

# 2022-2023 学年秋季学期《模式识别与机器学习》

## 期末考试试题 B 卷

承诺书：本人承诺考试过程中诚实守信，遵守考试纪律，不做任何违纪的事。

承诺人：\_\_\_\_\_（签名）

班级\_\_\_\_\_ 学号\_\_\_\_\_ 姓名\_\_\_\_\_

### 本试卷共七道题

题号	一	二	三	四	五	六	七	总分
成绩								

一、(10 分)判断下列说法的正误，在括号内画√表示正确，画×表示错误。

1. 特征选择的前向搜索算法可以得到最优特征组合。( )
2. 贝叶斯估计的结果只与数据有关。( )
3. 随机森林是一种 Boosting 方法。( )
4. 支持向量机是一种常见的线性分类器，它不能对数据做非线性分类。( )
5. 相似性度量学习为提升模型的性能，通常选用较难的三元组样本来计算损失函数。( )
6. 核主成分分析是一种线性方法。( )
7. 过拟合会增强模型的泛化性能。( )
8. 增加决策树的深度可能会导致决策树出现过拟合现象。( )
9. K 均值聚类算法的时间复杂度与聚类样本的数量成线性关系。( )
10. EM 算法的结果不依赖于初始值。( )

### 二、(20 分)选择题

1. (单选题) 下列哪类谓词逻辑问题不能用简单的感知器来解决？( )
  - A. 与
  - B. 或
  - C. 非
  - D. 异或
2. (单选题) 以下方法不属于子空间学习的数据降维方法是 ( )

- A. 拉普拉斯特征映射 LE
  - B. 局部保持投影 LPP
  - C. 主成分分析 PCA
  - D. 线性判别分析 LDA
3. (单选题) 关于参数估计, 下面哪个说法不正确? ( )
- A. 贝叶斯估计就是后验分布的最大值点
  - B. 最大似然估计比较简单直观
  - C. 若先验概率的信息是可靠的, 那么可认为贝叶斯估计比最大似然估计的结果更准确
  - D. 当训练样本数量无穷多时, 最大似然估计和贝叶斯估计的结果是一样的
4. (单选题) 以下机器学习方法, 可解释性最差的是 ( )
- A. 决策树
  - B. 支持向量机
  - C. 线性模型
  - D. 神经网络模型
5. (单选题) 支持向量机方法性能的有效性取决于 ( )
- A. 软边距参数
  - B. 核函数参数
  - C. 核函数选择
  - D. 以上所有
6. (多选题) 用 Parzen 窗法估计类概率密度函数时, 窗宽过窄导致波动过大的原因是 ( )
- A. 窗函数幅度过大
  - B. 窗函数幅度过小
  - C. 窗口中落入的样本数过少
  - D. 窗口中落入的样本数过多
7. (多选题) 关于决策树方法, 以下说法正确的是 ( )
- A. 基尼指数衡量了数据在类别上的不确定性的减小程度
  - B. 多变量决策树能得到与坐标轴呈一定夹角的分类面
  - C. 预剪枝和后剪枝策略都可以用于预防过拟合
  - D. 增益率准则对取值数目多的属性有所偏好
8. (多选题) 马氏距离和欧式距离都具有的特性包括 ( )
- A. 旋转不变性
  - B. 平移不变性
  - C. 尺度缩放不变性
  - D. 不受量纲影响的特性

9. (多选题) 关于特征选择方法, 以下说法正确的是 ( )
- A. 数据降维和特征选择都可以应对维度爆炸问题
  - B. 前向搜索、后向搜索和双向搜索都是贪心策略
  - C. 过滤式特征选择的计算开销通常比包裹式特征选择更大
  - D. 信息熵、类内类间距离、随机变量间的相关性都可以作为特征子集评价指标
10. (多选题) 关于数据聚类方法, 以下说法正确的是 ( )
- A. 高斯混合聚类问题可以使用 EM 算法来求解
  - B. K 均值聚类易受初始均值向量选取的影响
  - C. K 均值聚类、高斯混合聚类和层次化聚类方法都需要预先给定类簇个数
  - D. 设  $S_w$  是类内离散度矩阵,  $S_B$  是类间离散度矩阵, 则聚类目标可以是最大化  $|S_w S_B^{-1}|$

### 三、(20 分) 填空题

1. 深度学习训练过程中, 防止模型出现过拟合常采用的方法有: \_\_\_\_\_。
2. 数据降维与特征选择的主要区别是\_\_\_\_\_。
3. 构建决策树时, 选择最优划分属性的准则有\_\_\_\_\_。
4. 线性度量学习方法与深度度量学习方法的主要区别是\_\_\_\_\_。
5. 纯半监督学习和直推半监督学习的主要区别是 \_\_\_\_\_。
6. 贝叶斯网络中, 将有向图转变为无向图的过程叫做\_\_\_\_\_。
7. 基于最大可分性的子空间学习方法是\_\_\_\_\_。
8. 主动学习和半监督学习的主要区别是\_\_\_\_\_。
9. 对于开集人脸识别 (测试集中有训练集中没有出现过的人), 应将其建模为一个度量学习任务, 而不采用传统的多分类任务, 主要原因是\_\_\_\_\_。
10. 当采用反向传播算法训练神经网络时, 请简述如何根据网络的预测和样本的真实标签对网络参数进行优化\_\_\_\_\_。

### 四、计算题(12 分)

给定正例  $x_1 = (1, 3)$ ,  $x_2 = (-2, 3)$ , 负例  $x_3 = (2, 1)$ ,  $x_4 = (3, 4)$ , 计算支持向量、分类决策函数和分类超平面方程。如果使用基函数  $\phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ , 试求出核函数  $K(x, y)$ , 以及这个核函数下的支持向量、分类决策函数和分类超平面方程。



## 五、计算题(12 分)

训练数据如下表所示，假设弱分类器由竖直或水平分类面（ $x_1 < v_1$ 或 $x_1 > v_1$ 或 $x_2 < v_2$ 或 $x_2 > v_2$ ）产生，其阈值 $v_1$ 或 $v_2$ 使该分类器在训练数据集上分类误差率最低。试用 AdaBoost 算法学习一个强分类器。

序号	1	2	3	4	5	6	7	8	9	10
$(x_1, x_2)$	(2, 2)	(3, 1)	(4, 2)	(5, 2)	(1, 3)	(2, 3)	(3, 3)	(1, 4)	(2, 5)	(5, 4)
$y$	1	1	1	-1	1	-1	-1	1	-1	-1



## 六、计算题(13 分)

下表是一个由 15 个样本组成的贷款申请数据集。数据包括贷款申请人的 3 个特征(属性)：第 1 个特征是年龄，有 3 个可能值：青年，中年，老年；第 2 个特征是有自己的房子，有 2 个可能值：是，否；第 3 个特征是信贷情况，有 2 个可能值：好，一般。表的最后一列是类别，是否同意贷款，取两个值：是，否。

ID	年龄	有自己的房子	信贷情况	类别
1	青年	否	一般	否
2	青年	否	一般	否
3	中年	否	一般	否
4	中年	否	一般	否
5	老年	否	一般	否
6	青年	是	好	是
7	中年	是	一般	是
8	中年	是	好	是
9	老年	是	一般	是
10	老年	是	好	是
11	老年	否	好	是
12	青年	否	好	否
13	中年	否	好	否
14	中年	否	一般	是
15	老年	是	好	是

(1) 以 ID 为 1-11 的样本作为训练集，使用增益率准则建立决策树。

(2) 以 ID 为 12-15 的样本作为验证集，对 (1) 中的决策树进行后剪枝。





## 七、计算题(13 分)

给定前向传播神经网络如图.  $\{\mathbf{x}^{(i)}, y^{(i)}\}, i = 1, \dots, N$  为样本,  $\mathbf{x}^{(i)} \in \mathbb{R}^m$  为输入,

$$v_j^{(i)} = \sigma \left( \sum_{j=1}^3 w_{i,j} x_j^{(i)} \right), \quad \sigma(z) = \frac{1}{1 + \exp(-z)}$$

输出  $\hat{y}^{(i)} = \sigma(\boldsymbol{\theta}^\top \mathbf{v}^{(i)})$ , 损失函数为:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N (y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}))$$

(1) 求  $\frac{\partial \mathcal{L}}{\partial \theta_2}$

(2) 设学习率为  $\eta$ , 求用随机梯度下降法更新  $w_{1,2}$  的表达式



