

# 实验六 Decision Tree

## 一、说明

- 实验采用jupyter notebook, 请填写完代码后提交完整的ipynb文件
- 文件命名 规则: 班级\_姓名\_ML2019\_HW6.ipynb, 如计科\_1706\_张三\_ML2019\_HW6.ipynb
- 提交方式: 采用在线提交至:
- <http://pan.csu.edu.cn:80/invitation/076d4c3e-e3a3-401c-9d47-c3778f4e619f>
- 实验提交截至日期: 2019.12.10 23: 59

## 二、实验内容

本实验旨在逐步地指导实现常用的决策树算法, 包括ID3, C4.5和CART, 并用应用解决分类问题, 来直观地观察三种算法的不同。在实验中我们会进一步对python在机器学习中的应用进行实践。

## 三、实验目标

- 掌握决策树算法的思想
- 实现ID3, C4.5树和CART树的核心算法
- 学习python在机器学习中的应用

## 四、实验操作步骤

### 1. 启动 jupyter notebook

打开文件 DecisionTree.ipynb

### 2. 完成代码

- 任务 1 导入包和创建数据集
- 任务 2 ID3 决策树
- 任务 2.1 完成香农熵计算函数
- 任务 2.2 完成基本功能函数

- **任务 2.3** 用信息增益选择分类特征
- **任务 2.4** 生成 ID3 决策树
- **任务 3** C4.5 决策树
- **任务 3.1** 用信息增益比选择分类特征
- **任务 3.2** 生成 C4.5 树
- **任务 4** CART 决策树
- **任务 4.1** iris 数据集读取和预处理
- **任务 4.2** 完成基尼指数计算函数
- **任务 4.3** 完成基本功能函数
- **任务 4.4** 用基尼系数选择特征及划分点
- **任务 4.5** 生成 CART