**决策树分类实验报告**

**——ID3、C4.5、CART**

1. 实验环境：

系统：Windows 10

语言：Python3.6

IDE：Sublime Text3.

数据集：数据集是自己创建的，详情见代码或者是图片“数据集.png”。

1. 实验目的：

了解三种不同决策树的工作原理，使用自己创建的数据集，对数据进行分类，完成决策树的绘制。

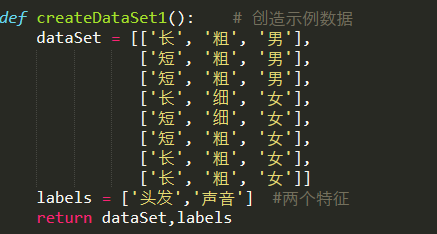
1. 实验过程：
2. 学习三种决策树的工作原理，了解到三种决策树的区别如下：

(1) ID3算法以信息增益为准则来进行选择划分属性，选择信息增益最大的进行划分；

(2) C4.5算法先从候选划分属性中找出信息增益高于平均水平的属性，再从中选择增益率最高的；

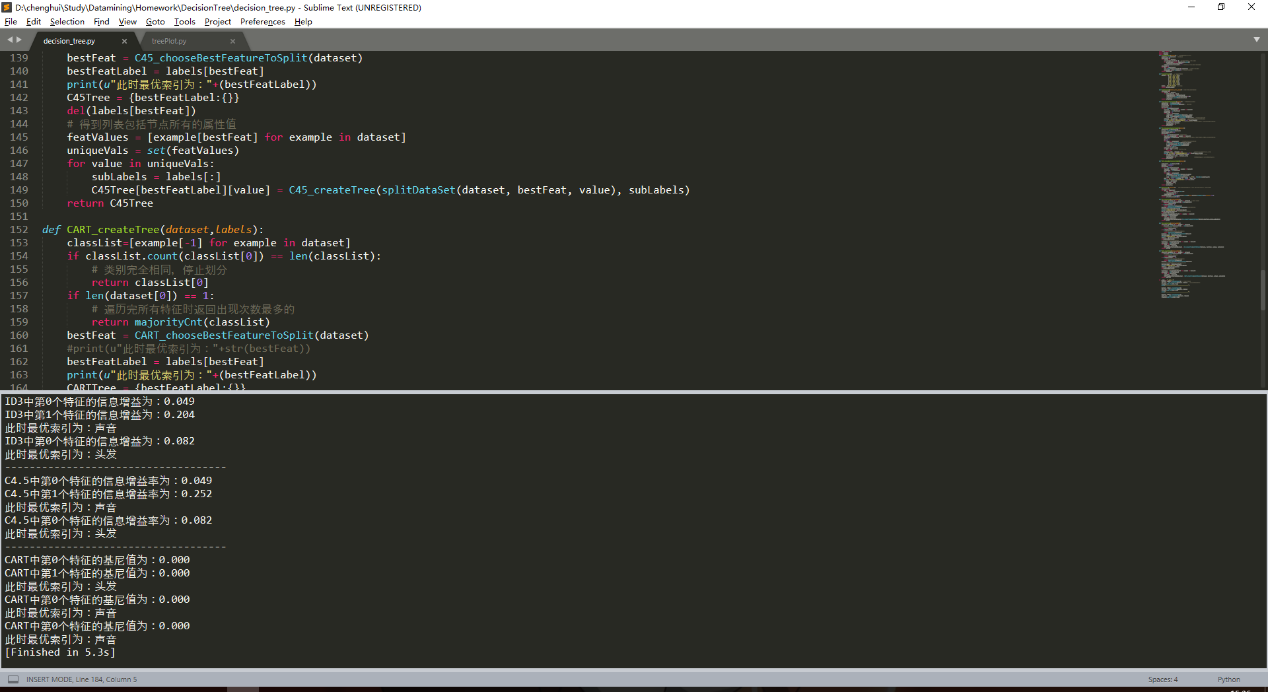
(3) CART算法使用“基尼指数”来选择划分属性，选择基尼值最小的属性作为划分属性。

1. 实验数据集：

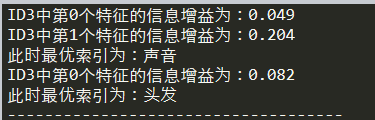


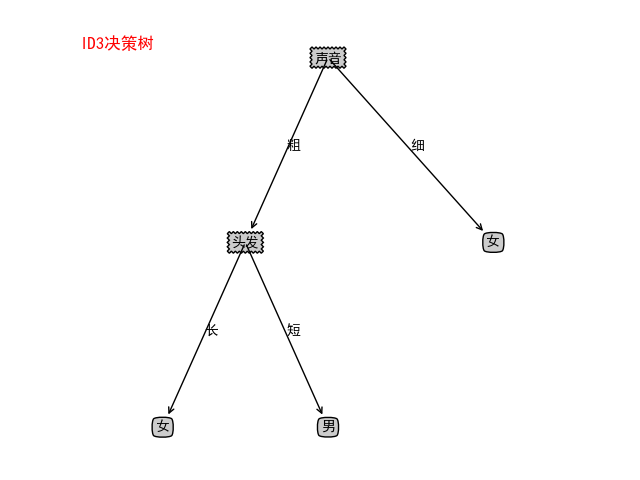
1. 编写代码，此次实验编程语言使用的是python。详细算法步骤见文件decision\_tree.py和treePlot.py中的注释。decision\_tree.py是用来实现属性划分的，treePlot.py是用来实现最终决策树绘制的。
2. 实验结果

实验结果如下：

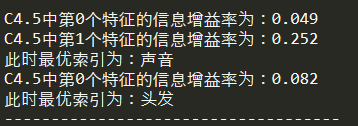


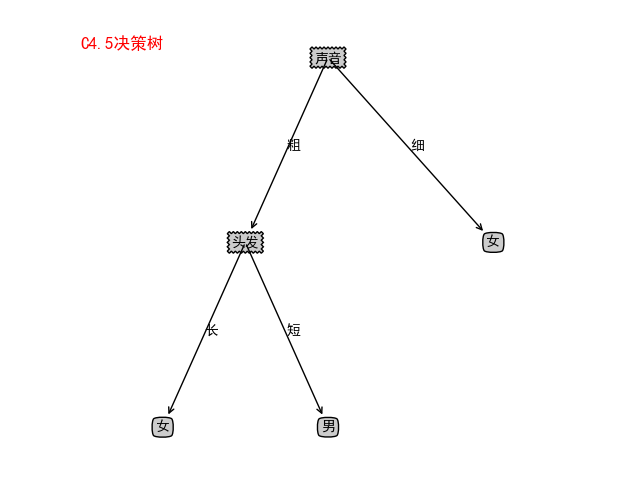
**ID3决策树：**

****

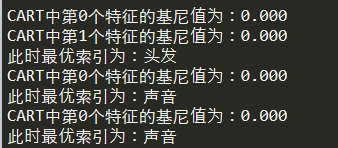
****

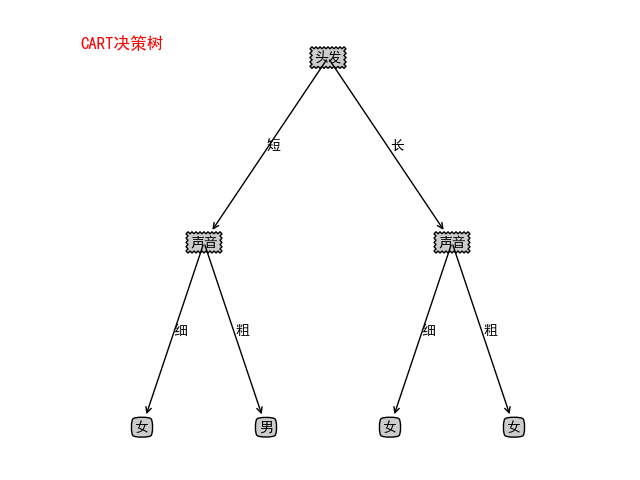
**C4.5决策树：**

****



**CART决策树：**

****

****

**实验结论：**

由以上实验结果我们可以看到，ID3和C4.5决策树的最优索引相同，所绘制出来的决策树是也是相同的，而CART绘制出来的决策树与前2种不同。