**决策树实验**

**实验任务：**

实现ID3决策树，并在给定的数据集上进行5折交叉验证。并观测训所得到的决策树在训练集和测试集的准确率，从而判断该决策树是否存在过拟合。在此基础上实现预剪枝和后剪枝，并比较预剪枝树与后剪枝树在训练集和测试集上的准确率。

**编程语言：**java, matlab, python, C++，C均可

**数据集：**

**鸢尾花卉Iris数据集描述：**

iris是鸢尾植物，这里存储了其萼片和花瓣的长宽，共4个属性，鸢尾植物分三类。所以该数据集一共包含4个特征变量，1个类别变量。共有150个样本，鸢尾有三个亚属，分别是山鸢尾 (Iris-setosa)，变色鸢尾(Iris-versicolor)和维吉尼亚鸢尾(Iris-virginica)。

也就是说我们的数据集里每个样本含有四个属性，并且我们的任务是个三分类问题。三个类别分别为： Iris Setosa（山鸢尾），Iris Versicolour（杂色鸢尾），Iris Virginica（维吉尼亚鸢尾）。

例如：

样本一： 5.1, 3.5, 1.4, 0.2, Iris-setosa

其中“5.1，3.5，1.4，0.2”代表当前样本的四个属性的取值，“Iris-setosa”代表当前样本的类别。

**实验完成要求：**

完成上述实验的编码，并把实验流程，算法思想和在给定数据集上得到的指标记录到实验报告里。向助教老师演示所实现代码，并解释核心代码的思想。