理解：
设信息量为信息总和
则：F(x,y) = F(x) + F(y) ， x, y 为互不相关两个变量
又：事件独立不相关；所以 p(x,y) = p(x)p(y)
对 p 取对数，得到 logp(x,y) = logp(x) + logp(y)
得到 F(x) = -logp(x)负号保证信息为正
所有信息的期望
即-p(x)logp(x)的总和就为 x 的熵

$$H(X)=-\sum_{x}p(x)logp(x)=-\sum_{i=1n}p(x_i)logp(x_i)$$

条件熵：

$$H(X|Y)=-\sum_{x,y}p(x,y)logp(x,y)=-\sum_{i=1n}\sum_{j=1m}p(x_i,y_i)logp(x_i,y_i)$$

表示 X 条件下，随机变量 Y 不确定性
信息增益：
问题(1)
给定条件下，信息熵-条件熵
信息增益率
加入了惩罚参数 HA(D)

$$惩罚参数 = \frac{1}{H_A(D)} = \frac{1}{-\sum_{i=1}^{n}\frac{|D_i|}{|D|}log_2\frac{|D_i|}{|D|}}$$

特征数越多，HA(D)越高，增益率越低
问题（2）
不会
问题（3）
min_samples_split , n_estimator , max_features ,max_depth
Q(4)
关联性越低越好

Python：
Padas
Get_dummies
join 主要用于基于索引的横向合并拼接；
merge 主要用于基于指定列的横向合并拼接；
concat 可用于横向和纵向合并拼接；
append 主要用于纵向追加；
combine 可以通过使用函数，把两个 DataFrame 按列进行组合。

groupby 过程拆解