

# 机器学习与数据挖掘- 第三次作业

## 选取最优划分准则构造决策树

学号: xxxxxxx, 姓名: xxxx, 邮箱: xxxxxxx@xxxxx.com

2023 年 3 月 30 日

### 1 决策树-信息增益准则

**题目 1:** 考虑下面的训练集: 共计 6 个训练样本, 每个训练样本有三个维度的特征属性和标记信息。详细信息如表 1 所示。

请通过训练集中的数据训练一棵决策树, 要求通过“信息增益”(information gain)为准则来选择划分属性。请参考《机器学习》(周志华)书中图 4.4, 给出详细的计算过程并画出最终的决策树。

表 1: 训练集信息

序号	特征 A	特征 B	特征 C	标记
1	0	1	1	0
2	1	1	1	0
3	0	0	0	0
4	1	1	0	1
5	0	1	0	1
6	1	0	1	1

**解答.** 此处用于写解答 (中英文均可)

(1) xxxxxxxxxxxx。

(2) xxxxxxxxxxxx。

**题目 2:** 模仿给出算法 ID3 的程序, 实现算法 C4.5 在西瓜数据集 2.0 上训练的程序, 给出训练完成后得到的决策树, 并且给出在两个测试样本 *test\_data\_1* 和 *test\_data\_2* 上的分类预测结果。

*test\_data\_1* = {'色泽': '青绿', '根蒂': '蜷缩', '敲声': '浊响', '纹理': '稍糊', '脐部': '凹陷', '触感': '硬滑'}

*test\_data\_2* = {'色泽': '乌黑', '根蒂': '稍蜷', '敲声': '浊响', '纹理': '清晰', '脐部': '凹陷', '触感': '硬滑'}

**解答.** 此处用于写解答 (中英文均可)

(1) xxxxxxxx。

(2) xxxxxxxx。