**机器学习报告名称：**决策树的ID3和C4.5的算法实现及运用

**一、基本信息**

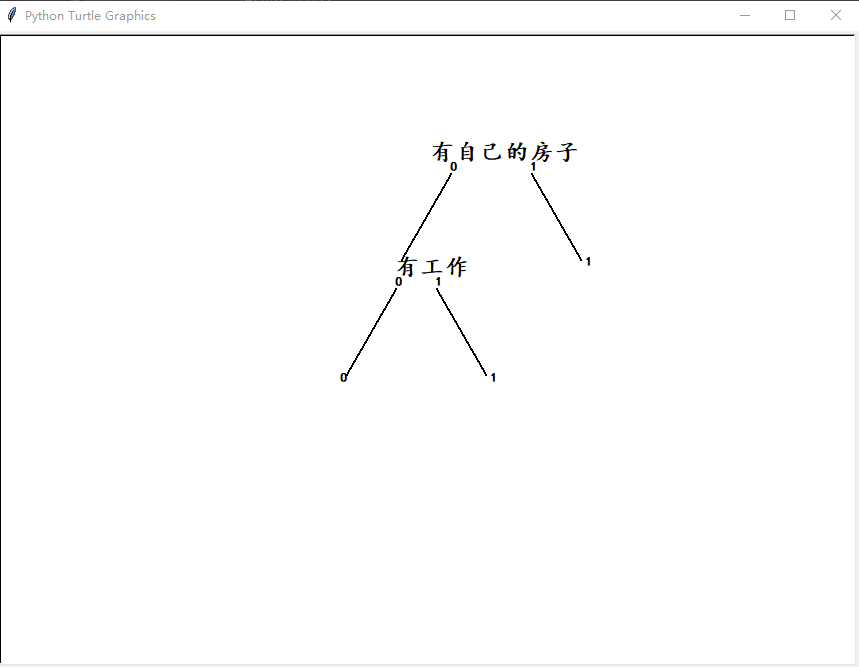
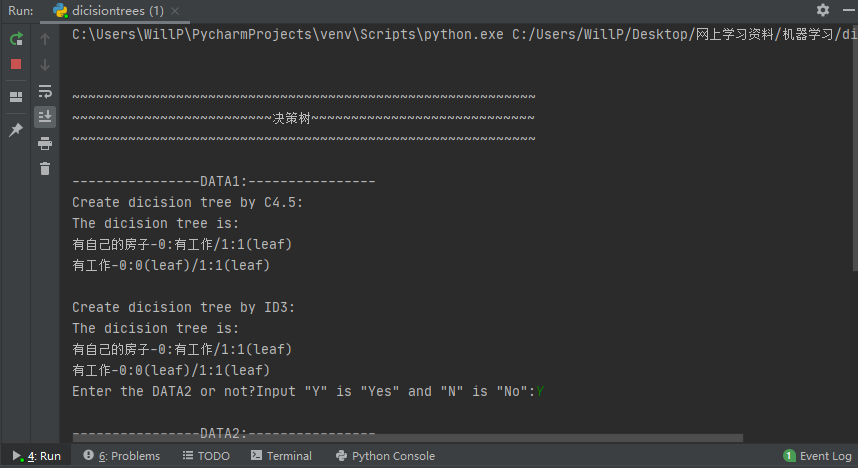
**姓名:**刘俊

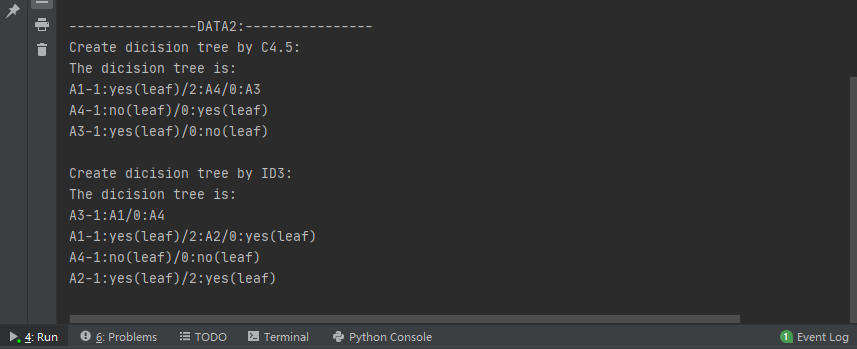
**学号：**41821153

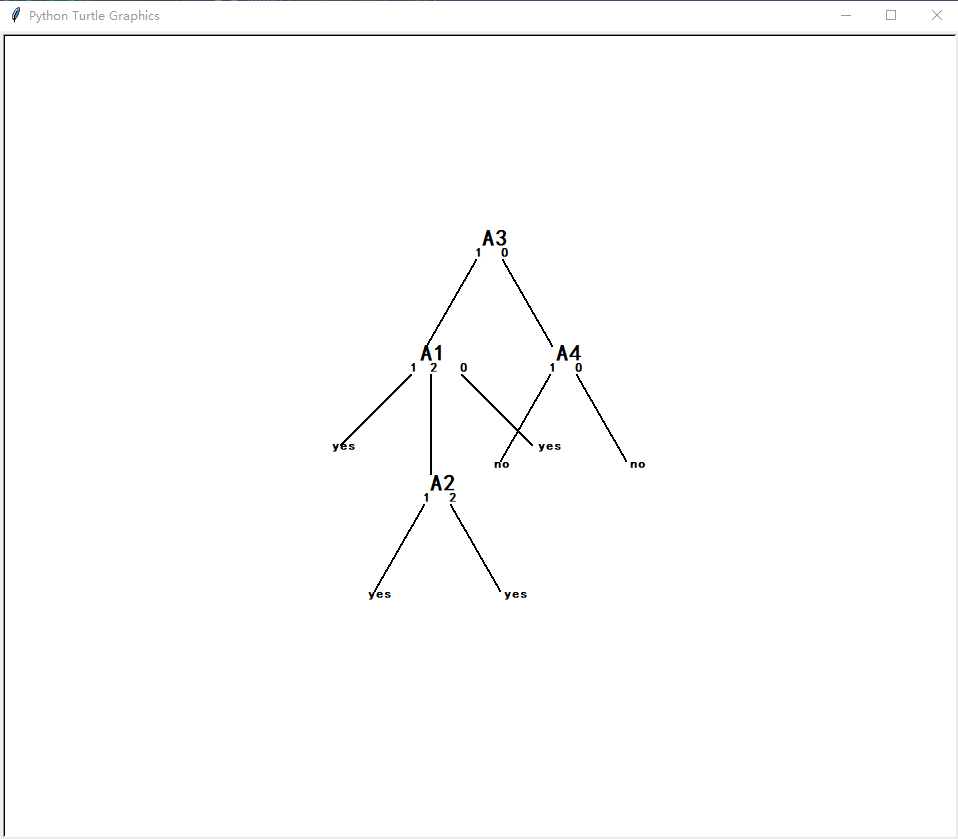
**二、上机作业内容与结果（包括图形截图）：**

1目的：分别用ID3算法和C4.5算法实现决策树的构建，并用其对数据集进行分类。

2 结果：







**三、文件与代码说明：**

<机器学习课程-code>

**decisiontree.py:**主要包括C45类和ID3类，ID3类为C45的继承类。其中类的方法主要包括：

def h(self, data: list): # 计算单数据集data的熵

def gda(self, data: list, feature: str): # 计算数据集data中特征feature的信息增益

def grda(self, data: list, feature: str): # 计算datadict数据集中feature特征的信息增益比

def bestfeature(self, data: list, features: list): # 返回数据集data下的特征features中最大信息增益比的特征

def creatdicisiontree(self, data: list, features: list): # 建树

def printtree(self): # 打印树

def drawtree(self): # 画树

分别对两个数据集分别进行了C4.5和ID3算法决策树的构建和分类，其中DATA1为课本中的数据-“贷款申请样本数据”，DATA2为老师所给数据-“是否购买计算机（ID3data.csv）”。

**ID3data.csv**：“是否购买计算机”的数据。

**四、备注：**

代码用到了numpy库和turtle库画树，所有代码均自己编写，没有进行二次的修改和优化，可能有可以优化的地方。