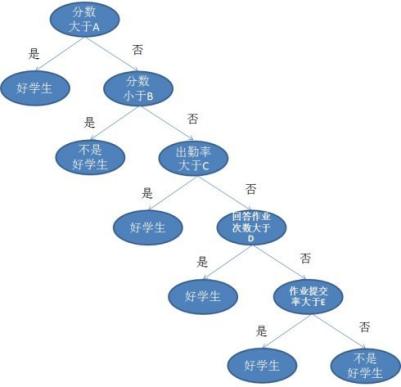
# 原理阶段

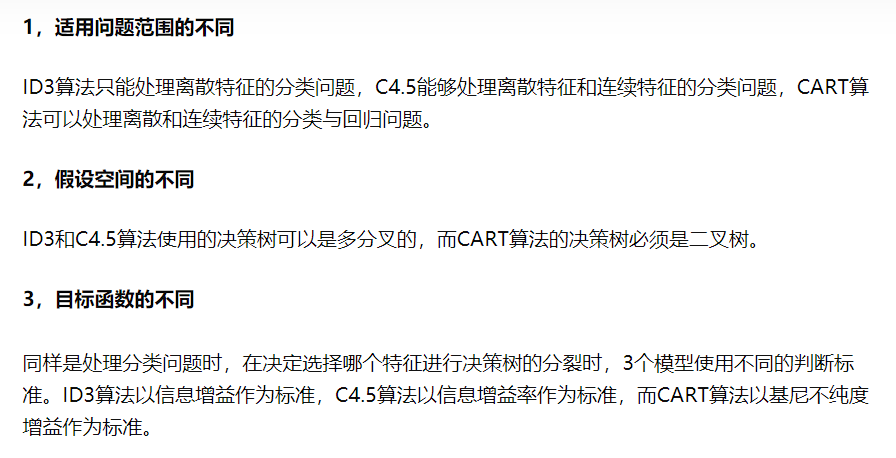
1. 什么是决策树?

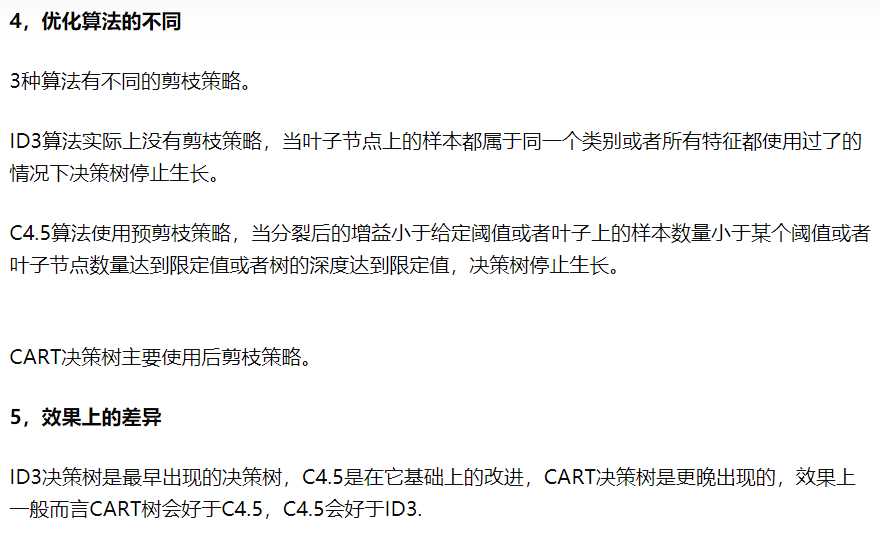




1. 决策树的生成算法

ID3, C4.5和cart

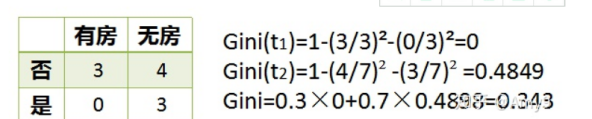






1. CART原理

https://zhuanlan.zhihu.com/p/139523931



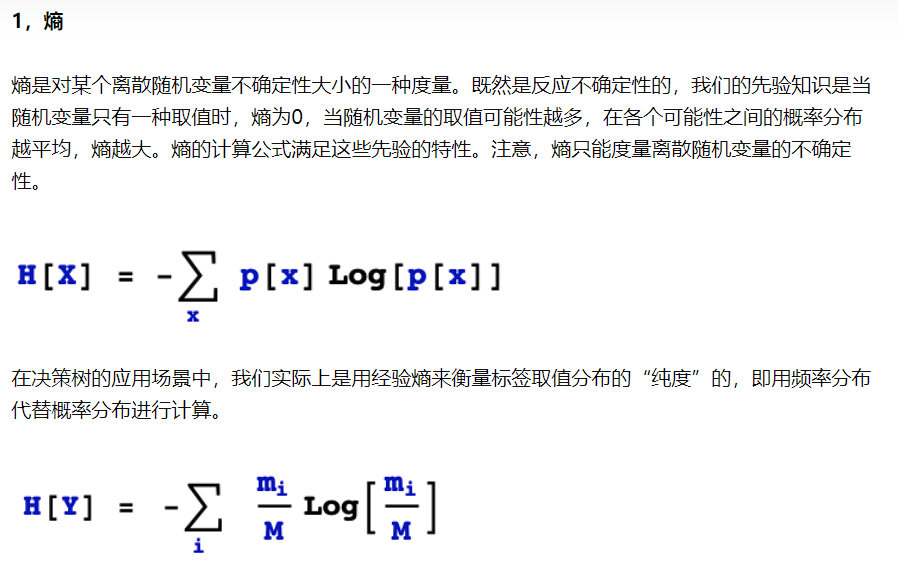
G(有房)

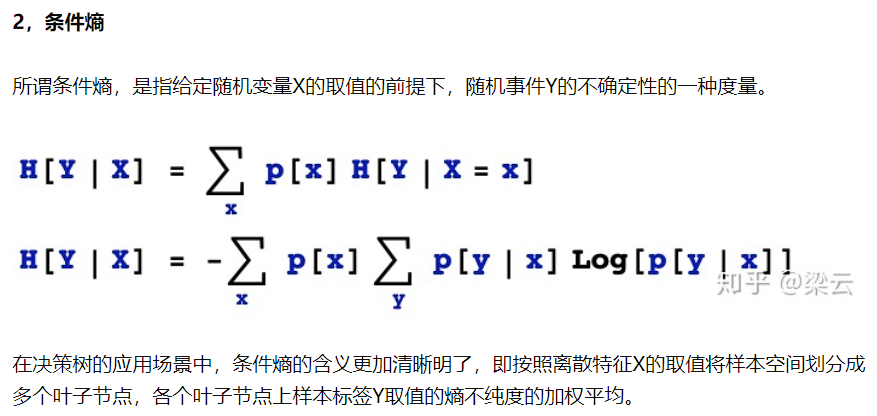
G(没房)

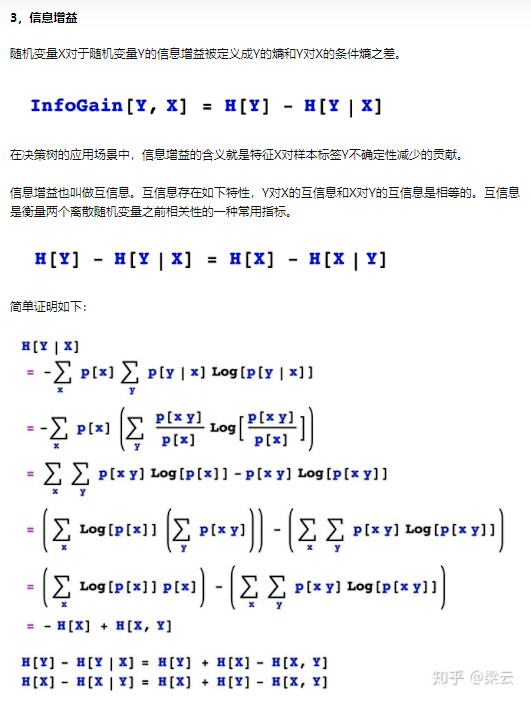
1. 递归

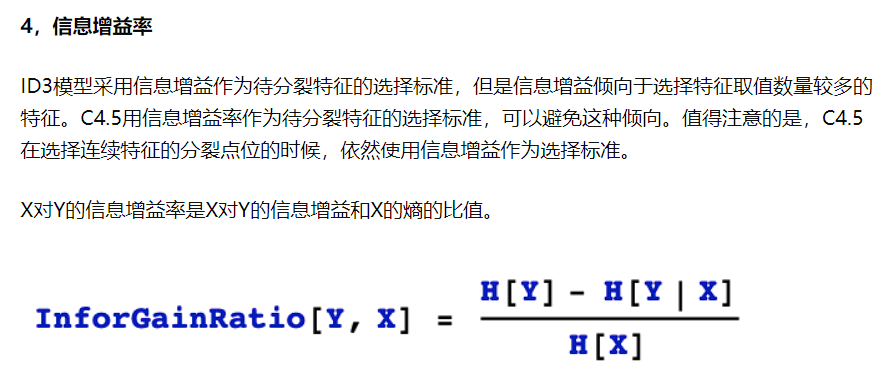


1. 熵，条件熵，信息增益，信息增益率

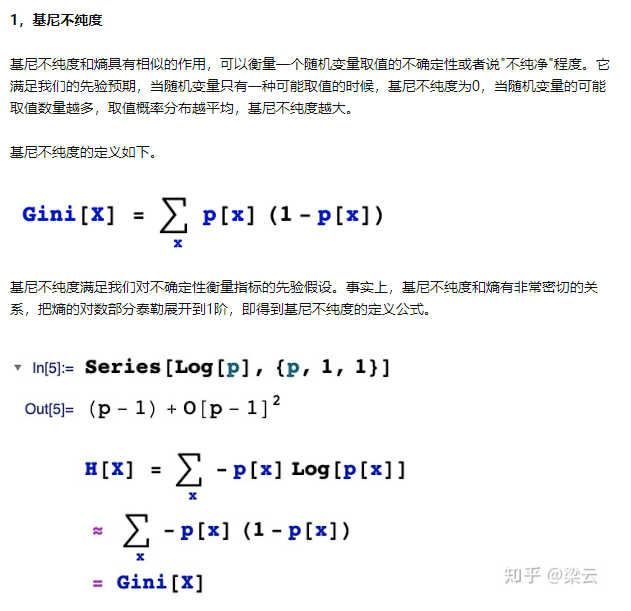


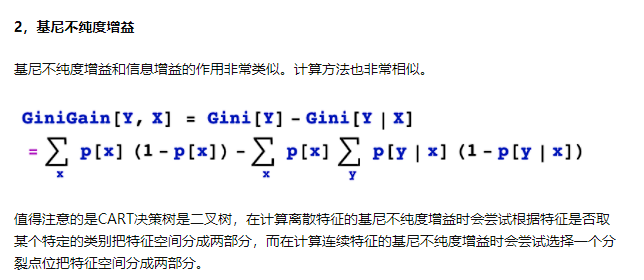






1. 基尼不纯度和基尼不纯度增益





1. 何时停止分裂？

每次分裂的Gini系数需要比不分裂的时候好

1. 前剪枝 vs 后剪枝

前剪枝是指在构造树的过程中就知道哪些节点可以剪掉 。 后剪枝是指构造出完整的决策树之后再来考查哪些子树可以剪掉

1. 后剪枝方法？

代价复杂性剪枝、最小误差剪枝、悲观误差剪枝

1. CPP[代价复杂度剪枝]

<https://zhuanlan.zhihu.com/p/368502685>

# 代码阶段



思考题

1. 后剪枝合适认为决策树已经合适不需要再后剪枝？