### 实验四 决策树算法

**一、实验目的**

1. 掌握决策树算法的原理
2. 根据给定数据集使用sklearn实现决策树
3. 学习使用graphviz可视化决策树
4. 使用决策树实现线性二分类和非线性二分类

**二、实验要求**

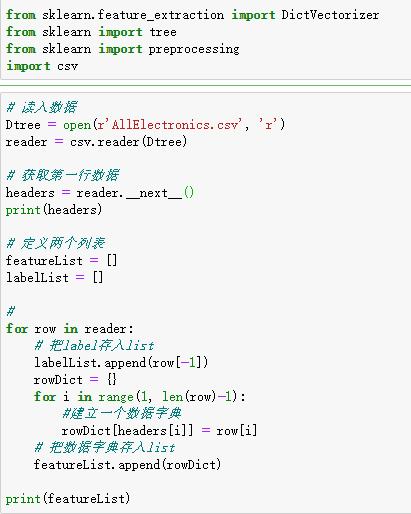
1.独立完成实验

2.书写实验报告书

**三、实验内容和步骤**

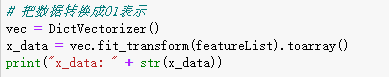
实验①根据给定数据集使用sklearn实现决策树

1. 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。

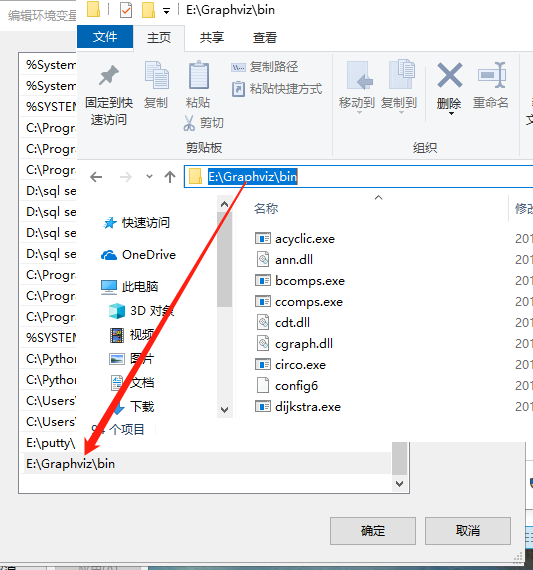
然后为了把我们的数据信息转化为向量形式输入到决策树中，我们使用DictVectorizer将数据转化为01形式的数据。



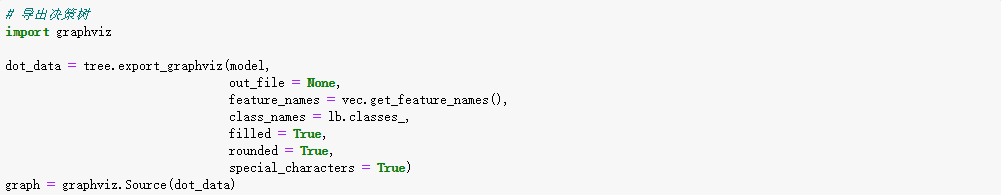
1. 用sklearn创建决策树并进行训练（这部分需要得出C4.5算法和CART算法两种结果的模型）
2. 从训练集中选取数据进行简单的测试
3. 使用graphviz可视化决策树

Graphviz的用法：

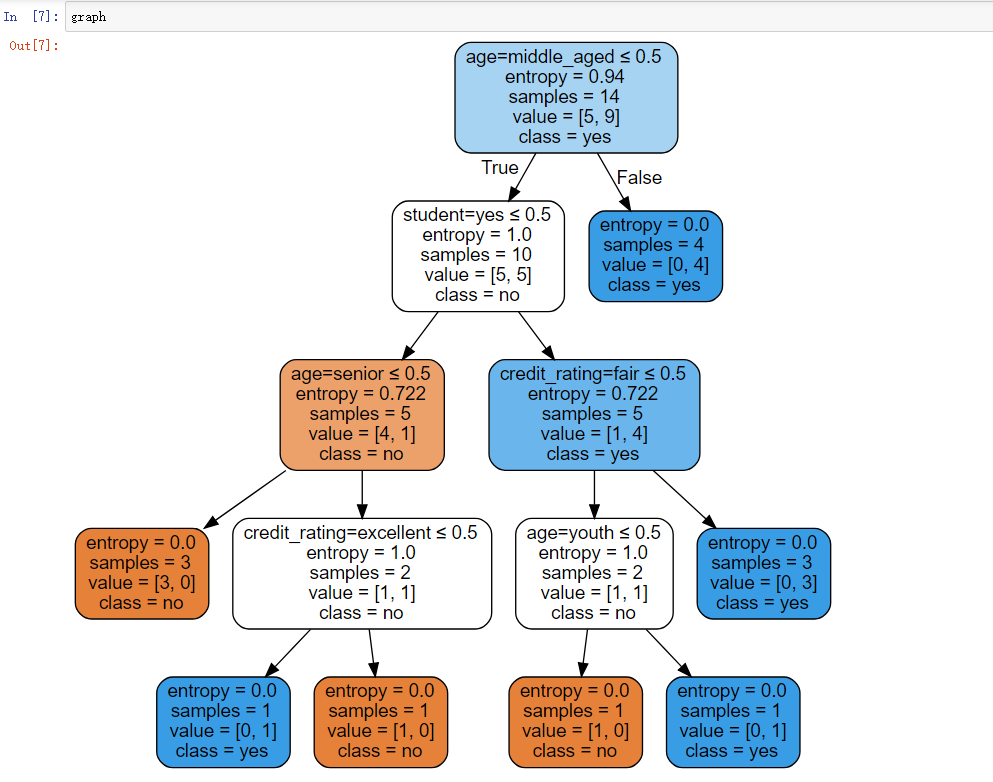
首先通过pip命令安装graphviz，然后安装graphviz-2.38.msi于某路径下，最后将该路径的bin目录加入环境变量中。



调用graphviz实现可视化：

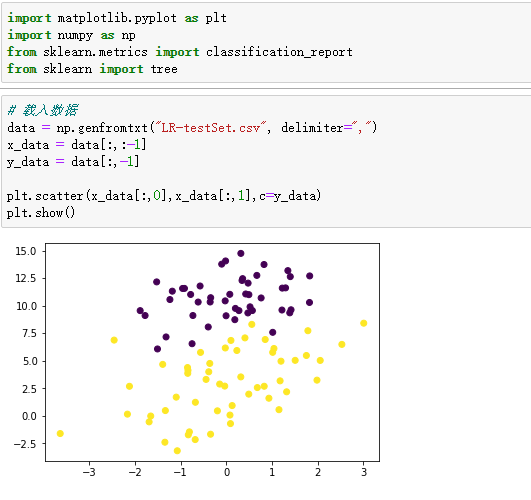


到处的决策树如下图（需要导出出C4.5和CART的两种）



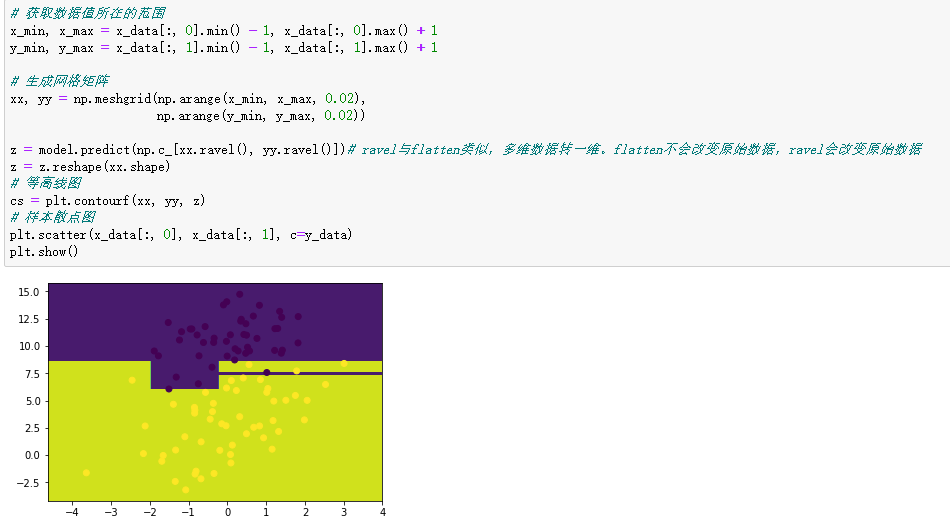
实验② 使用决策树做线性二分类

1. 导入数据(数据集为LR-testSet.csv)



(2)创建决策树并训练（使用CART算法）

(3)导出分类结果散点图



实验③ 使用决策树做非线性二分类

1. 导入数据(数据集为LR-testSet2.txt)
2. 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

1. 导出分类结果散点图
2. 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）