分类树：classification tree，Y是无序离散变量

回归树：regression tree，Y是连续变量或有序变量

|  |  |  |  |
| --- | --- | --- | --- |
| 算法 | 评价函数 | 停止条件 | 备注 |
| THAID | 计算节点内部不同Y值的掺杂比例，争取各子节点的掺杂比例之和最低 --- 即尽可能单纯 | 掺杂率的相对降幅< 阀值 | X连续：二叉  X离散：||X||叉 |
| C4.5 | 熵 | 1. 先构造一个很大的树，然后再剪枝 2. 估计误差率的启发式公式 |  |
| CART | Gini index（一种二项式方差） | 1. 先构造一个很大的树，然后再剪枝 2. 验证集 + 10-fold的交叉验证 | R中rpart |
| FACT | 逐个测试X和Y的关系，取最充分的那个X： |  |  |
| CRUISE | 逐个测试X和Y的关系，取最充分的那个X：  chi squared tests | 1. 先构造一棵很大的树，然后剪枝（剪枝使用*CART*） |  |
| GUIDE | 1、逐个测试X和Y的关系，取*significance probability*最小的X\*：  chi squared tests  2、划分为2个子集，使得*the sum of Gini indexes*最小 | 1. 先构造一棵很大的树，然后剪枝（剪枝使用*CART*） |  |
| QUEST | 逐个测试X和Y的关系，取最充分的那个X：  有序离散变量：chi squared tests  无序离散变量： analysis of variance  (ANOVA) tests | 1. 先构造一棵很大的树，然后剪枝（剪枝使用*CART*） |  |
| CTree |  |  |  |