

实验报告

实验名称： 实验四 ：决策树算法

专业班级： 数据1706 学号： 2017007790

学生姓名： 成泳甫

指导教师： 王 莉

2019年 12 月 17 日

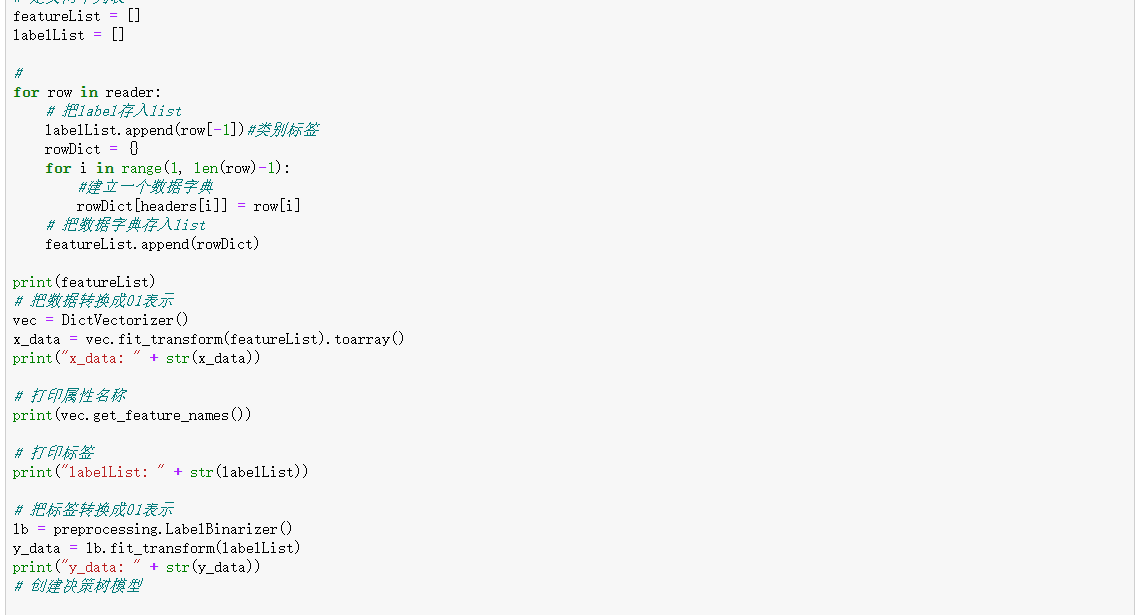
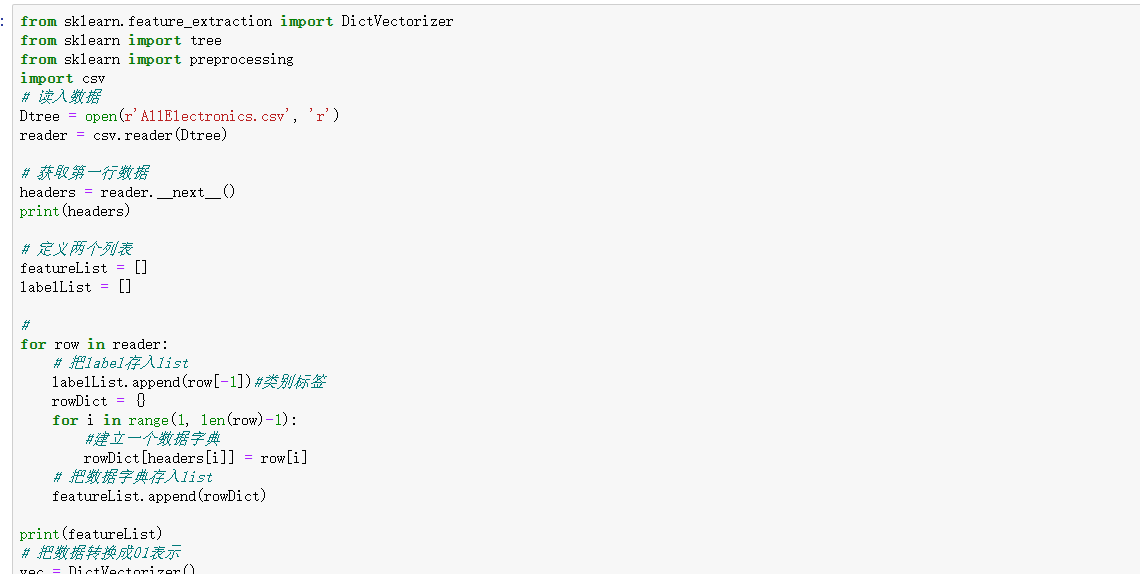
实验四：决策树算法

1. **实验目的**
2. 掌握决策树算法的原理
3. 根据给定数据集使用sklearn实现决策树
4. 学习使用graphviz可视化决策树
5. 使用决策树实现线性二分类和非线性二分类
6. **实验要求**
7. 独立完成实验
8. 书写实验报告书
9. **实验内容和步骤**
10. **实验①根据给定数据集使用sklearn实现决策树**
11. 读入并处理给定数据（数据为ALLElectronics.csv）

这里使用csv库方法读入，reader是一个迭代器，可以按行遍历整个数据，header为表头信息也就是字段的信息。For循环把所有数据分为字段信息（字典形式）和标签。然后为了把我们的数据信息转化为向量形式输入到决策树中，我们使用DictVectorizer将数据转化为01形式的数据。

1. 用sklearn创建决策树并进行训练（这部分需要得出C4.5算法和CART算法两种结果的模型）
2. 从训练集中选取数据进行简单的测试
3. 使用graphviz可视化决策树

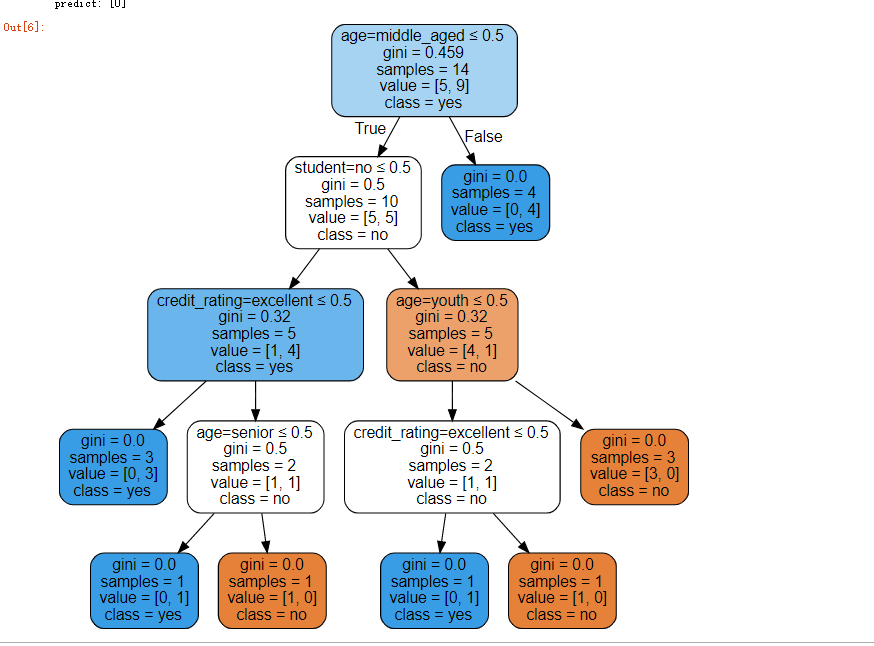
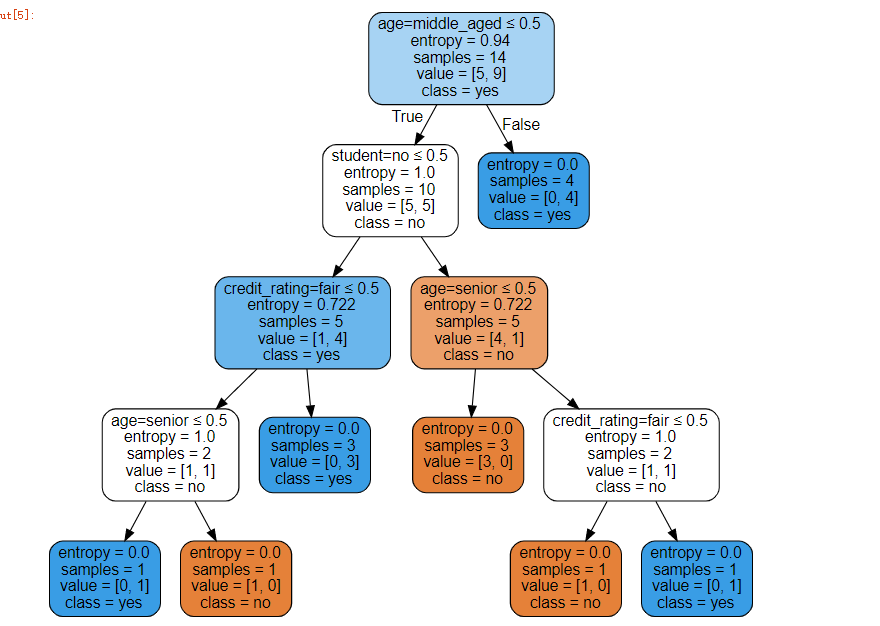
实验截图：



C4.5: 

Cart: 

运行结果：

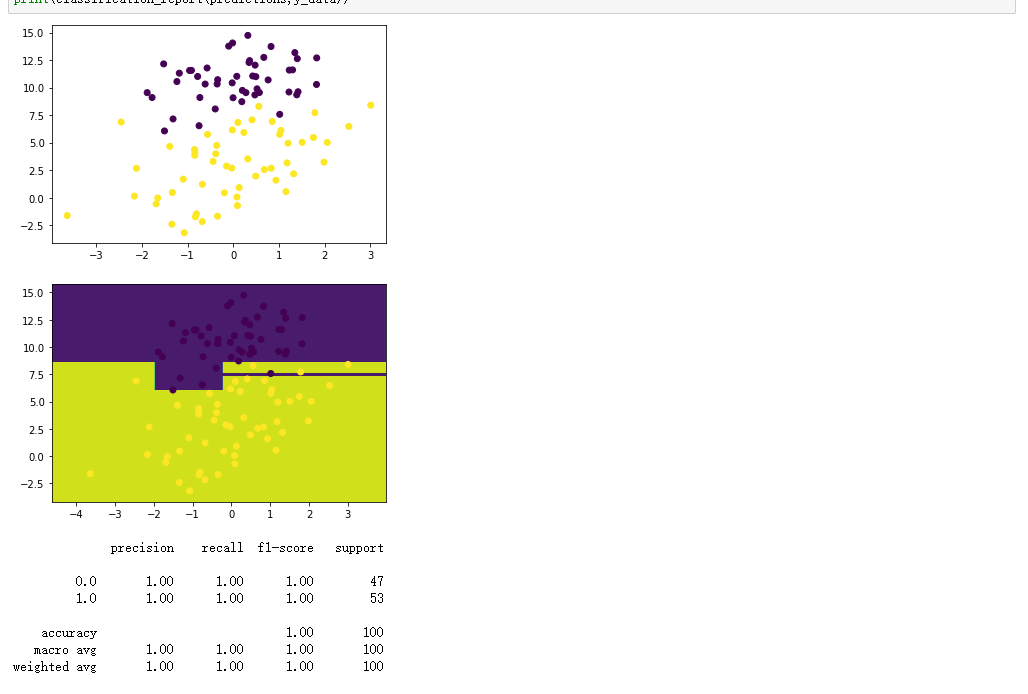


1. **实验② 使用决策树做线性二分类**
2. 导入数据(数据集为LR-testSet.csv)
3. 创建决策树并训练（使用CART算法）
4. 导出分类结果散点图

实验截图：

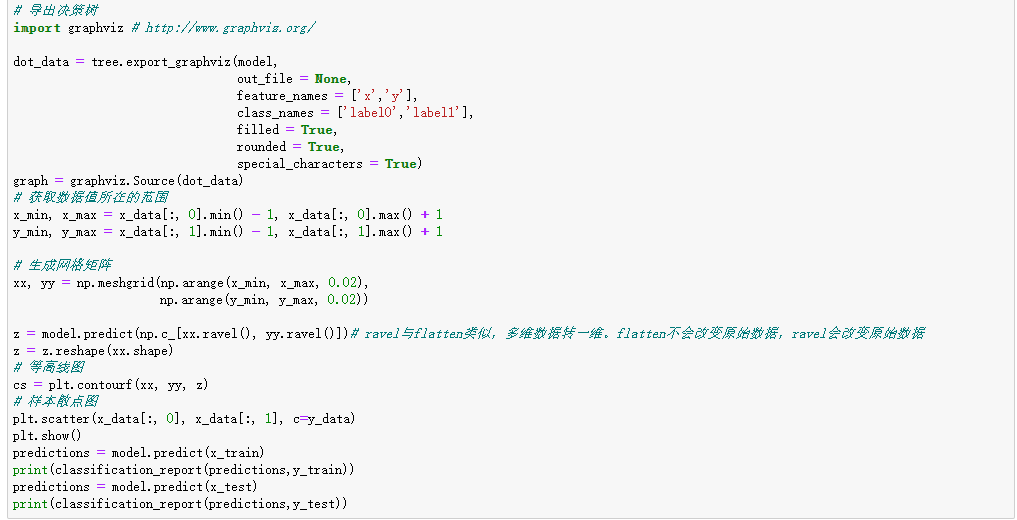
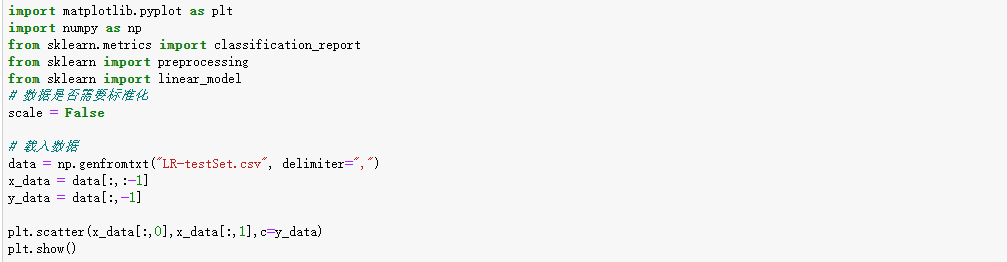


运行结果：

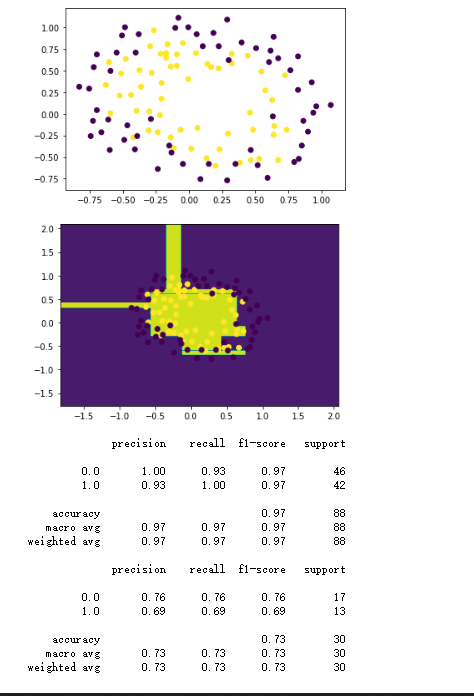


1. **实验③ 使用决策树做非线性二分类**
2. 导入数据(数据集为LR-testSet2.txt)
3. 创建决策树并训练（使用CART算法）需要将所有数据分割为训练集和测试集（分割比例3：1），训练集用于此处训练
4. 导出分类结果散点图
5. 将(2)留出的测试集用于模型测试，得出测试结果（用classification\_report得出结果即可）

实验截图：



运行结果：



实验代码链接：<https://github.com/yongfucheng/data-mining>