

4.1 基本流程

概念

熵

宏观态的不确定性叫做熵。

e.g.

ABCD四个选项选择哪个很不确定

信息

消除不确定的事物

- 调整概率
- 排除干扰
- 确定情况

噪音

不能消除某人对某时间啊不确定性的事物（即信息量为0的事物）

数据

信息+噪音

熵如何量化

类似1kg的参照，熵参照一个不确定的事件作为单位，记一次抛硬币的不确定为1bit。

抛1次硬币，有2种情况

抛2次硬币，有4种情况

抛3次硬币，有8种情况

抛n次硬币，有 2^n 种情况

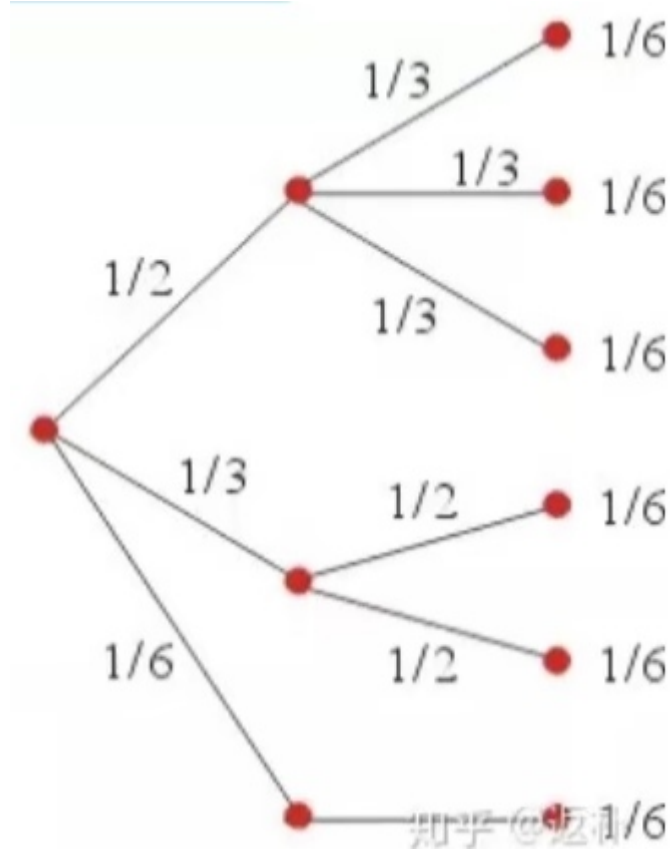
等概率分布

$n = \log_2 m$, 其中m为等概率不同情况的数量

每种情况概率不相等一般分布

公式: $Ent(D) = \sum_{k=1}^{|y|} p_k \log_2 \frac{1}{p^k} = - \sum_{k=1}^{|y|} p_k \log_2 p^k$

e.g.



$$\begin{aligned} & \frac{1}{2}(\log_2 6 - \log_2 3) + \frac{1}{3}(\log_2 6 - \log_2 2) + \frac{1}{6}(\log_2 6 - \log_2 1) \\ &= \frac{1}{2}\log_2 \frac{6}{3} + \frac{1}{3}\log_2 \frac{6}{2} + \frac{1}{6}\log_2 \frac{6}{1} \\ &= \frac{1}{2}\log_2 2 + \frac{1}{3}\log_2 3 + \frac{1}{6}\log_2 6 \end{aligned}$$

信息如何量化

得知信息前后熵的差额，就是信息量

e.g.

ABCD四个选项种每个选项的正确率为 $\frac{1}{4}$ ，则原始熵为 $4 * \frac{1}{4}\log_2 4 = 2bit$

若告知C是正确答案的肯性为 $\frac{1}{2}$

则现在的熵为 $3 * \frac{1}{6}\log_2 6 + \frac{1}{2}\log_2 2 = 1.79bit$

那么此条信息的信息量为 $2-1.79=0.21bit$