**熵：表示随机变量的不确定性。**

**条件熵：在一个条件下，随机变量的不确定性。**

**信息增益：熵 - 条件熵。表示在一个条件下，信息不确定性减少的程度。**

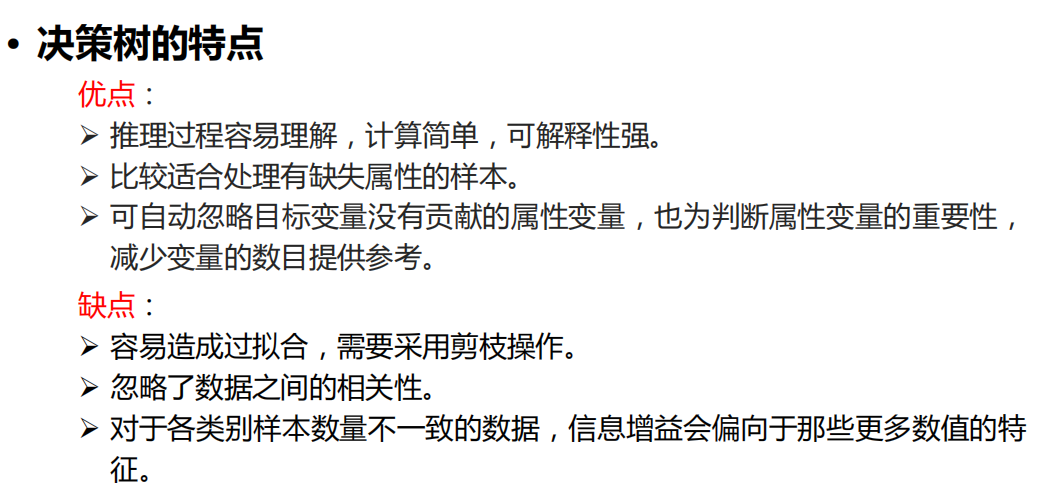
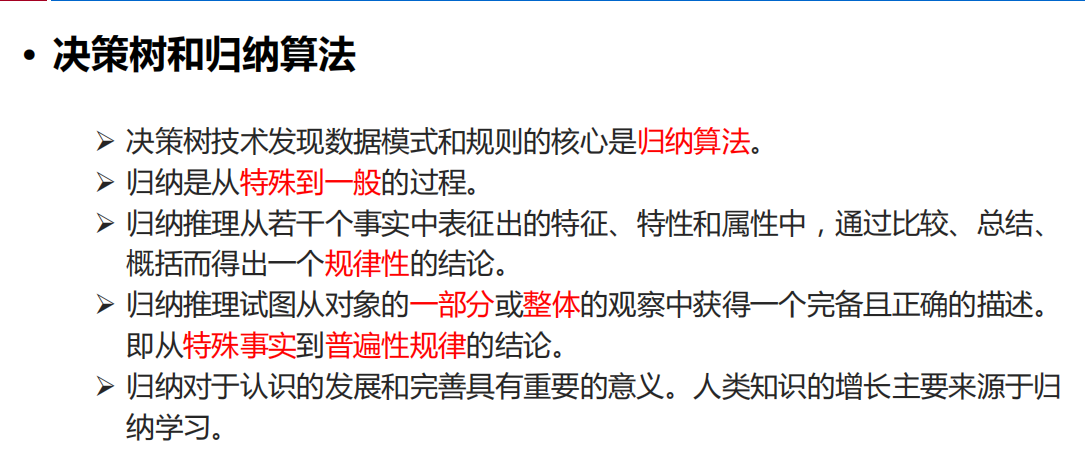
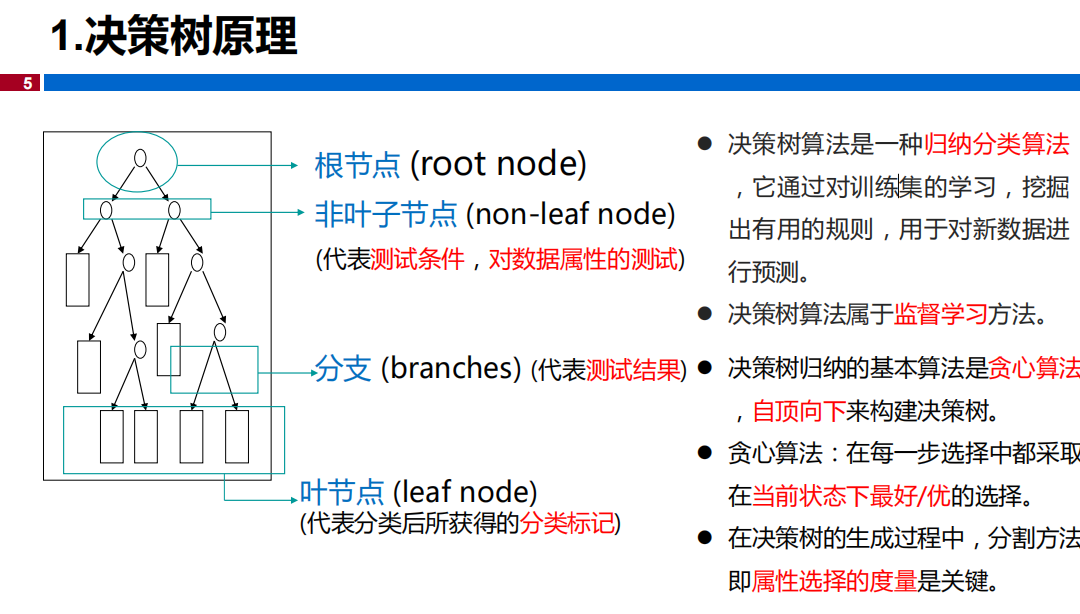
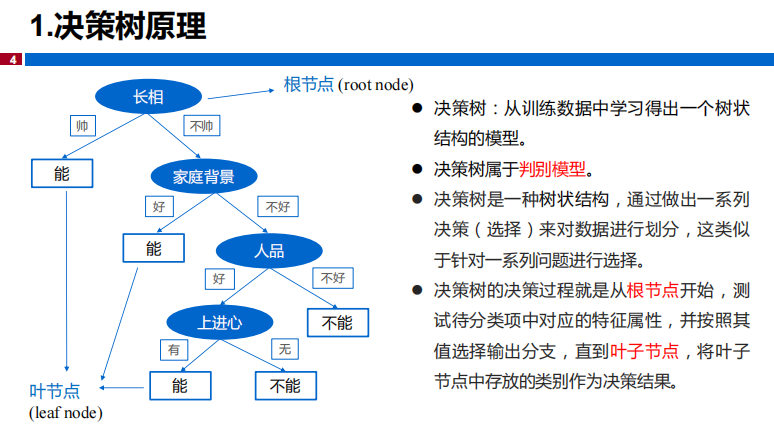
**通俗地讲，X(明天下雨)是一个随机变量，X的熵可以算出来， Y(明天阴天)也是随机变量，在阴天情况下下雨的信息熵我们如果也知道的话（此处需要知道其联合概率分布或是通过数据估计）即是条件熵。**

**X的熵减去Y条件下X的熵，就是信息增益。具体解释：原本明天下雨的信息熵是2，条件熵是0.01（因为如果知道明天是阴天，那么下雨的概率很大，信息量少），这样相减后为1.99。在获得阴天这个信息后，下雨信息不确定性减少了1.99，不确定减少了很多，所以信息增益大。也就是说，阴天这个信息对明天下午这一推断来说非常重要。**

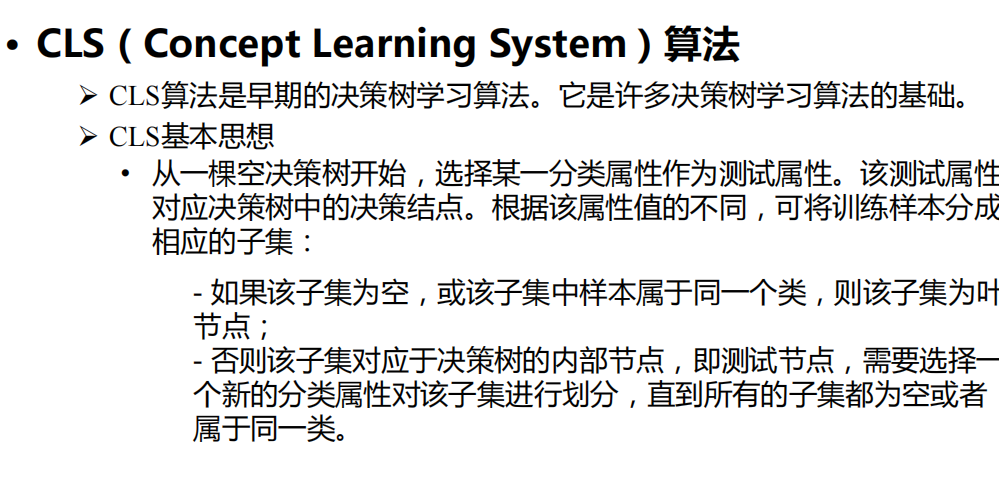
**所以在特征选择的时候常常用信息增益，如果IG（信息增益大）的话那么这个特征对于分类来说很关键，决策树就是这样来找特征的。**

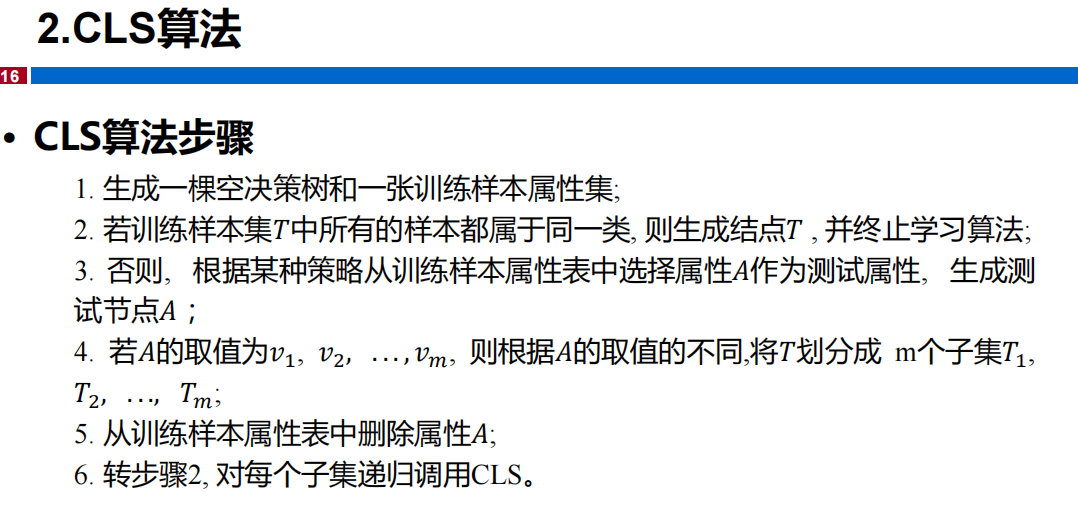
**相当于 加上了某个条件之后， 熵的值（不确定性）降低了很多，那么就代表这个条件跟结果的切合度很高，或者叫 只要知道了这个条件，那么更容易判断出结果。**

# **决策树原理**

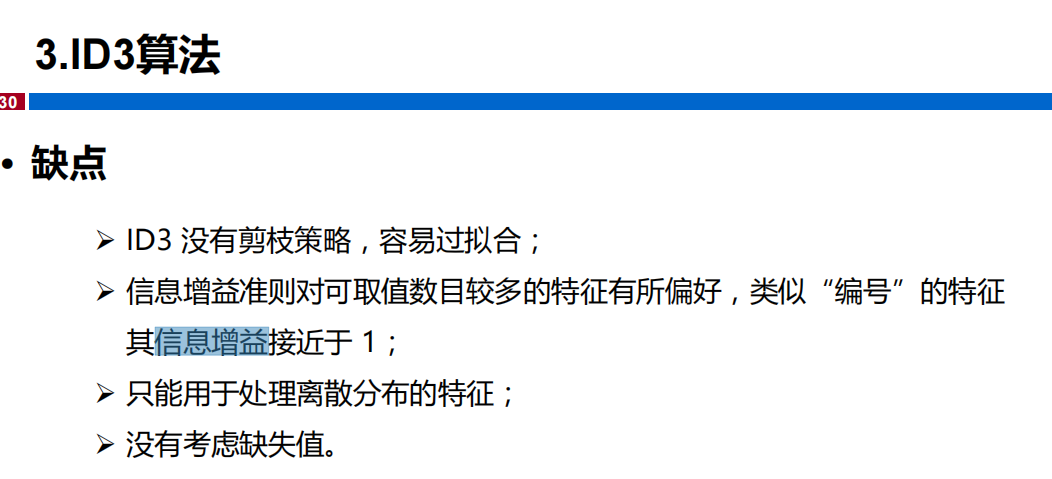
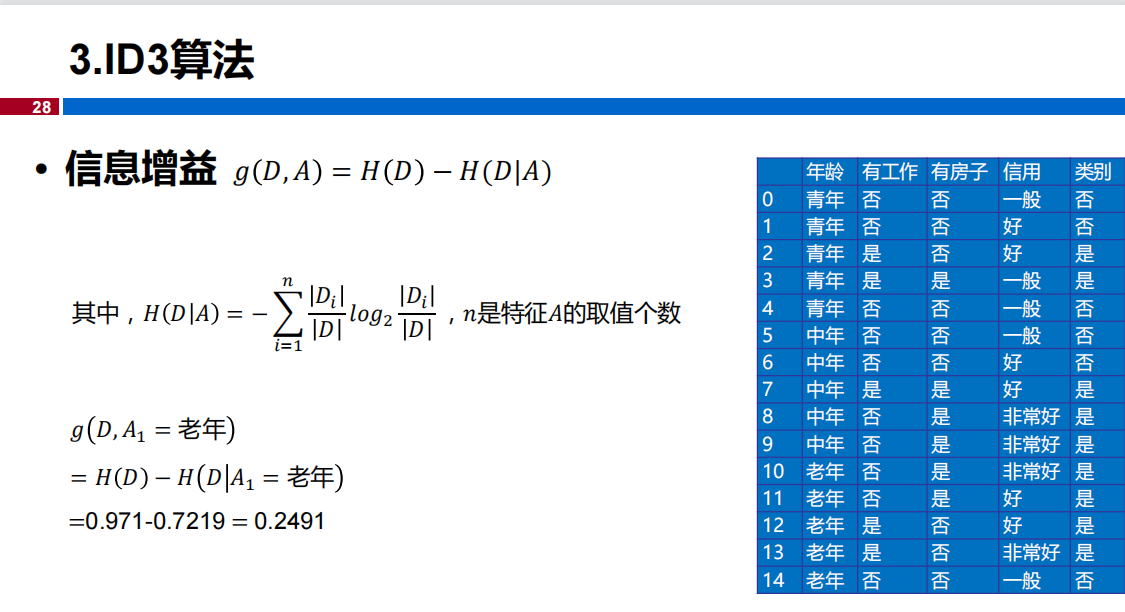
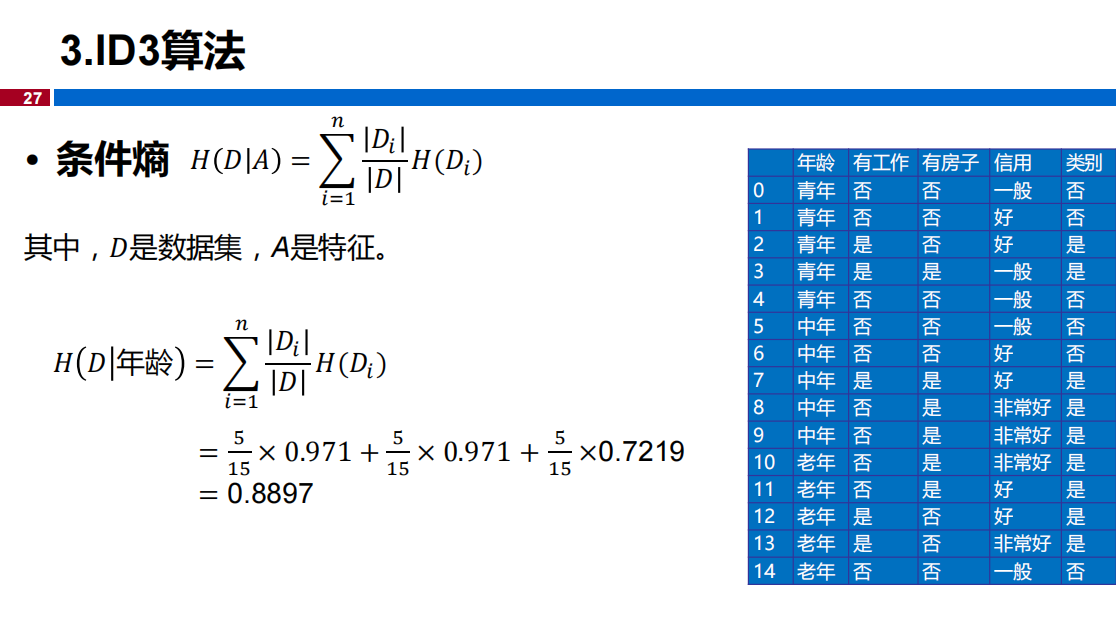
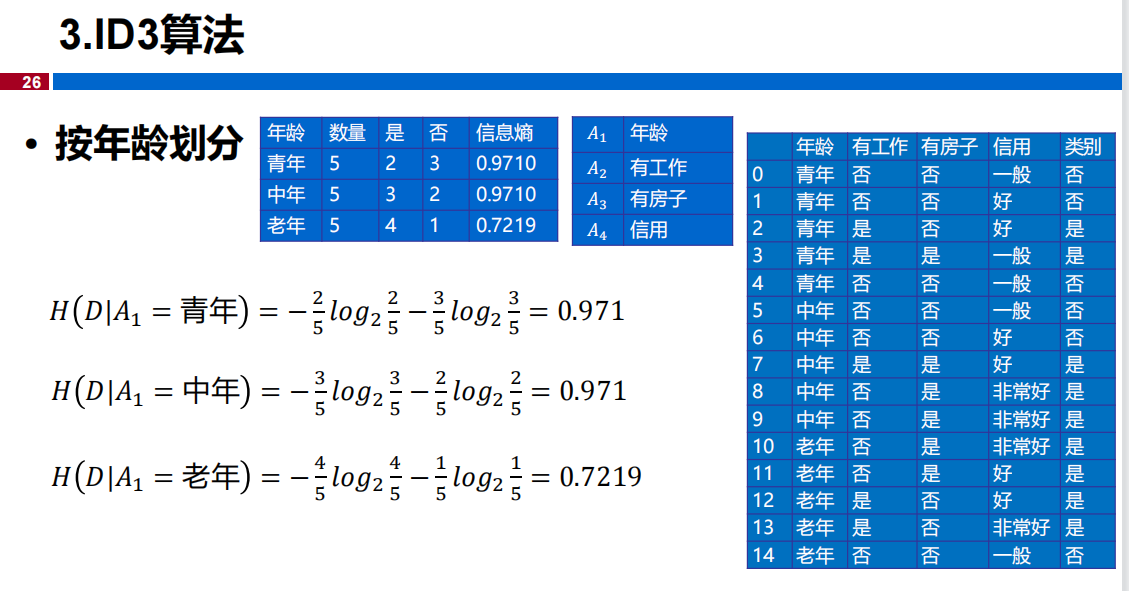
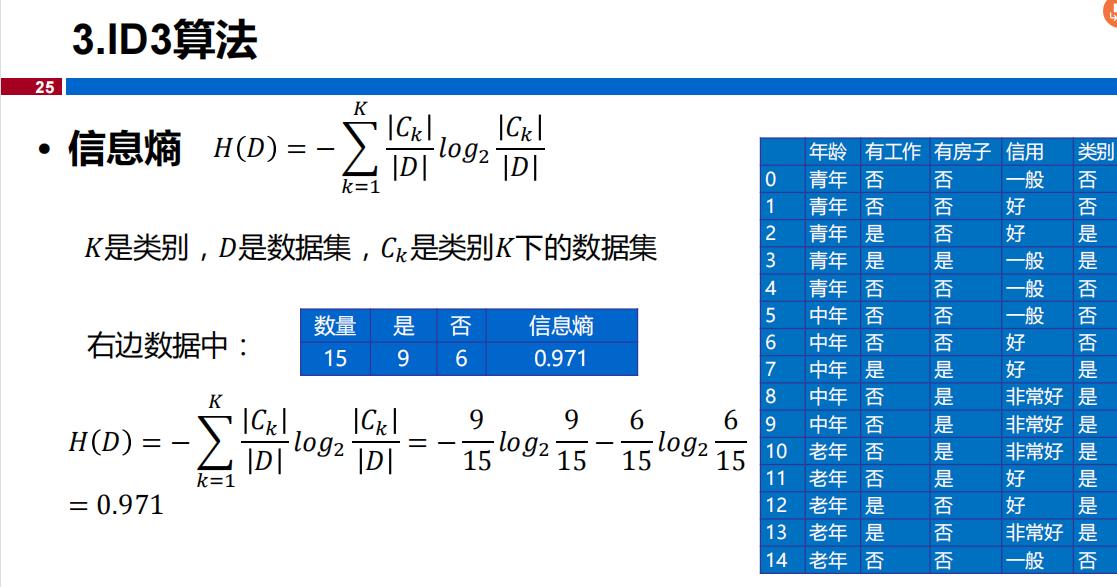
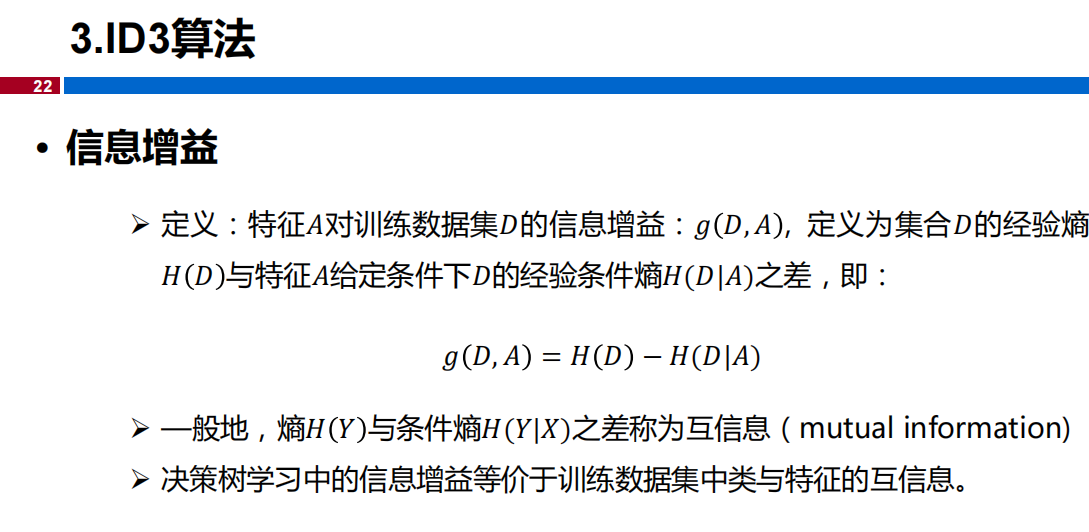
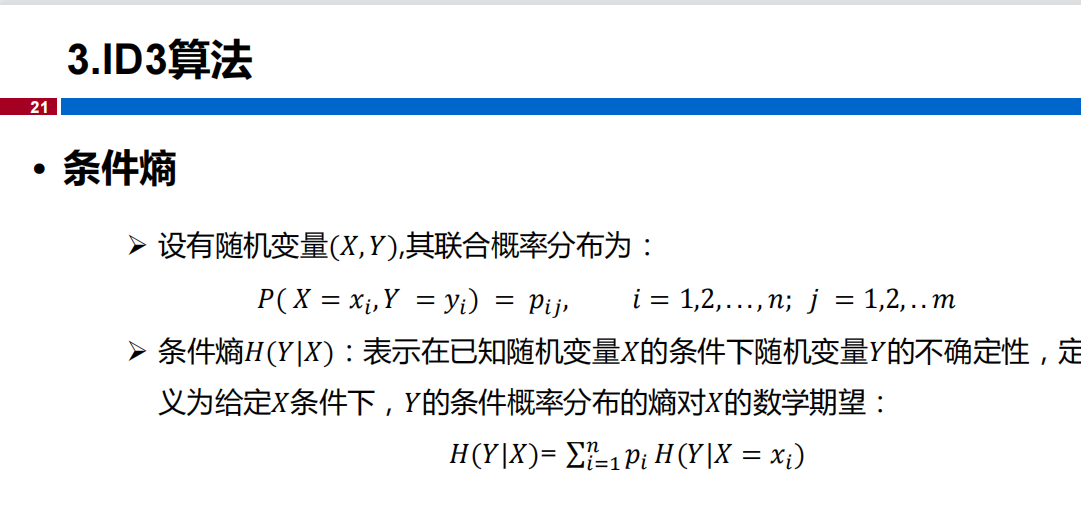
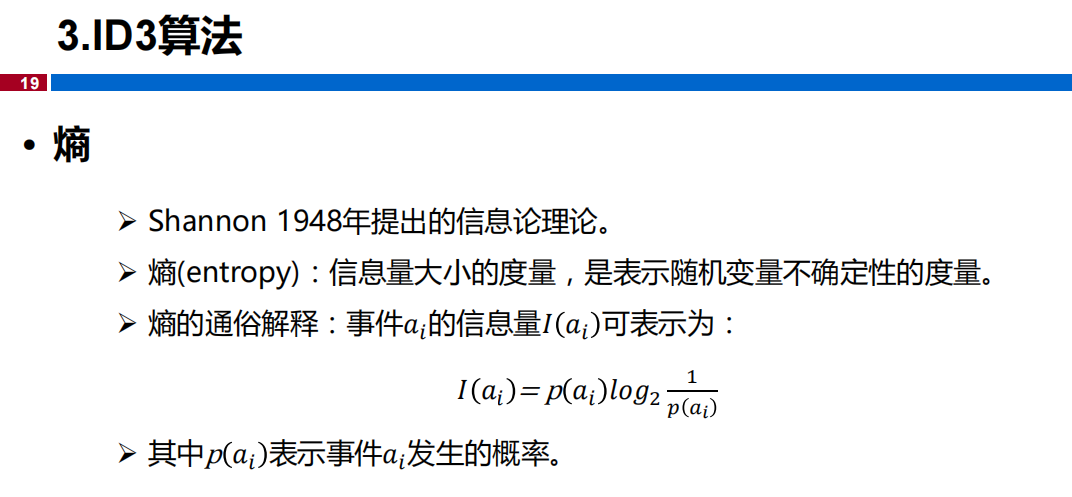
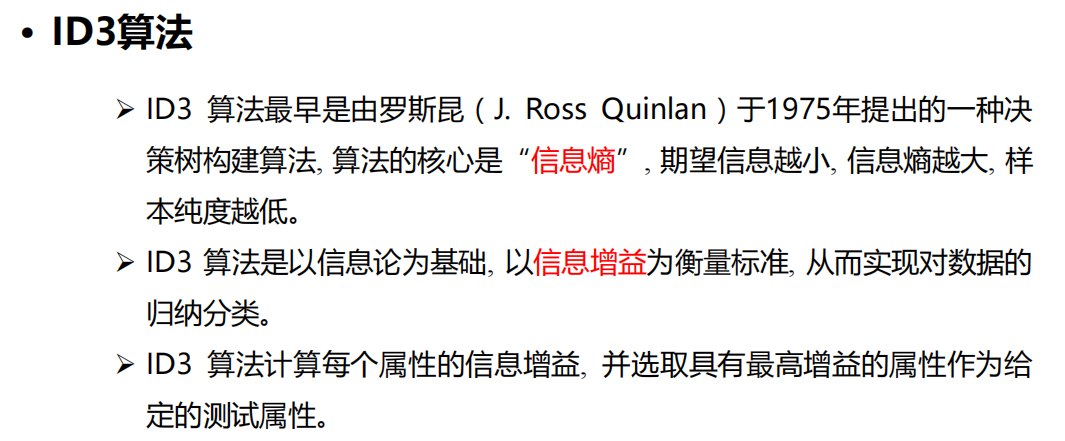


1. **CLS算法**



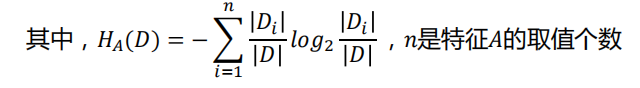
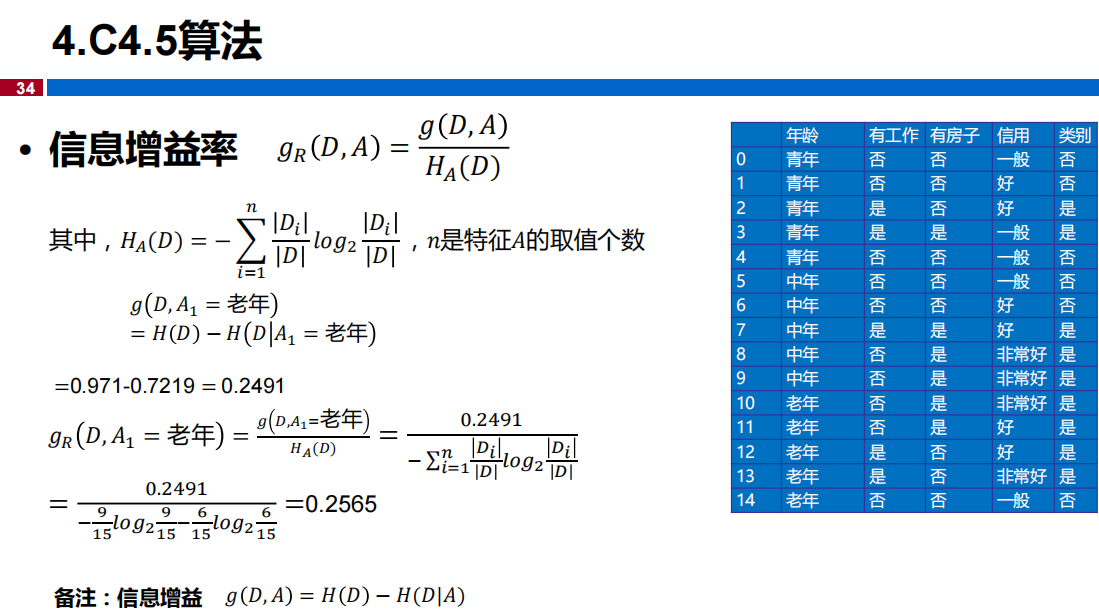
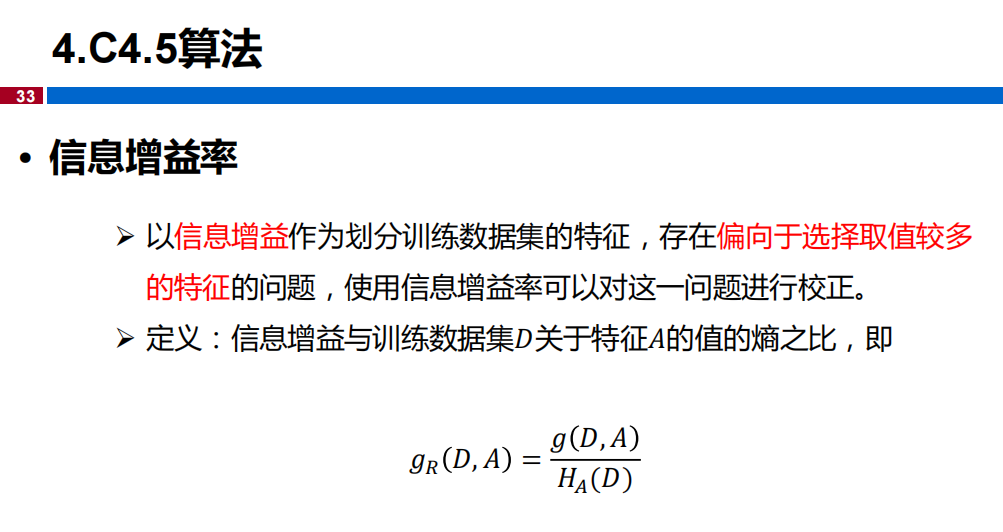
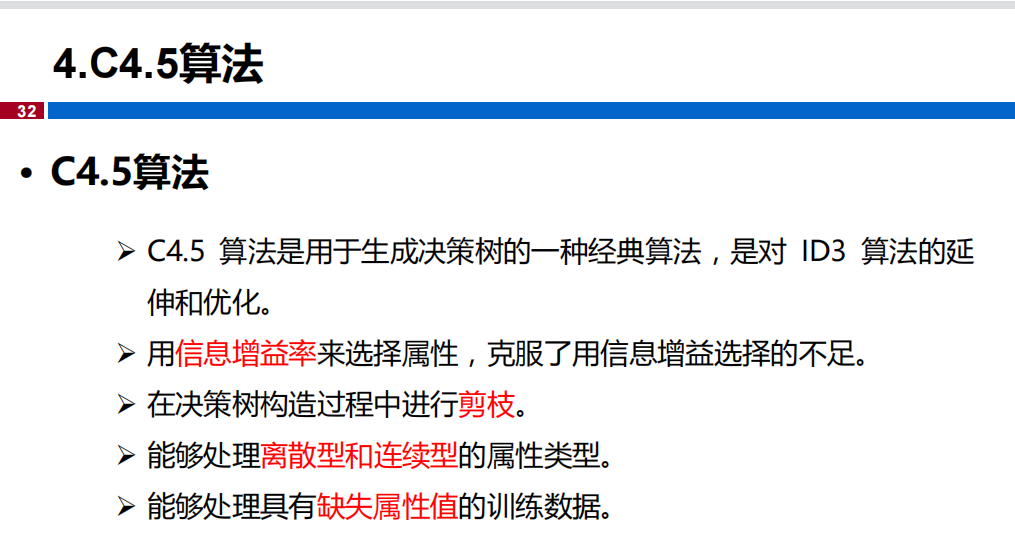


1. **ID3算法**

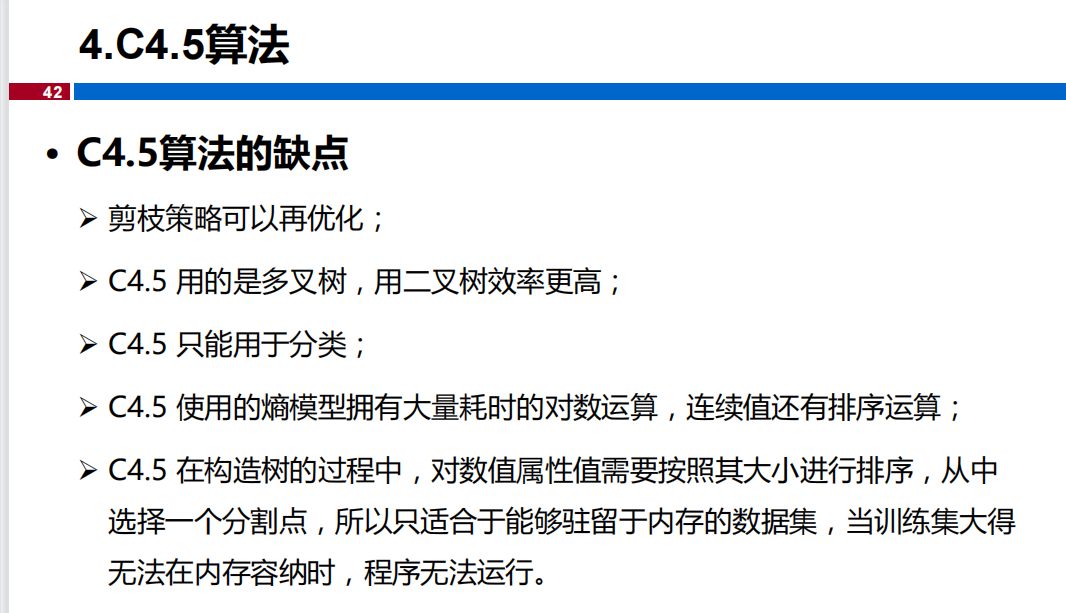
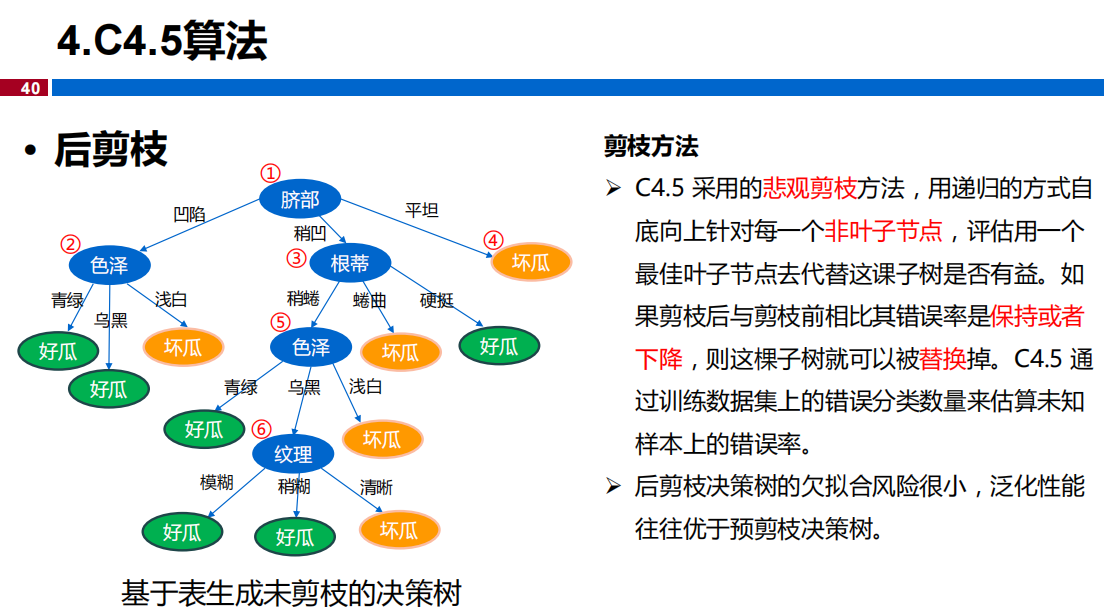
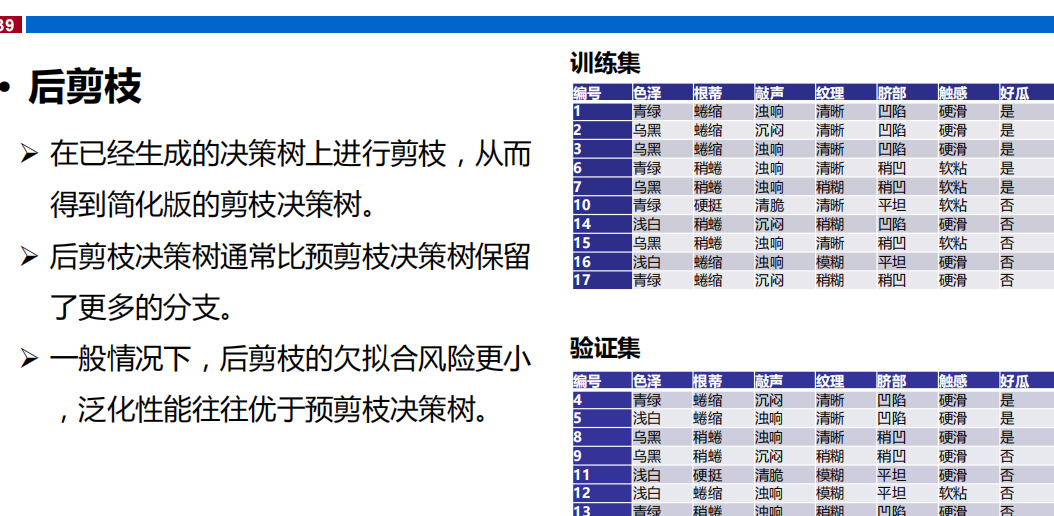
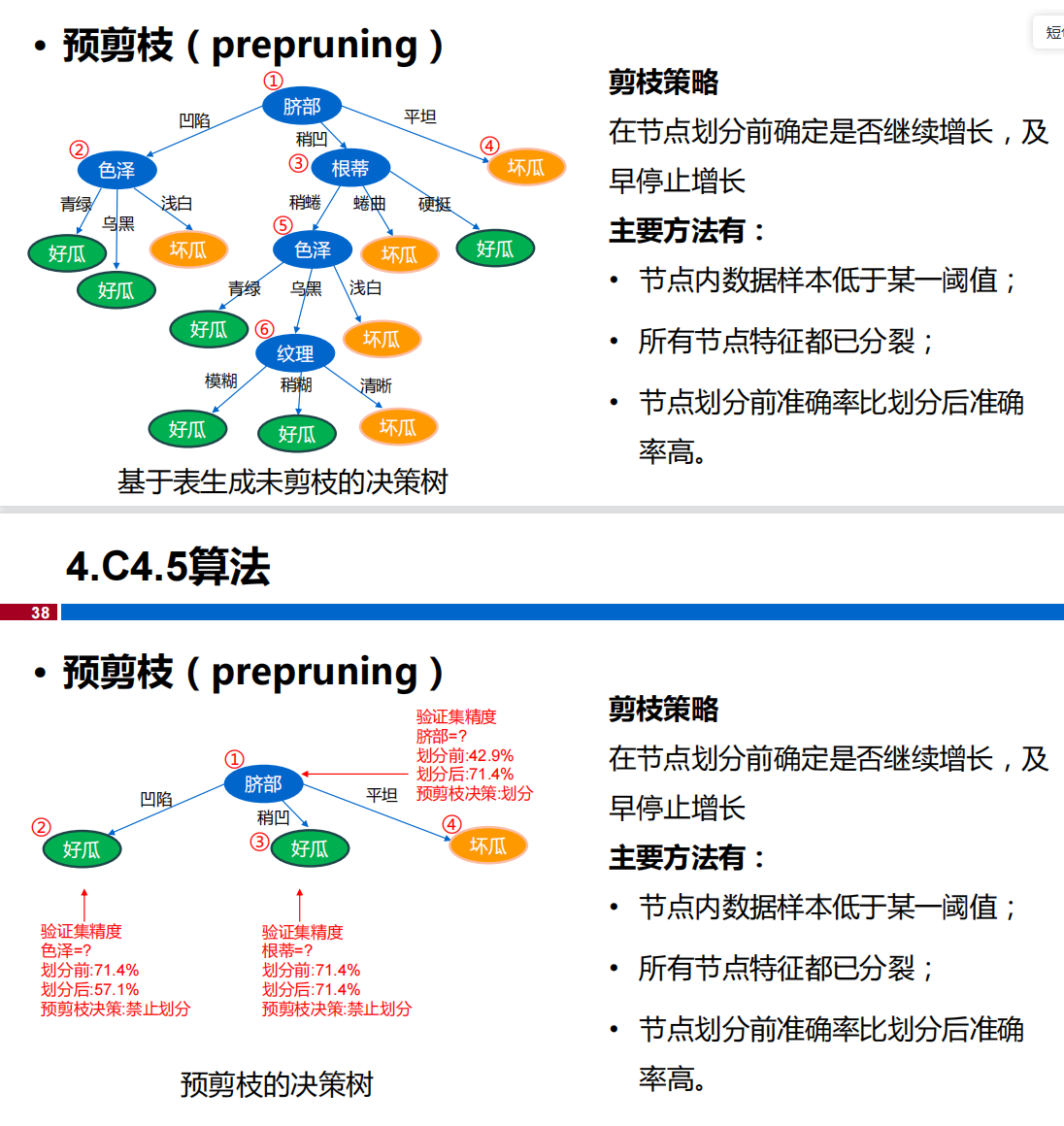


1. **C4.5算法**

C4.5 完美的尝试解决ID3的问题



在Di表示其中为Ai的个数 比如上题就应该为 三个三分之一 因为青年、中年、老年的数量一样。



1. **CART算法**

