决策树剪枝原理——Cost-Complexity-Pruning

c1p

**后剪枝：代价复杂剪枝-CCP(Cost-Complexity-Pruning)**



**算法公式:**

**计算所有非叶子节点的表面误差率增益值**

**选择表面误差率增益值最小的非叶子节点（若多个非叶子节点具有相同小的表面误差率增益值，选择节点数最多的非叶子节点）**

R(T)=

**示例计算这部分的**



T2这个节点误差率，也就是把这部分的60个分成一类，里面有20个是错误的；

T2在这个途中细分成几部分之后的错误率是R(T)；现在的R(T)=2/18

T2这个节点下有3片叶子，所以

T2的=0.203472222

计算每个T节点的剪掉最小的那个，把下面的部分直接变成一片叶；如果很多节点的相等，就找下面分出叶子最多的剪掉，也就是最大的。

参考资料：<http://blog.csdn.net/zhang4418876/article/details/47022821>

<http://www.cnblogs.com/yonghao/p/5064996.html>

<http://blog.sina.com.cn/s/blog_4e4dec6c0101fdz6.html>

<http://www.academia.edu/7032066/How_to_compute_the_complexity_parameter_in_CART>

<http://gormanalysis.com/decision-trees-in-r-using-rpart/>