Взаимная информация и информация о случайных величинах

Информация в сообщениях и взаимная информация

- Пришло сообщение о значении случайной величины Y. Какая тогда информация в конкретном сообщении о случайной величине X?

  1. $J(\mathcal{X} = x_i, \mathcal{Y} = y_j) = -\log P(\mathcal{X} = x_i, \mathcal{Y} = y_j)$

- Информация о значении случайной величины X, без сообщения о Y:

  2. $J(\mathcal{X} = x_i) = -\log P(\mathcal{X} = x_i)$

- Разность (2) – (1):

$$-\log P(\mathcal{X} = x_i) + \log P(\mathcal{X} = x_i, \mathcal{Y} = y_j)$$

- Эта разность обозначается как:

$$J(\mathcal{X} = x_i \leftarrow \mathcal{Y} = y_j)$$

  и представляет собой количество информации в сообщении $\mathcal{Y} = y_j$ о том, что $\mathcal{X} = x_i$.

Преобразуем выражение

$$J(A \leftarrow B) = -\log P(A) + \log P(A|B)$$

Так как:

$$P(A|B) = \frac{P(AB)}{P(B)}$$

То:

$$J(A \leftarrow B) = -\log P(A) - \log P(B) + \log P(AB)$$

Таким образом:

$$J(B \leftarrow A) = -\log P(B) - \log P(A) + \log P(AB)$$

$$J(A \leftarrow B) = J(B \leftarrow A) \Rightarrow J(A \Leftrightarrow B) = \log\left(\frac{P(AB)}{P(A) \cdot P(B)}\right)$$

- Видим, что $J(\mathcal{X} = x_i \Leftrightarrow \mathcal{Y} = y_j)$ — это функция от пары (X, Y), то есть тоже случайная величина.
- Используем двойную стрелку, так как $J(X \leftarrow Y) = J(Y \leftarrow X)$.

Определение (Взаимная информация)

$$I(X \Leftrightarrow Y) = \mathbb{E}[J(\mathcal{X} = x_i \Leftrightarrow \mathcal{Y} = y_j)]$$

— взаимная информация между случайными величинами X и Y.

- Это среднее количество информации, которое мы получаем в сообщении о значении одной случайной величины из сообщения о значении другой.

Теорема (о взаимной информации)

1. $0 \leq I(X \Leftrightarrow Y)$
2. $I(X \Leftrightarrow Y) = H(X) + H(Y) - H(X, Y)$

Доказательство (вторая часть теоремы)

$$I(X \Leftrightarrow Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \log\left(\frac{P(x_i, y_j)}{P(x_i) \cdot P(y_j)}\right) \cdot P(x_i, y_j)$$

Разделим логарифм:

$$= \sum_{i,j} (\log P(x_i, y_j) - \log P(x_i) - \log P(y_j)) \cdot P(x_i, y_j)$$

Распишем по частям:

- $-\sum_{i,j} \log P(x_i, y_j) \cdot P(x_i, y_j) = H(X, Y)$
- $-\sum_{i,j} \log P(x_i) \cdot P(x_i, y_j) = -\sum_{i} \log P(x_i) \cdot P(x_i) = H(X)$
- $-\sum_{i,j} \log P(y_j) \cdot P(x_i, y_j) = -\sum_{j} \log P(y_j) \cdot P(y_j) = H(Y)$

Итак:

$$I(X \Leftrightarrow Y) = H(X) + H(Y) - H(X, Y) \qquad \blacksquare$$

Доказательство (первой части теоремы)

$$-I(X \Leftrightarrow Y) = -\sum_{i,j} \log\left(\frac{P(x_i, y_j)}{P(x_i) \cdot P(y_j)}\right) \cdot P(x_i, y_j)$$

$$= \sum_{i,j} \log\left(\frac{P(x_i) \cdot P(y_j)}{P(x_i, y_j)}\right) \cdot P(x_i, y_j)$$

$$= \log(e) \cdot \sum_{i,j} \ln\left(\frac{P(x_i) \cdot P(y_j)}{P(x_i, y_j)}\right) \cdot P(x_i, y_j)$$

Используем неравенство: $\ln x \leq x - 1$:

$$\leq \log(e) \cdot \sum_{i,j} \left(\frac{P(x_i) \cdot P(y_j)}{P(x_i, y_j)} - 1\right) \cdot P(x_i, y_j)$$

Раскрываем скобки:

$$\log(e) \cdot \left(\sum_{i,j} P(x_i)P(y_j) - \sum_{i,j} P(x_i, y_j)\right) = \log(e) \cdot (1 - 1) = 0$$

$$\Rightarrow -I(X \Leftrightarrow Y) \leq 0 \Rightarrow I(X \Leftrightarrow Y) \geq 0 \qquad \blacksquare$$

Следствие

Если X, Y — независимы, то:

$$P(x_i, y_j) = P(x_i) \cdot P(y_j) \Rightarrow I(X \Leftrightarrow Y) = 0$$