CHƯƠNG 2: ENTROPY & MUTUAL INFORMATION

TS. TRỊNH VĂN CHIẾN (SOICT-HUST)

TÍNH KHÔNG CHẮC CHẮN

- Xem xét biến ngẫu nhiên $X \in \{x_1, ..., x_M\}$ với xác suất tương ứng $p_1, ..., p_M, \forall p_i \geq 0, \sum_{i=0}^{M} p_i = 1$

- Tính không chắc chắn của một biến ngẫu nhiên thể hiện qua phép đo từ không gian xác suất sang tập các giá trị có thể xuất hiện
  - Tính không chắc chắn **không phụ thuộc** vào các mà chúng ta gán nhãn các giá trị trong tập $\{x_1, ..., x_M\}$

ĐẶC TÍNH

- Tính không chắc chắn là một hàm của xác suất $p_1,\ldots,p_M$
- Tính không chắc chắn được định nghĩa bởi một hàm

$$H(p_1,\ldots,p_M)$$

- Tính đơn điệu (monotonicity): Với $f(M)=H(1/M,\ldots,1/M)$ và $M<M'$, ta có

$$f(M)<f(M')$$

  - ➔ Không gian càng lớn khả năng xuất hiện càng cao

- Tính cộng (additivity): Xét 02 biến ngẫu nhiên độc lập *X*, *Y* nằm trong hai không gian có độ lớn *M* và *N*. Vậy tính không chắc chắn của cặp biến ngẫu nhiên *(X,Y)* tỉ lệ với *MN*
  - Do *X* và *Y* độc lập, sự xuất hiện của *X* không ảnh hưởng đến tính chắc chắn của *Y*

$$f(MN)-f(M)=f(N)$$

- Qui tắc nhóm (Grouping rule): Chia đầu ra thành 02 nhóm, ngẫu nhiên chọn một nhóm. Từ nhóm được chọn, ngẫu nhiên chọn một phần tử ➔ Phương pháp lựa chọn này không làm thay đổi tính chất của đầu ra

ENTROPY (1)

- Entropy (self-information-lượng tin riêng):

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

  - Lượng thông tin trong một biến ngẫu nhiên
  - Độ không chắc chắn trung bình của một biến ngẫu nhiên
  - Cho biết không gian của biến ngẫu nhiên (bao gồm giá trị có thể xuất hiện và xác suất xuất hiện)
- Đặc tính của entropy:
  - $H(X) \geq 0$. Nếu $H(X) = 0$ có nghĩa là không có thông tin mới
  - Từ tính đơn điệu (monotonicity): Entropy (lượng tin riêng) tăng khi mà chiều dài bản tin tăng

VÍ DỤ

Ví dụ 1: Cho phổ điểm của một lớp như sau

Điểm	A	B+	B	C+	C	D+	D
Tỉ lệ	1/16	1/8	1/4	1/8	1/4	1/8	1/16

- Tính điểm trung bình của lớp
- Tính lượng tin riêng (entropy) từ phổ điểm của lớp

Ví dụ 2: Cho biến ngẫu nhiên X được định nghĩa như sau: $X = 0$ với xác suất p và $X = 1$ với xác suất $(1-p)$.

- Định nghĩa lượng tin riêng (entropy) của X
- Khảo sát biên thiên của entropy theo xác suất p
- Khi nào thì entropy đạt cực đại

JOINT ENTROPY

- Mở rộng khái niệm cho nhiều biến ngẫu nhiên. Ví dụ một cặp biến ngẫu nhiên *(X,Y)*
- Joint entropy được định nghĩa như sau

$$H(X,Y) = -\sum_{x\in\tilde{X}}\sum_{y\in\tilde{Y}} p(x,y)\log_2 p(x,y)$$

  - Đây là khái niệm mở rộng của entropy cho một biến. Nếu thiết lập một véc-tơ: $Z = [X,Y]^T$ thì quay về khái niệm entropy cho một biến
  - Lượng thông tin trung bình để biểu diễn 02 biến ngẫu nhiên

CONDITIONAL ENTROPY (1)

- Xác suất có điều kiện

$$p(Y \mid X) = \frac{p(X,Y)}{p(X)}$$

- Một cặp biến ngẫu nhiên *(X,Y)*, conditional entropy (entropy có điều kiện) được định nghĩa

$$H(Y \mid X) = \sum_{x \in X} p(x) H(Y \mid X = x) = \sum_{x \in X} p(x) \left[-\sum_{y \in Y} p(y \mid x) \log_2 p(y \mid x) \right] = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y \mid x)$$

- Lượng thông tin thêm cần để truyền *Y* nếu đã biết *X*

CONDITIONAL ENTROPY (2)

- Ví dụ: Có 03 bạn nam (X_1, X_2, X_3) và 03 bạn nữ (Y_1, Y_2, Y_3) tham gia trò chơi ghép đôi

$p(X,Y)$	X_1	X_2	X_3	$p(Y)$
Y_1	1/16	3/8	1/16	1/2
Y_2	1/16	3/16	0	1/4
Y_3	0	3/16	1/16	1/4
$p(X)$	1/8	3/4	1/8	

a)Tính xác suất có điều kiện

$$p(Y_1 | X_1), p(Y_1 | X_2), p(Y_1 | X_3), p(Y_2 | X_1), p(Y_2 | X_2), p(Y_2 | X_3), p(Y_3 | X_1), p(Y_3 | X_2), p(Y_3 | X_3)$$

b) Tính entropy có điều kiện: $H(Y | X)$

c) Tính lượng tin riêng: entropy *H(X)*

d) Tính lượng tin chung: joint entropy H(X,Y)

CONDITIONAL ENTROPY (3)

Xác suất có điều kiện được tính như sau

$$p\left(Y_1 \mid X_1\right) = \frac{p(X_1, Y_1)}{p\left(X_1\right)} = \frac{1/16}{1/8} = \frac{1}{2} \qquad p\left(Y_1 \mid X_2\right) = \frac{p(X_2, Y_1)}{p\left(X_2\right)} = \frac{3/8}{3/4} = \frac{1}{2} \qquad p\left(Y_1 \mid X_3\right) = \frac{p(X_3, Y_1)}{p\left(X_3\right)} = \frac{1/16}{1/8} = \frac{1}{2}$$

Entropy có điều kiện

$$H(Y \mid X) = -\sum_{X \in \tilde{X}} \sum_{Y \in \tilde{Y}} p\left(X, Y\right) \log_2 p(Y \mid X) = 1.375 \text{ [bits]}$$

Entropy (lượng tin riêng)

$$H(X) = -\sum_{x \in \tilde{X}} p\left(x\right) \log_2 p(x) = 1.061 \text{ [bits]}$$

Lượng tin chung (Joint entropy)

$$H(X, Y) = H(Y) + H(Y \mid X) = 2.436 \text{ [bits]}$$

QUY TẮC CHUỖI

- Quy tắc chuỗi (chain rule)
  - Đối với cặp biến ngẫu nhiên (X,Y): $H(X,Y) = H(X) + H(Y \mid X)$
  - Đối với nhiều biến ngẫu nhiên $\left(X_1, \ldots, X_n\right)$

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \ldots + H\left(X_n \mid X_1, \ldots, X_{n-1}\right)$$

- Tính chất

$$H(X \mid Y) \neq H(Y \mid X)$$

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

RELATIVE ENTROPY

- Relative entropy (entropy tương đối):

$$D(p \| q) = \sum_{X \in \tilde{X}} p(x) \log_2 \frac{p(x)}{q(x)}$$

  - Tên gọi khác: Khoảng cách Kullback-Leibler
  - Đo độ không hiệu quả nếu giả sử phân bố xác suất là *q* trong khi phân bố xác suất đúng là *p*
  - Nếu chúng ta dùng phân bố *q* để mã hóa dữ liệu, cần số bits để biểu diễn biến ngẫu nhiên $H(p) + D(p \| q)$

MUTUAL INFORMATION (1)

- Mutual information (thông tin tương hỗ):

$$I(X;Y)=\sum_{x\in\tilde{X}}\sum_{y\in\tilde{Y}}p(x,y)\log_2\frac{p(x,y)}{p(x)q(y)}=D\left(p(x,y)\,\|\,p(x)p(y)\right)$$

  - Giảm tính không chắc chắn của một biến ngẫu nhiên khi biết thông tin của một biến ngẫu nhiên khác
- Mối quan hệ giữa thông tin tương hỗ và entropy

$$I(X;Y)=H(Y)-H(Y\mid X)$$

MUTUAL INFORMATION (2)

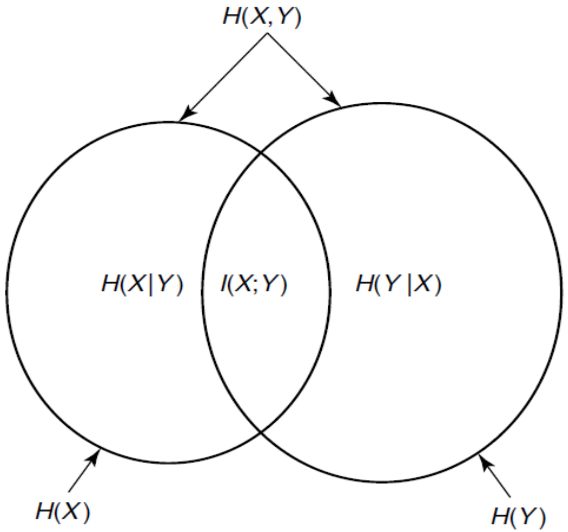
- Ta có: $H(X,Y) = H(X) + H(Y|X)$
- Do đó $I(X;Y) = H(X) + H(Y) - H(X,Y)$

$$I(X;X) = H(X) - H(X|X) = H(X)$$

➔ Entropy là lượng tin riêng

- Từ biểu đồ Vien: Thông tin tương hỗ của X và Y

là giao thoa của thông tin của X và Y.



Biểu đồ Vien