Chapter 2

Xuan

2020 年 9 月 10 日

1　离散信源熵与互信息

1.1　信源熵

$$H(X) = \sum_i p(x_i) \log_2 \frac{1}{p(x_i)} = -\sum_i p(x_i) \log_2 p(x_i)$$

性质：

1. $H(X) \geq 0$
2. $H(X) \leq \log_2 |X|$

当某一符号 x_i 的概率为 p_i 为零时，$p_i \log_2 p_i$ 在熵公式中无意义，为此规定这时的 $p_i \log_2 p_i$ 为零。当信源 X 中只含一个符号 x 时，必定有 $p(x) = 1$，此时信源熵 $H(x)$ 为零，是确定信源。

1.1.1　证明 $H(x) \leq \log_2 |X|$:

方法一：　求偏导，略

方法二： 利用相对熵

假设 q 服从均匀分布，即 $q_i = \frac{1}{|X|}$

$$\begin{aligned} D(p||q) &= \sum_i p_i \log_2 \frac{p_i}{q_i} && (1) \\ &= \sum_i p_i \log_2 p_i - \sum_i p_i \log_2 q_i && (2) \\ &= -H(x) - \sum_i p_i \log_2 q_i \geq 0 && (3) \\ &= -H(x) + \sum_i p_i \log_2 |X| \geq 0 && (4) \end{aligned}$$

所以我们可以得到

$$H(x) \leq \log_2 |X|$$

1.2 相对熵

p 相对于 q 的相对熵定义为

$$D(p||q) = \sum_i p_i \log_2 \frac{p_i}{q_i}$$

相对熵也成为交叉熵或 Kullback-Leibler 距离 (KL 距离)。它满足两个要求：

1. 非负性
2. 当且仅当对所有 i,$p_i = q_i$ 时，相对熵为零

1.2.1 证明 $D(p||q) \geq 0$

下凸函数的性质 $pf(x_1) + (i-p)f(x_2) \geq f(px_1 + (1-p)x_2)$

Jensen's Inequality: 对于下凸函数而言，

$$Ef(X) \geq f(EX)$$

上凸函数反之。

1. $|X| = 2$ 时，由下凸函数性质可证

2. Jensen's Inequality

下面先证 Jensen's Inequality：

Assume $|X| = k-1$, Jensen's Inequality holds

Prove $|X| = k$, Inequality holds as well.

Assume $\sum_{i=1}^{k-1} p(x_i) = q = 1 - p(x_k)$

$$\begin{aligned} Ef(X) &= p(x-k)f(x_k) + \sum_{i=1}^{k-1} p(x_i)f(x_i) && (5)\\ &= p(x_k)f(x_k) + q\sum_{i=1}^{k-1}\frac{p(x_i)}{q}f(x_i) && (6)\\ &= p(x_k)f(x_k) + (1-p(x_k))f(\sum_{i=1}^{k-1}\frac{p(x_i)}{1-p(x_k)}x_i) && (7)\\ &\geq f(p(x_k)x_k + \sum_{i=1}^{k-1} p(x_i)x_i) && (8)\\ &= f(EX) && (9)\end{aligned}$$

下面证 $D(p||q) \geq 0$

$$\begin{aligned} -D(p||q) &= -\sum_{x\in A} p(x)\log_2\frac{p(x)}{q(x)} && (10)\\ &= \sum_{x\in A} p(x)\log_2\frac{q(x)}{p(x)} && (11)\\ &\leq \log_2\sum_{x\in A} p(x)\frac{q(x)}{p(x)} && (12)\\ &= \log_2\sum_{x\in A} q(x) && (13)\\ &\leq \log_2\sum_{x\in Z} q(x) && (14)\\ &= \log_2 1 && (15)\\ &= 0 && (16)\end{aligned}$$

1.3 联合熵、条件熵

条件熵

$$
\begin{aligned}
H(X|Y) &= \sum_j p(y_j)H(X|y_j) && (17)\\
&= \sum_{ij} p(y_j)p(x_i|y_j)I(x_i|y_j) && (18)\\
&= \sum_{ij} p(x_i,y_j)\log_2 p(x_i|y_j) && (19)
\end{aligned}
$$

联合熵

$$H(X,Y) = \sum_{ij} p(x_i,y_j)I(x_i,y_j) = -\sum p(x_i,y_j)\log_2 p(x_i,y_j)$$

$$H(X|Y) = \sum p(y)H(X|y)$$

1.3.1 证明 $H(X|Y) = H(X,Y) - H(Y)$

$$\begin{aligned}
H(X|Y) &= \sum p(x,y)\log_2\frac{1}{p(x|y)} && (20)\\
H(X,Y)-H(Y) &= \sum p(x,y)\log_2\frac{1}{p(x,y)} - \sum p(y)\log_2\frac{1}{p(y)} && (21)\\
&= \sum p(y)\sum p(x|y)\log_2\frac{1}{p(x,y)} - \sum p(y)\log_2\frac{1}{p(y)} && (22)\\
&= \sum p(y)(\sum p(x|y)\log_2\frac{1}{p(x,y)} - \log_2\frac{1}{p(y)}) && (23)\\
&= \sum p(y)(\sum p(x|y)\log_2\frac{1}{p(x,y)} - \sum p(x|y)\log_2\frac{1}{p(y)}) && (24)\\
&= \sum p(y)\sum p(x|y)\log_2\frac{p(y)}{p(x,y)} && (25)\\
&= \sum p(y)\sum p(x|y)\log_2\frac{1}{p(x|y)} && (26)\\
&= \sum p(x,y)\log_2\frac{1}{p(x|y)} && (27)\\
&= H(X|Y) && (28)
\end{aligned}$$

1.4 互信息

$$\begin{aligned}
I(X;Y) &= \sum_{i,j} p(x_i,y_j)\log_2\frac{p(x_i|y_j)}{p(x_i)} && (29)\\
&= \sum_{i,j} p(x_i,y_j)\log_2\frac{p(x_i,y_j)}{p(x_i)p(y_j)} && (30)\\
&= \sum_{i,j} p(x_i,y_j)\log_2\frac{p(y_j|x_i)}{p(y_j)} && (31)\\
&= I(Y;X) && (32)
\end{aligned}$$

互信息的一些性质：

1. X || Y 时，$I(X;Y) = 0$

2. X = Y 时，$I(X;Y) = H(X) = H(Y)$

3. X = f(Y) 时，$I(X;Y) = H(X)$

互信息有如下表达形式：

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) && (33) \\ &= H(X) - H(X|Y) && (34) \\ &= H(Y) - H(Y|X) && (35) \end{aligned}$$

证明 $I(X;Y) = H(X) + H(Y) - H(X,Y)$

有如下公式成立：

$$p(x_i) = \sum_{y \in Y} p(x,y)$$

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) && (36) \\ &= \sum p(x)\log_2 \frac{1}{p(x)} + \sum p(y)\log_2 \frac{1}{p(y)} - \sum p(x,y)\log_2 \frac{1}{p(x,y)} && (37) \\ &= \sum p(x,y)\log_2 \frac{1}{p(x)} + \sum p(x,y)\log_2 \frac{1}{p(y)} - \sum p(x,y)\log_2 \frac{1}{p(x,y)} && (38) \\ &= \sum p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)} && (39) \end{aligned}$$

证明 $I(X;Y) = H(X) - H(X|Y)$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (40) \\ &= \sum p(x)\log_2 \frac{1}{p(x)} - \sum p(x,y)\log_2 \frac{1}{p(x|y)} && (41) \\ &= \sum p(x,y)\log_2 \frac{p(x|y)}{p(x)} && (42) \end{aligned}$$

1.5 Chain Rule

二元的情况下有 $H(X,Y) = H(X|Y) + H(Y)$

对于三元的情况 $H(X_1, X_2, X_3)$，设 $(X_2, X_3) = Z$，则有

$$\begin{aligned} H(X_1, X_2, X_3) &= H(X_1, Z) && (43) \\ &= H(X_1|Z) + H(Z) && (44) \\ &= H(X_1|X_2, X_3) + H(X_2|H_3) + H(X_3) && (45) \end{aligned}$$

推广到 n 元的情况下，得到 $H(X_1, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i+1}, ..., X_n)$

1.6 Conditional Matual Information