

Entropy:

x : 字符 (e.g., a, b, c, d)

X : 字符集

D : 码字集 (e.g., {0,1})

ℓ_x : 字符 x 被编码的长度

满足 nonprefix 要求:

$$\sum_{x \in X} |D|^{-\ell_x} \leq 1.$$

最小平均编码长度:

$$\min \sum_{x \in X} p_x \ell_x$$

$$\text{s.t. } \sum_{x \in X} |D|^{-\ell_x} \leq 1.$$

最优解 $\ell_x^* = \log_{|D|} p_x^{-1}$.

最短平均编码长度 $L^* = \sum_{x \in X} p_x \log_{|D|} p_x^{-1}$.

当 $|D|=2$ 时 记 L^* 为 Shannon Entropy. $H(X)$.

$$\sum_i x_i \ln x_i^{-1}$$

2021.11.12.

Cross Entropy

p : X 的真实分布

q : X 的估计分布

此时的期望编码长度为

$$\sum_i x_i \ln q_i^{-1}$$

2021.11.12.

$$H(p, q) = \sum_{x \in X} p_x \log q_x^{-1}.$$

称之为交叉熵.

KL-Divergence

$$\sum_i x_i \ln \frac{x_i}{y_i}$$

$$D_{KL}(p, q) \triangleq \sum_{x \in \mathcal{X}} p_x \log p_x - \sum_{x \in \mathcal{X}} p_x \log q_x$$

relative entropy

$$= H(p, q) - H(p)$$

Cross entropy, entropy.

In machine learning, $p(\text{Model})$, $p(\text{Data})$, $p(\text{Truth})$.

$$\text{minimize } D_{KL}(p(\text{Data}), p(\text{Model}))$$

$$= \text{minimize } H(p(\text{Data}), p(\text{Model})). \quad (H(\text{Data}) \text{ is fixed})$$

When Data is large enough, $p(\text{Data}) \approx p(\text{Truth})$ and thus
 $p(\text{Model}) \rightarrow p(\text{Truth})$.