概率分布
5. 下列算法不属于决策树的是 ([请将答案写在前面表格中])  
A. ID3 B. C4.5 C. 核方法 D. CART
6. 下列关于 SVM 说法不正确的是 ([请将答案写在前面表格中])

- A. SVM 是一种二分类模型
- B. SVM 不能解决非线性分类问题
- C. 核方法可解决非线性分类问题
- D. 支持向量可以在间隔边界上

7. 关于 AdBoost 描述不正确的是 ([请将答案写在前面表格中])

- A. 它是弱分类器的加权组合
- B. 其训练误差是以指数率下降
- C. 它的训练误差边界不存在
- D. 提升树以分类树或回归树为分类器

8. 关于 EM 算法描述错误的是 ([请将答案写在前面表格中])

- A. EM 算法是含有隐变量的概率模型极大似然估计
- B. EM 算法可保证收敛到全局最优
- C. EM 算法是高斯混合模型参数估计的有效方法
- D. EM 算法的一次迭代可由 F 函数的极大-极大算法实现

9. 不属于隐马尔可夫模型的概率计算方法的是 ([请将答案写在前面表格中])

- A. 直接算法
- B. 前向算法
- C. 后向算法
- D. Baum-Welch 算法

10. 关于奇异值分解说法正确的是 ([请将答案写在前面表格中])

- A. 任何矩阵均存在奇异值分解
- B. 矩阵奇异值分解形式唯一
- C. 紧奇异值分解的秩与原矩阵相同
- D. 矩阵截断奇异值分解形式唯一

11. 关于主成分分析描述不准确的是 ([请将答案写在前面表格中])

- A. 总体主成分方差之和与随机变量主成分方差之和相同
- B. 主成分的协方差矩阵不一定是对角阵
- C. 随机变量 X 的第 k 主成分的方差是 X 协方差矩阵的第 k 个特征值
- D. 主成分分析是对数据进行正交变换

12. 关于聚类说法不正确的是 ([请将答案写在前面表格中])

- A. 聚类属于无监督学习
- B. 层次聚类是自上而下的
- C. 分裂聚类是自上而下的
- D. K 均值聚类的质量可以用类的平均直径来衡量

13. 设矩阵  $A$  的奇异值分解为  $A = U \Sigma V^T$ ，则下面不正确的是：([请将答案写在前面表格中])

- A.  $AV = U \Sigma$       B.  $A^T U = V \Sigma^T$       C.  $A^T A = V^T \Sigma \Sigma^T V$       D.  $AA^T = U \Sigma \Sigma^T U^T$

14. 关于逻辑斯蒂回归说法不正确的是 ([请将答案写在前面表格中])

- A. 逻辑斯蒂回归是分类方法      B. 它假设数据服从逻辑斯蒂分布  
C. 其参数可以通过极大似然法估计      D. 它仅适用于二分类

15. 关于核函数说法不正确的是 ([请将答案写在前面表格中])

- A. 核函数是原始数据空间到新空间的映射  
B. 核函数是原始数据空间到新空间的映射的内积  
C. 核函数是正定核的充要条件是其 Gram 矩阵为半正定矩阵  
D. 核函数即可定义在欧式空间又可定义在离散数据集

## 二、填空题 (10 小题, 每题 2 分, 共 20 分)

1. 统计学习方法的三要素为\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
2. 监督学习的三个重要应用方面是\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
3. 假设输入空间  $X \subseteq \mathcal{R}^n$ ，输出空间  $Y = \{+1, -1\}$ ， $x \in X$  表示实例特征向量， $y \in Y$  为实例类别，则输入空间到输出空间的函数  $f(x) =$  \_\_\_\_\_ 称为感知机。
4. 已知 3 个点  $x_1 = (1, 1)^T$ ,  $x_2 = (5, 1)^T$ ,  $x_3 = (4, 4)^T$ ， $L_2$  距离下  $x_1$  的最近邻点是\_\_\_\_\_。
5. 决策树中样本集合  $D$  对特征  $A$  的信息增益(ID3)可表示为:  $g(D, A) =$ \_\_\_\_\_。
6. 二项逻辑斯蒂回归模型可表示为条件概率分布:  $P(Y=1|x) =$ \_\_\_\_\_ 和  $P(Y=0|x) =$ \_\_\_\_\_。
7. AdaBoost 算法中弱分类器  $G_m(x)$  的系数  $\alpha_m$  与训练误差率  $e_m$  的关系为: \_\_\_\_\_。
8. EM 算法通过迭代求  $L(\theta) = \log P(Y|\theta)$  的极大似然估计。每次迭代包含两步，第一步为: \_\_\_\_\_; 第二步为: \_\_\_\_\_。
9. 隐马尔可夫模型的三要素为\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
10. 对任意一个实矩阵  $A$ ，设  $u_k, v_k^T, \sigma_k$  分别为矩阵  $A$  的第  $k$  个左奇异向量，右奇异向量，奇异值，则其外积展开式表示为: \_\_\_\_\_。

### 三、简答题（5 小题，每题 4 分，共 20 分）

1. 写出统计学习的策略中常用四种损失函数的名称及具体形式。
2. 试写出特征选择中信息增益算法的具体实现步骤。

3. 试写出非线性支持向量机中常用的核函数及其具体形式。

4. 试写出聚类方法中四种距离定义及具体形式。

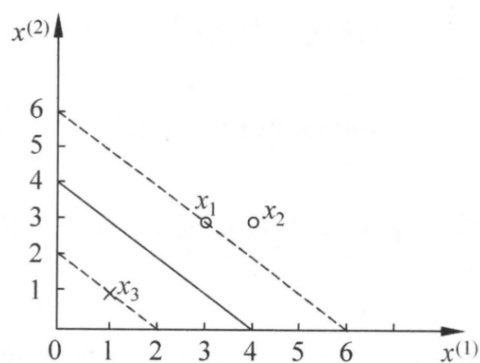
5. 试写出利用奇异值分解的主成分分析算法流程。

输入：  $m \times n$  的样本矩阵  $\mathbf{X}$ ，其每行元素的均值为 0；

输出：  $k \times n$  的样本主成分矩阵  $\mathbf{Y}$ ；

参数：主成分个数  $k$ 。

四、（10 分）已知一个如下图所示的训练数据集，其正例点是  $X_1 = (3, 3)^T$ ， $X_2 = (4, 3)^T$ ，负例点是  $X_3 = (1, 1)^T$ ，试求利用线性可分支持向量机学习算法计算最大间隔分离超平面。



五、(10 分) 假设有 4 个盒子，每个盒子里都装有红、白两种颜色的球，盒子里的红、白球数由下表给出，并按照下面规则抽球，产生一个球的颜色观测序列。

盒子编号	1	2	3	4
红球	7	4	2	9
白球	3	6	8	1

开始，从 4 个盒子里以等概率随机选取 1 个盒子，从这个盒子里随机抽出 1 个球，记录其颜色后，放回，然后，从当前盒子随机转移到下一个盒子。

**规则:**如果当前盒子是盒子 1，那么下一盒子一定是盒子 2;

如果当前是盒子 2 或 3，分别以概率 0.3 和 0.7 转移到左边或右边的盒子;

如果当前是盒子 4，以 0.5 的概率停留在盒子 4 或转移到盒子 1;

确定转移的盒子后，再从这个盒子里随机抽出 1 个球，记录其颜色，放回;

如此下去，重复进行 5 次，得到一个球的颜色观测序列:

$O = (\text{红}, \text{红}, \text{白}, \text{白}, \text{红})$

在此过程中，观察者只能观测到球的颜色序列，观测不到球是从哪个盒子取出的，即观测不到盒子序列。

试写出上例的隐马尔可夫模型  $\{\pi, A, B\}$

六、(10 分) 给定 5 个样本的集合，样本之间的欧氏距离由如矩阵  $M$  表示：

$$M = [d_{i,j}]_{5 \times 5} = \begin{bmatrix} 0 & 7 & 2 & 9 & 3 \\ 7 & 0 & 5 & 4 & 6 \\ 2 & 5 & 0 & 8 & 1 \\ 9 & 4 & 8 & 0 & 5 \\ 3 & 6 & 1 & 5 & 0 \end{bmatrix}$$

其中  $d_{i,j}$  表示第  $i$  个样本与第  $j$  个样本之间的欧氏距离。显然  $M$  为对称矩阵。  
应用聚合层次聚类法对这 5 个样本进行聚类。