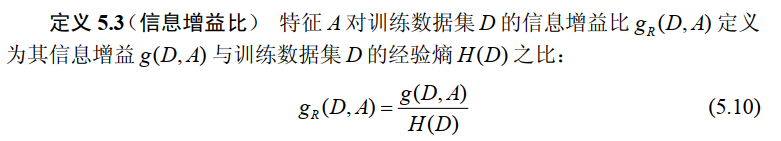
**《决策树模型四》**

**决策树生成算法-C4.5算法**

ID3 算法的核心是在决策树各个结点上应用信息增益准则选择特征，进行决策点的划分，从而构建决策树。而C4.5使用信息增益比来进行特征的选择：



用比值往往更稳定，更能说明效果。

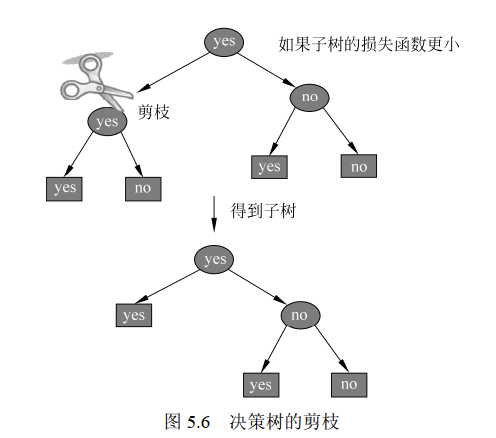
具体的算法详情：略过

**为什么要剪枝：**

决策树生成算法递归地产生决策树，直到不能继续下去为止．这样产生的树往往对训练数据的分类很准确，但对未知的测试数据的分类却没有那么准确，即出现过拟合现象**（过拟合：学习好考试不好）**

过拟合的原因在于学习时过多地考虑如何提高对训练数据的正确分类，从而构建出过于复杂的决策树。**（可以参考机器学习常用解决过拟合的方法）**

对于决策树而言：解决这个问题的办法是考虑决策树的复杂度，对已生成的决策树进行简化。



决策树剪枝通过优化损失函数还考虑了减小模型复杂度．决策树生成学习局部的模型而决策树剪枝学习整体的模型。