

实验六 决策树

1、实验描述

1.决策树算法：决策树是一种树形结构，其中每个内部节点表示一个特征，分支表示该特征的可能取值，叶节点表示类别。决策树的构建过程是递归地选择最优特征进行数据划分，直到所有数据属于同一类别或没有更多特征可供选择。

2.特征选择：在每一步，算法选择一个特征来划分数据。选择标准可以是信息增益、信息增益比或基尼指数。在这个例子中，使用的是基尼指数（gini）。

3.递归划分：算法递归地在每个节点上应用特征选择和数据划分，直到满足停止条件（如所有样本属于同一类别或没有更多特征可供选择）。

4.分类：构建好的决策树可以用于对新样本进行分类。根据样本的特征值，从根节点开始，沿着决策树的分支进行判断，直到到达叶节点，叶节点的类别即为样本的预测类别。

2、实验及结果分析

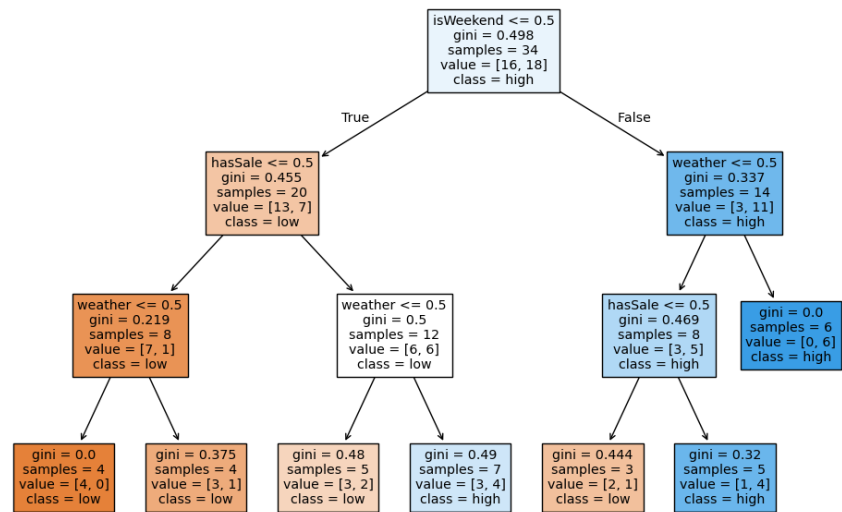
（1） 开发语言及运行环境；

python3.9.19, Mindspore2.4.1, , scikit-learn1.5.2

（2） 实验的具体步骤；

- 1.加载数据：从 CSV 文件中读取数据。
- 2.数据预处理：将类别特征转换为数值。
- 3.定义特征和目标变量：分离特征和目标变量。
- 4.训练模型：使用 CART 建立决策树模型。
- 5.可视化决策树：使用 matplotlib 绘制决策树。

（3） 根据实验数据集，按实验要求给出相应的结果（截图）；



(4) 对实验结果进行简要分析。

该决策图符合基本事实

3、实验心得

- (1) 学习了sklearn库的使用
- (2) 学习了CART算法的思想，了解了基尼指数的计算过程
- (3) 学会了使用matplotlib绘制决策树

5、程序文件名的清单

源码/exp6. py