**实验五 决策树分类算法分析与实现**

【实验目的】

1 理解决策树分类算法的基本原理。

2 掌握决策树分类算法的编码实现方法。

3 理解决策树分类算法的优劣势。

【实验类型】

设计型

【实验学时】

2学时

【实验环境】

Windows 7以上操作系统

Python3.0以上版本

Pycharm开发环境

Spyder开发环境

【实验要求】

决策树分类算法是最为常见的一种分类算法，通过属性划分来建立一棵决策树，测试对象通过在树上由顶向下搜索确定所属的分类。决策树的构建主要需要解决两个问题：

**（1）树的每次成长，选择哪个属性进行划分，可以参考下面几个标准：**

**A Gini系数**



**多分支Gini系数的组合方法**



**B 基于熵的信息增益或信息增益率**

**熵的定义**



**信息增益的定义**



**信息增益率的定义**





**C 误分率**



**（2）什么时候在一个节点上停止生长（继续划分），可考虑包含下面两个原则：**

**A 当节点分配的所有记录都属于同一类时停止**

**B 当没有属性可以使用时停止**

**请使用python编写程序实现典型的决策树分类算法并自行设计数据进行测试，最后将实验报告（包含附录源代码）上交到QQ作业。**