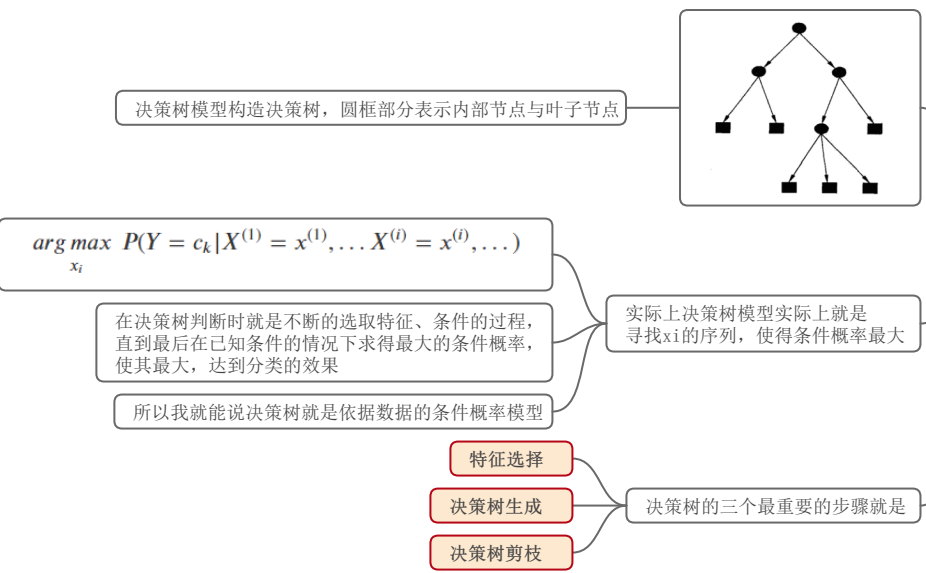
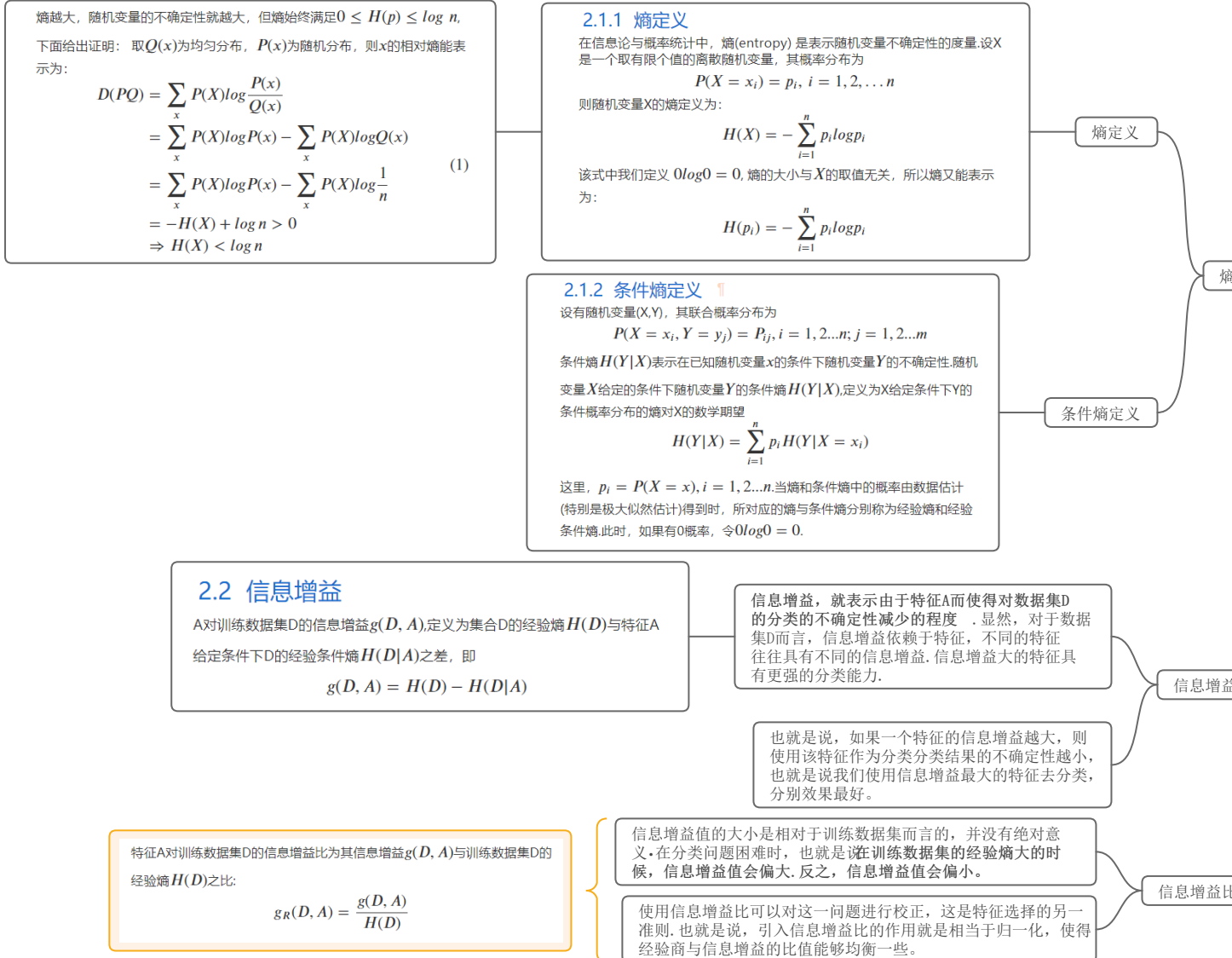


决策树

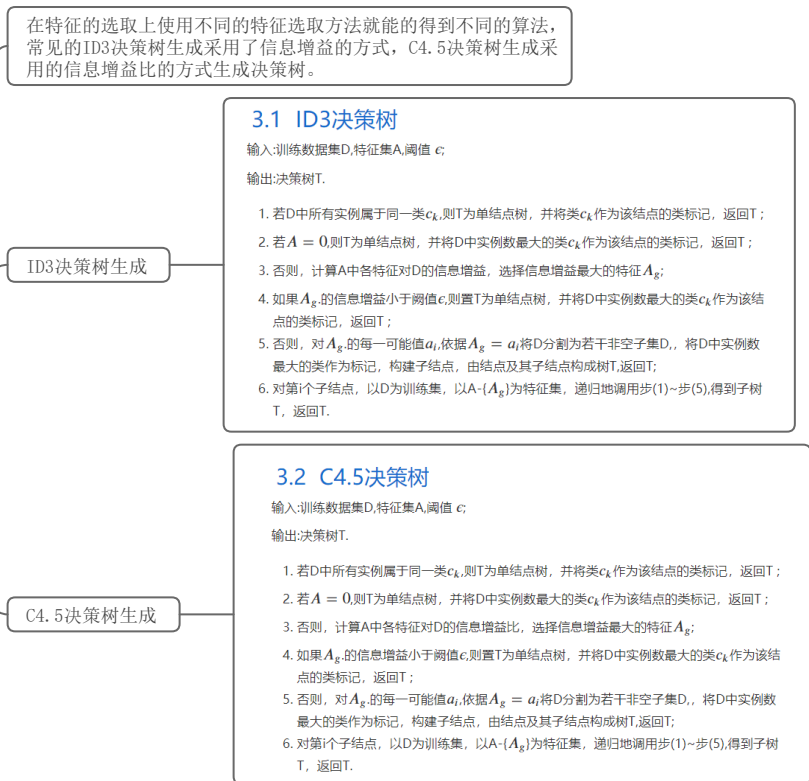
1 决策树模型



2 特征选择

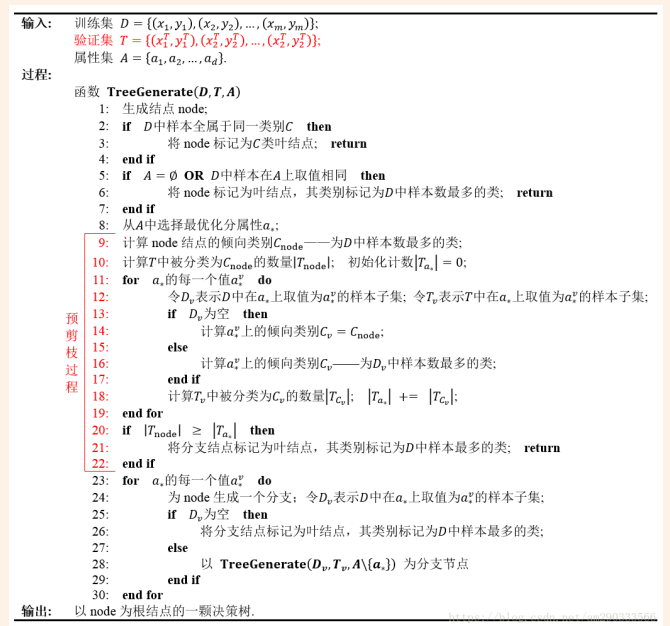


3 决策树生成



在以我们通常通过树搜索剪枝往往是想提高搜索速度，但是在决策树剪枝当中，不仅有这个作用，更大的作用是为了防止过拟合。通过无限划分的方式，让这个决策树在训练集上面有很好的分类效果，但是对未知数据，我们就不能够保证也有好的分类效果，也就是说，它的泛化能力比较差。就是我们要通过增加阈值、提前终止、剪枝等方式来提高决策树的泛化能力。

- 每一个结点所包含的最小样本数目，例如 10，则该结点总样本数小于 10 时，则不再分
- 指定树的高度或者深度，例如树的最大深度为 4；
- 指定结点的熵小于某个值，不再划分。



4 决策树剪枝



5 CART 决策树

