决策树

假设是一个离散型随机变量，其概率分布函数为，则的信息量为：概率对数的负数，概率越小，信息量越大



信息熵：对于某个事件，有n种可能性，每一种可能性的概率为，则信息熵为所有信息量的期望：



联合熵：如果随机变量有多个，则信息熵就会变化成联合熵：



条件概率：



<https://www.cnblogs.com/fantasy01/p/4581803.html>

条件熵：为已知随机变量的条件下，随机变量的不确定性：



信息增益：待分类的集合的熵-按某个条件分类后的条件熵



在决策树算法中，分类效果越好，就会越小，相应的信息增益就会越大。