تئوری اطلاعات: جلسه 4 و 5 : 99/7/12

مرور جلسه قبل:

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

↑ متغیر تصادفی گسسته

$$= \mathop{\mathbb{E}}_{\sim P_X}\left[\log_2 \frac{1}{P_X(X)}\right]$$

$\log \frac{1}{P_X(x)}$: میزان شگفتی ما از دیدن وقوع $x \in \mathcal{X}$

آنتروپی شرطی:

$$H(X|Y=y) = -\sum_{x \in \mathcal{X}} P(x|Y=y) \log_2 P(x|Y=y)$$

$$H(X|Y) = \mathop{\mathbb{E}}_{\sim P_Y} H(X|Y=Y) = \sum_{y \in \mathcal{Y}} P_Y(y) \cdot H(X|Y=y)$$

$$= -\mathbb{E}_{P_{XY}}\left[\log_2 P_{X|Y}(X|Y)\right]$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \log_2 P_{X|Y}(x|y)$$

قاعده زنجیره‌ای در آنتروپی : Chain Rule :

$$H(X,Y) = H(X) + H(Y|X)$$
$$\hookrightarrow H(Y) + H(X|Y)$$

$$P(x,y) = P(x)\cdot P(y|x) = P(y)\cdot P(x|y)$$

$$H(X_1,\dots,X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1,\dots,X_{n-1})$$

مثال : P_{XY} :

Y \ X	1	2	3	4	
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
4	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$
					P_Y

$$\Rightarrow P_X = \left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8}\right) \longrightarrow H(X) = \frac{7}{4} \text{ bit}$$

$$P_{X|Y}(X|3) = \left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right) \Rightarrow H(X|Y=3) = 2 \text{ bit}$$

$$H(X|Y) = \frac{1}{4}\Big[\underbrace{H\left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8}\right)}_{=7/4} + \underbrace{H\left(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{8}, \tfrac{1}{8}\right)}_{7/4} + \underbrace{H(X|Y=1)}_{=2} + \underbrace{H\left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right)}_{=2}, \underbrace{H(1,0,0,0)}_{=0}\Big] = \frac{11}{8}$$

$$P_{X|Y}(x|3) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \Rightarrow H(X|Y=3) = 2 \text{ bit}$$

$$H(X|Y) = \frac{1}{4}\Big[\overbrace{H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)}^{=7/4} + \overbrace{H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right)}^{7/4}$$

$$+ \overbrace{H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)}^{H(X|Y=1)\ =2}, \overbrace{H(1,0,0,0)}^{=0}\Big] = \frac{11}{8}$$

$$H(Y|X) = \frac{13}{8} \neq H(X|Y) = \frac{11}{8}$$

رابطه کلی : $H(X) \geq H(X|Y)$

شرطی کردن آنتروپی را کاهش می دهد.

* Relative Entropy (آنتروپی نسبی) : Kullback-Leibler divergence (distance)

بین دو توزیع احتمالاتی که هر دو بر روی الفبای $\mathcal{X}$ تعریف شده‌اند:

$P(x) : \mathcal{X}$

$Q(x) : \mathcal{X}$

$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} = \mathop{\mathbb{E}}_{\sim P}\left[\log \frac{P(x)}{Q(x)}\right]$$

① $D(P\|Q) \geq 0$

② $D(P\|Q)=0 \Rightarrow P=Q$

③ $D(P\|Q) \neq D(Q\|P)$: تقارن ندارد

④ $P(P\|Q)$ در نامساوی مثلث صدق می‌کند:

$\not\times$ $D(P\|R) + D(R\|Q)$

① $0 \cdot \log \frac{0}{Q} = 0$ ② $P \log \frac{P}{0} = \infty$

③ $0 \cdot \log \frac{0}{0} = 0$

* اطلاعات متقابل Mutual Information:

بین دو متغیر تصادفی X و Y با توزیع P_{XY} می‌خواهیم محاسبه کنیم:

$$P_{XY} \longrightarrow \begin{matrix} P_X \\ P_Y \end{matrix}$$

$$I(X;Y) = D\left(P_{XY}(x,y) \,\|\, P_X(x)\cdot P_Y(y)\right)$$

$$= \sum_{x,y \in \mathcal{X},\mathcal{Y}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)\,P_Y(x)}$$

$$= \mathbb{E}_{P_{XY}}\left[\log \frac{P_{XY}(X,Y)}{P_X(X)\cdot P_Y(Y)}\right]$$

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y)\log P_{XY}(x,y) - \sum_{x,y} P_{XY}(x,y)\log P_X(x) - \sum_{x,y} P_{XY}(x,y)\log P_Y(y)$$

($\sum_x P_X(x)$ روی عبارت دوم، $\sum_y P_Y(y)$ روی عبارت سوم)

$$= H(X) + H(Y) - H(X,Y)$$

$$I(X;Y) = D\left(P_{XY}(x,y) \,\|\, P_X(x)\cdot P_Y(y)\right)$$

$$= \sum_{x,y \in \mathcal{X},\mathcal{Y}} P_{XY}(x,y) \ln \frac{P_{XY}(x,y)}{P_X(x)\,P_Y(x)}$$

$$= \mathbb{E}_{P_{XY}}\left[\log \frac{P_{XY}(X,Y)}{P_X(X)\cdot P_Y(Y)}\right]$$

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y)\log P_{XY}(x,y) - \sum_{x,y} P_{XY}(x,y)\log P_X(x) - \sum_{x,y} P_{XY}(x,y)\log P_Y(y)$$

($\sum_x P_X(x)$ over the second term; $\sum_y P_Y(y)$ over the third term)

$$= H(X) + H(Y) - H(X,Y)$$

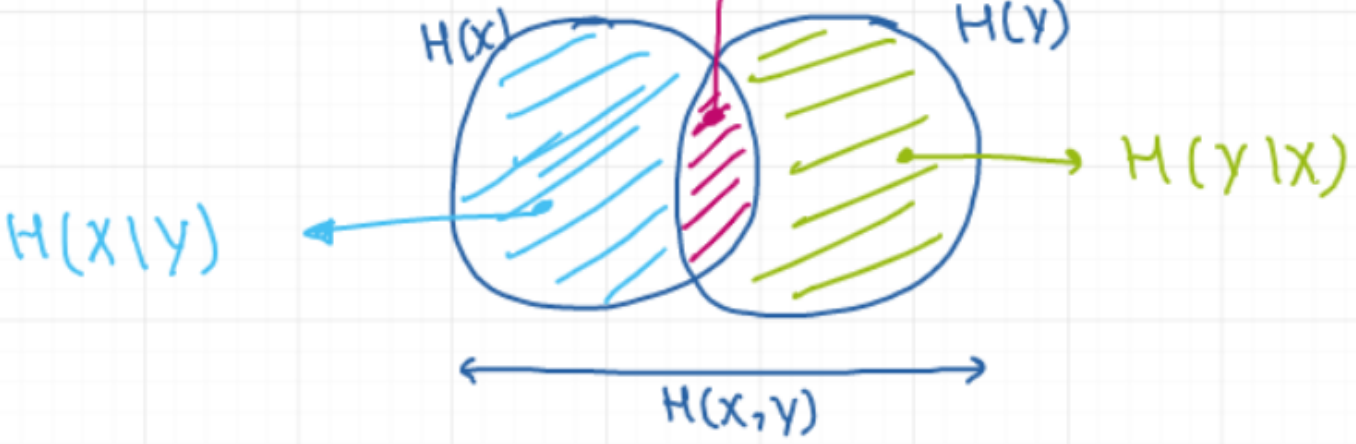
$$H(X,Y) \begin{cases} = H(X) + H(Y|X) \\ = H(Y) + H(X|Y) \end{cases}$$

$$= \begin{cases} H(Y) - H(Y|X) \\ H(X) - H(X|Y) \end{cases} \rightarrow I(X;Y)$$

$$= \begin{cases} H(X) + H(Y|X) \\ H(Y) + H(X|Y) \end{cases}$$

$$= \begin{cases} H(Y) - H(Y|X) \\ H(X) - H(X|Y) \end{cases}$$

$I(X;Y)$



---

* مقدار متقابل شرطی: Conditional Mutual Information

$$I(X;Y|Z) = E_{P_{XYZ}}\left[\log \frac{P_{XY|Z}(x,y|z)}{P_{X|Z}(x|z) \cdot P_{Y|Z}(y|z)}\right]$$

$x_1 \cdots x_n$ $y_1 \cdots y_n$ $z_1 \cdots z_n$

$$= \begin{cases} H(X|Z) - H(X|Y,Z) \\ H(Y|Z) - H(Y|X,Z) \end{cases}$$

قاعده زنجیره‌ای برای اطلاعات متقابل:

$$I(\overbrace{X_1, \cdots X_n}^{X}\ ;\ Y) = H(X) - H(X|Y)$$

$$= H(X_1, \cdots, X_n) - H(X_1, \cdots, X_n | Y)$$

$$= \sum_{i=1}^{n} H(X_i | X_{i-1}, \cdots, X_1) \quad - \sum_{i=1}^{n} H(X_i | X_{i-1}, \cdots, X_1, Y)$$

$$= \sum_{i=1}^{n} \left[H(X_i | X_{i-1}, \cdots, X_1) \quad - H(X_i | X_{i-1}, \cdots, X_1, Y) \right]$$

$$= \sum_{i=1}^{n} I(X_i\ ;\ Y \mid X_{i-1}, \cdots, X_1)$$

$$= I(X_1; Y) + I(X_2; Y | X_1) + \cdots I(X_3; Y | X_1, X_2)$$

$$\cdots + I(X_n; Y | X_1, \cdots, X_{n-1})$$

$P(x,y) : \mathcal{X} \times \mathcal{Y}$: Conditional Relative Entropy

$Q(x,y) : \mathcal{X} \times \mathcal{Y}$

$P(x|y)$

$Q(x|y)$

$$D(P(x|y) \| Q(x|y)) \triangleq \sum_{y} P(y) \left[\sum_{x} P(x|y) \log \frac{P(x|y)}{Q(x|y)} \right]$$

$$= \mathbb{E}_{P_{XY}} \left[\log \frac{P(X|Y)}{Q(X|Y)} \right]$$

$$D(P(x,y) \| Q(x,y)) = D(P(x) \| Q(x)) + D(P(y|x) \| Q(y|x))$$

توابع محدب و مقعر :

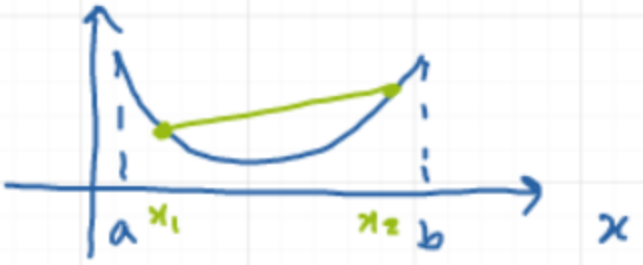
Convex ↑ محدب ↓ ∪

Concave ↑ مقعر ↓ ∩

تابع محدب : تابع f بر روی بازه (a,b) محدب است اگر که برای هر $x_1, x_2 \in (a,b)$ و هر λ که $0 \le \lambda \le 1$

$$f(\underbrace{\lambda x_1 + (1-\lambda)x_2}) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

Convex combination of x_1 and x_2



تابع f اکیدا محدب است اگر که شرط $\le$ ما به صورت نامساوی باشد

Strictly

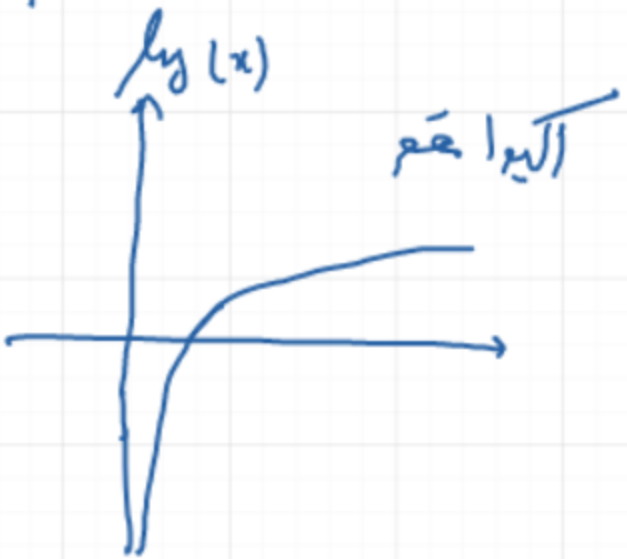
غیر از دو نقطه $\lambda = 0$ و $\lambda = 1$.

* تابع f محدب است اگر $-f$ محدب باشد.

مثال: x^2 محدب

محدب

اکیدا محدب



اکیدا محدب

* اگر تابع $f(x)$ در بازه (a,b) دو بار مشتق پذیر باشد و

تابع f محدب باشد $\iff \forall x \in (a,b): f''(x) \geq 0$

اگر $f''(x) > 0 \quad \forall x \in (a,b)$ $\longleftarrow$ تابع f اکیدا محدب است.

مثال: $f(x)=x\ln x$ برای $x\in(0,+\infty)$

$f'(x)=\ln x+1 \longrightarrow f''(x)=\frac{1}{x}>0$

پس $f(x)$ اکیدا محدب.