



3.4 其他决策树算法

- ID3算法
 - C4.5的前身
 - J.罗斯·昆兰 1986年提出的。
 - 与C4.5最大的不同——ID3使用信息增益来选择分裂属性。
- CART (Classification And Regression Tree, 分类回归树)
 - 1984年雷奥·布莱曼 (Leo Breiman) 等人提出的。
- CHAID决策树算法
 - 戈登V.凯斯 (Gordon V. Kass) 于1980年提出的。
 - CHAID与C4.5和CART不同, 它要求所有属性为分类类型, 且使用 χ^2 显著性检验来选择分裂属性。
 - CHAID具有统计学特色, 在SAS和SPSS等商业统计软件中应用很好。



3.5 决策树小结

- **优点**

- (1) 容易被理解和被解释，并且可以被映射到一组更具吸引力的产生式规则。
- (2) 不需要对数据的性质作预先的假设。
- (3) 能够使用数值型数据和分类类型数据的数据集建立模型。

- **局限性**

- (1) 输出属性必须是分类类型，且输出属性必须为一个。
- (2) 决策树算法是不稳定 (Unstable) 的。
- (3) 用数值型数据集创建的树较为复杂（如例2.3中的未剪枝的决策树），因为数值型数据的属性分裂通常是二元分裂。