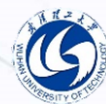


联合熵与条件熵

武汉理工大学

Information theory
and
coding



武汉理工大学

联合熵和条件熵

联合熵：联合自信息量的统计平均。

条件熵：条件自信息量的统计平均。

各类熵之间的关系：与各类自信息量之间的关系对应。

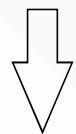
联合熵

设联合概率空间为

$$[XY, P_{XY}] = [(x_k, y_j), P(x_k, y_j) | k = 1, 2, \dots, K; j = 1, 2, \dots, J]$$

联合符号 (x_k, y_j) 的先验不确定性称为联合自信息量：

$$I(x_k, y_j) = \log \frac{1}{P(x_k, y_j)} \quad k = 1, 2, \dots, K; j = 1, 2, \dots, J$$



统计平均

联合熵

$$\begin{aligned} H(XY) &= \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) \log \frac{1}{P(x_k, y_j)} \\ &= - \sum_{k,j} P(x_k, y_j) \log P(x_k, y_j) \end{aligned}$$

熵 $H(XY)$ 的物理意义: 信源 的平均不确定性。

条件熵

设联合概率空间为

$$[XY, P_{XY}] = [(x_k, y_j), P(x_k, y_j) | k = 1, 2, \dots, K; j = 1, 2, \dots, J]$$

条件自信息量：

$$I(x_k | y_j) = \log \frac{1}{P(x_k | y_j)} \quad k = 1, 2, \dots, K; j = 1, 2, \dots, J$$

↓ 统计平均

条件熵

$$\begin{aligned} H(X | Y) &= \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) I(x_k | y_j) \\ &= \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) \log \frac{1}{P(x_k | y_j)} \\ &= - \sum_{k,j} P(x_k, y_j) \log P(x_k | y_j) \end{aligned}$$

条件熵（续一）

$$\begin{aligned} H(X | Y) &= \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) \log \frac{1}{P(x_k | y_j)} \\ &= \sum_{k=1}^K \sum_{j=1}^J P(y_j) P(x_k | y_j) \log \frac{1}{P(x_k | y_j)} \\ &= \sum_{j=1}^J P(y_j) \left[\sum_{k=1}^K P(x_k | y_j) \log \frac{1}{P(x_k | y_j)} \right] \\ &= \sum_{j=1}^J P(y_j) H(X | Y = y_j) \end{aligned}$$

式中 $H(X | Y = y_j) = \sum_{k=1}^K P(x_k | y_j) \log \frac{1}{P(x_k | y_j)}$

解释： $H(X | Y = y_j)$ 是另一种条件熵，它只对 $Y = y_j$ 求了统计平均，而未对 X 求统计平均，代表在给定条件 $Y = y_j$ 下有关 X 的（平均）不确定性。

条件熵（续二）

$$H(X | Y) = \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) \log \frac{1}{P(x_k | y_j)} = \sum_{j=1}^J P(y_j) H(X | Y = y_j)$$

$$H(X | Y = y_j) = \sum_{k=1}^K P(x_k | y_j) \log \frac{1}{P(x_k | y_j)}$$

$$H(Y | X) = \sum_{k=1}^K \sum_{j=1}^J P(x_k, y_j) \log \frac{1}{P(y_j | x_k)} = \sum_{k=1}^K P(x_k) H(Y | X = x_k)$$

$$H(Y | X = x_k) = \sum_{j=1}^J P(y_j | x_k) \log \frac{1}{P(y_j | x_k)}$$

》》》 各类熵之间的关系

$$I(x_k, y_j) = I(x_k) + I(y_j | x_k) = I(y_j) + I(x_k | y_j)$$



$$\begin{aligned} H(XY) &= \sum_{k,j} P(x_k, y_j) I(x_k, y_j) \\ &= \sum_{k,j} P(x_k, y_j) I(x_k) + \sum_{k,j} P(x_k, y_j) I(y_j | x_k) \\ &= \sum_k \left[I(x_k) \sum_j P(x_k, y_j) \right] + H(Y | X) \\ &= \sum_k [I(x_k) P(x_k)] + H(Y | X) = H(X) + H(Y | X) \end{aligned}$$

$$\sum_j P(x_k, y_j) = P(x_k)$$

同理 $H(XY) = H(Y) + H(X | Y)$

总之 $H(XY) = H(X) + H(Y | X) = H(Y) + H(X | Y)$

熵的强可加性

推广 $H(X_1 X_2 \cdots X_N) = H(X_1) + H(X_2 | X_1) + H(X_3 | X_1 X_2) + \cdots + H(X_N | X_1 \cdots X_{N-1})$

》》》 各类熵之间的关系（续）

当 X 与 Y 相互独立，则

$$P(x_k | y_j) = P(x_k) \quad P(y_j | x_k) = P(y_j) \quad P(x_k, y_j) = P(x_k)P(y_j)$$

于是 $I(x_k | y_j) = I(x_k) \quad I(y_j | x_k) = I(y_j)$

$$I(x_k, y_j) = I(x_k) + I(y_j)$$

因此，熵之间的关系简化：

$$H(X | Y) = H(X) \quad H(Y | X) = H(Y)$$

$$H(XY) = H(X) + H(Y)$$

熵的可加性

推广： $\{X_1, X_2, \dots, X_N\}$ 统计独立时，有

$$H(X_1 X_2 \cdots X_N) = H(X_1) + H(X_2) + \cdots + H(X_N)$$

感谢观看!

Information theory
and
coding



武汉理工大学