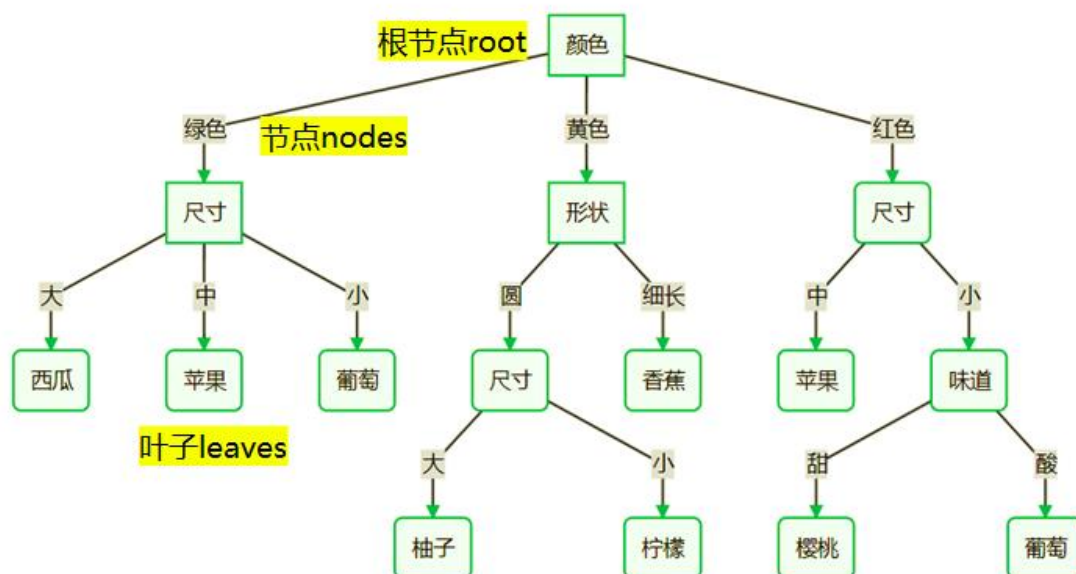


决策树

Decision Tree

决策树是一种机器学习的方法。决策树的生成算法有 ID3, C4.5 和 C5.0 等。决策树是一种树形结构，其中每个内部节点表示一个属性上的判断，每个分支代表一个判断结果的输出，最后每个叶节点代表一种分类结果。



决策树的生成主要分两步，这两步通常通过学习已经知道分类结果的样本来实现。

1. 节点的分裂：一般当一个节点所代表的属性无法给出判断时，则选择将这一节点分成 2 个子节点（如不是二叉树的情况会分成 n 个子节点）
2. 阈值的确定：选择适当的阈值使得分类错误率最小（Training Error）。

比较常用的决策树有 ID3, C4.5 和 CART（Classification And Regression Tree），CART 的分类效果一般优于其他决策树。

决策树是一种十分常用的分类方法，属于监管学习，监管学习就是给出一堆样本，每个样本都有一组属性和一个分类结果，也就是分类结果已知，那么通过学习这些样本得到一个决策树，这个决策树能够对新的数据给出正确的分类。

决策树易于理解和实现，人们在在学习过程中不需要使用者了解很多的背景知识，这同时是它能够直接体现数据的特点，只要通过解释后都有能力去理解决策树所表达的意义。