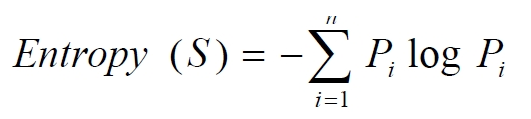
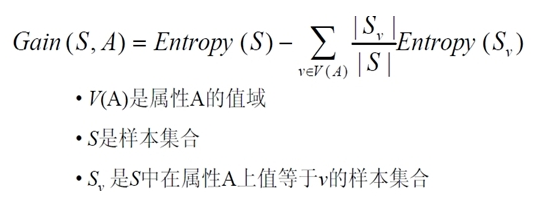
**决策树算法**

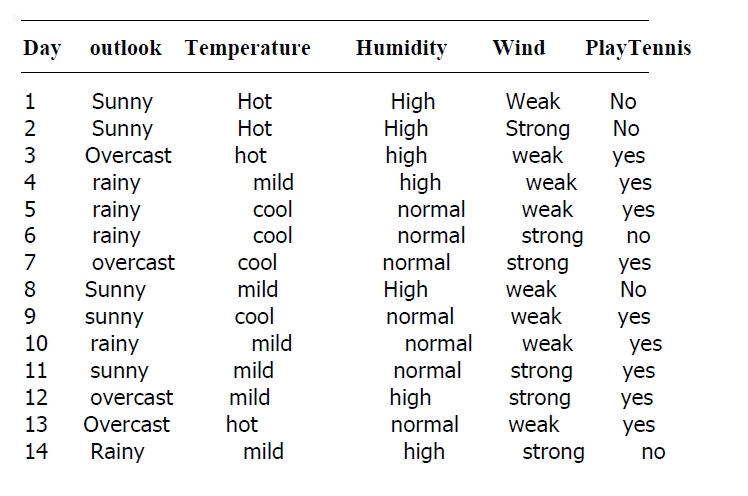
决策树算法是非常常用的分类算法，是逼近离散目标函数的方法，学习得到的函数以决策树的形式表示。其基本思路是不断选取产生信息增益最大的属性来划分样例集和，构造决策树。信息增益定义为结点与其子结点的信息熵之差。信息熵是香农提出的，用于描述信息不纯度(不稳定性)，其计算公式是



Pi为子集合中不同性(而二元分类即正样例和负样例)的样例的比例。这样信息收益可以定义为样本按照某属性划分时造成熵减少的期望，可以区分训练样本中正负样本的能力，其计算公式为：



样例集合为：



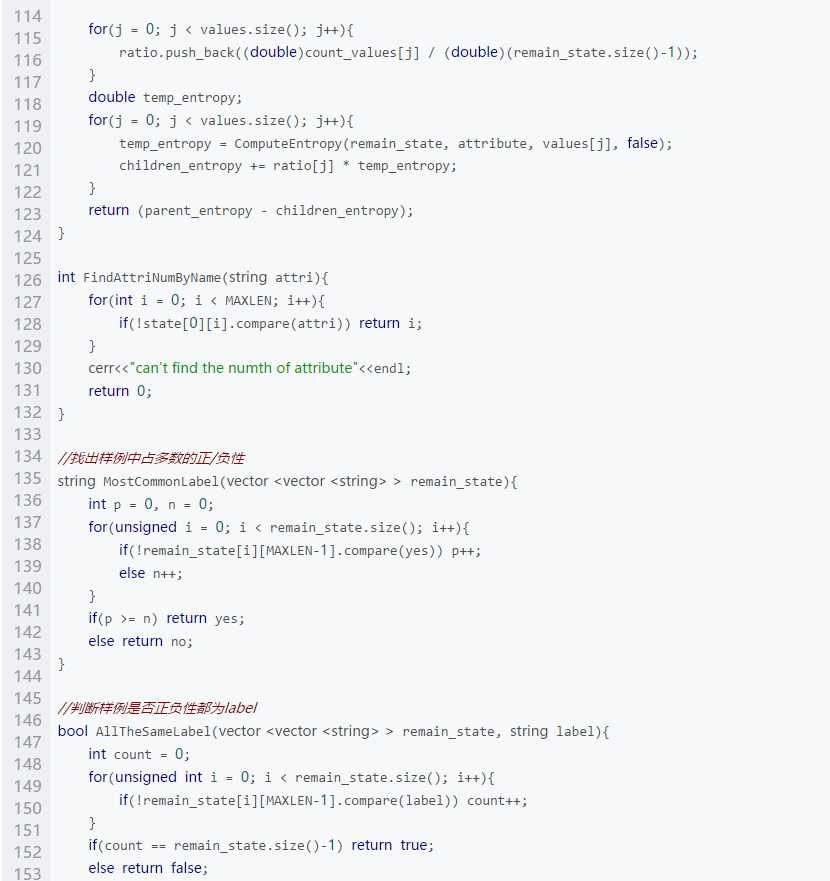
该表记录了在不同气候条件下是否去打球的情况，根据该表用程序输出决策树

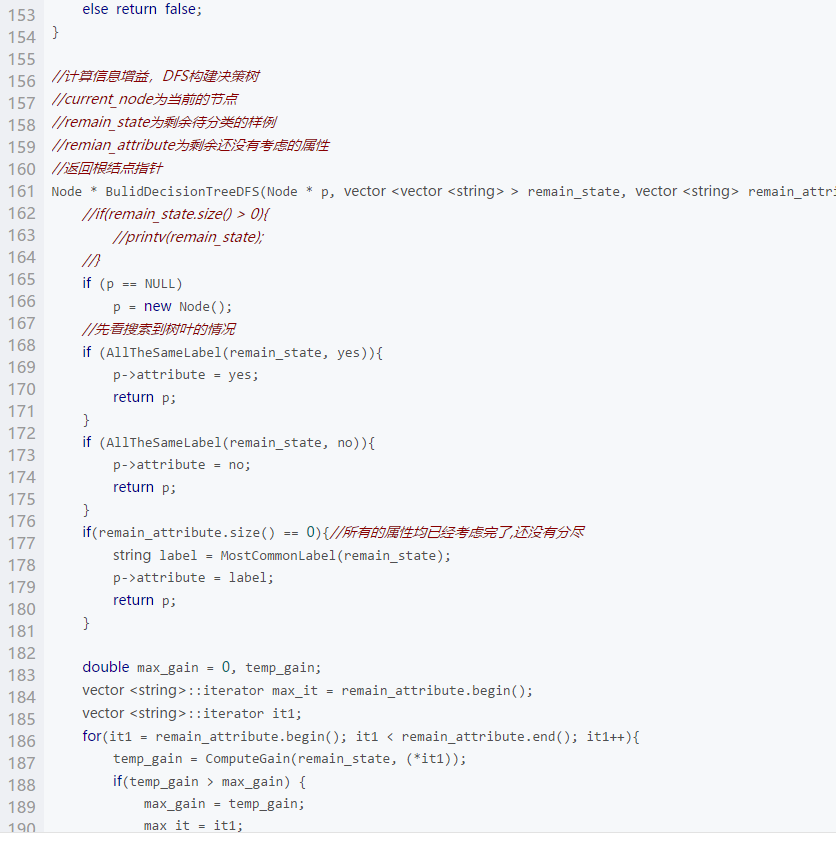
以下为代码实现：

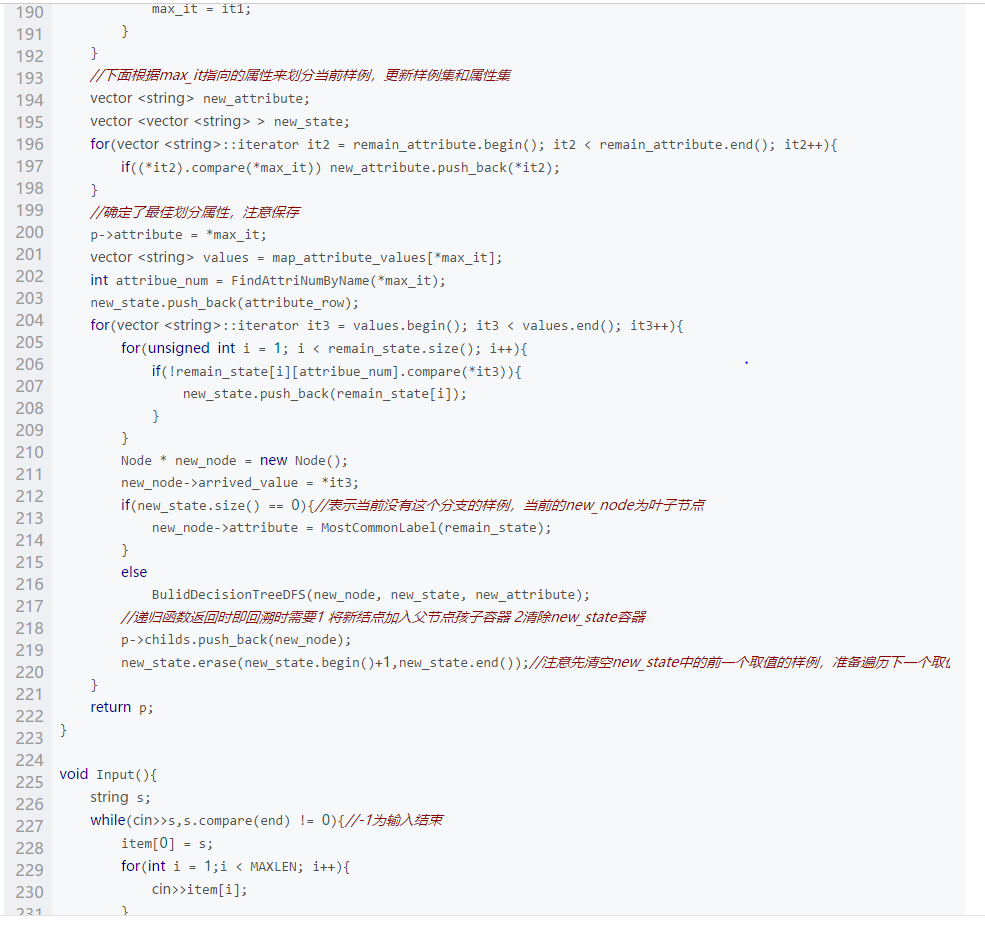




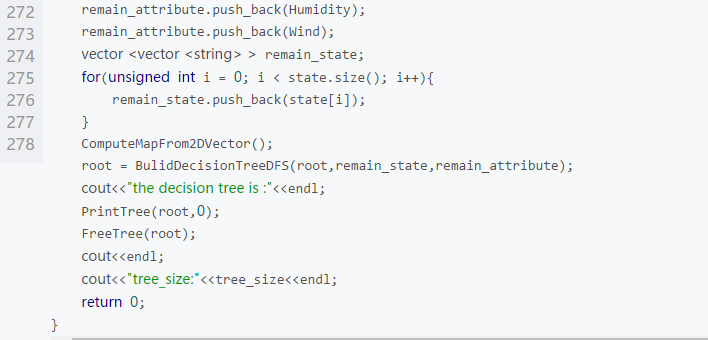




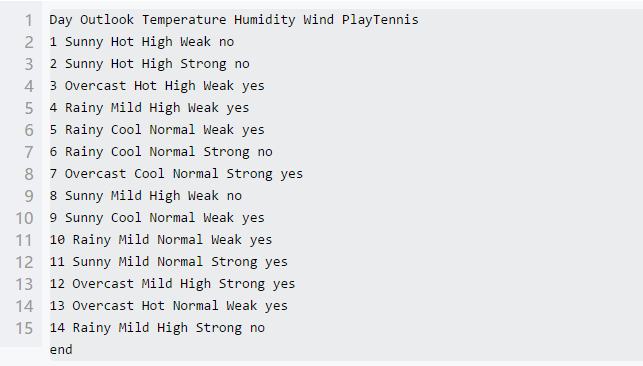




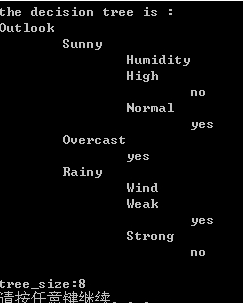




此为输入训练数据



输出决策树如下：



而图形表示为：

