Whiteboard 2231407-8211-0005

تئوری اطلاعات : جلسه ۷ ، ۹۹/۷/۱۹ :

---

* نامساوی log-sum
  - نامنفی بودن KL-D
  - KL-D محدب است
  - ۲ تئوری مهم

* نامساوی پردازش اطلاعات (Data processing Ineq)

* آمار کافی (Sufficient Statistics)

* نامساوی Fano

نامساوی log-sum

$a_1, \ldots, a_n \geq 0$

$b_1, \ldots, b_n \geq 0$

Whiteboard 2231407-8211-0005

$b_1, \dots, b_n \geq 0$

$$\sum_{i=1}^{n} a_i \log\left(\frac{a_i}{b_i}\right) \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\left(\sum a_i\right)}{\left(\sum b_i\right)}$$

برقراری تساوی $\iff \frac{a_i}{b_i} = \text{Const} : \forall i$

کاربردهای نامساوی log-sum :

قضیه: برای دو توزیع P, Q بر روی الفبای $\mathcal{X}$: $D(P \| Q) \geq 0$

$$D(P\|Q) = \sum_{x \in \mathcal{X}} \overbrace{P(x)}^{a_i \geq 0} \log \frac{\overbrace{P(x)}^{a_i \geq 0}}{\underbrace{Q(x)}_{b_i \geq 0}} \geq \left(\sum_x P(x)\right) \log \frac{\left(\sum_x P(x)\right)}{\left(\sum_x Q(x)\right)}$$

Whiteboard 2231407-8211-0005

$b_i \geq 0$

$= 1 \cdot \log \frac{1}{1} = 0$ ✓

log-sum Ineq.

$$\forall x \in \mathcal{X} : \frac{p(x)}{q(x)} = c \iff D(p\|q) = 0$$ شرط تساوی جای

$$\longrightarrow \quad p(x) = c\, q(x) \xrightarrow{\sum_x} \overbrace{\sum_{x \in \mathcal{X}} p(x)}^{=1} = c \cdot \overbrace{\sum_{x \in \mathcal{X}} q(x)}^{=1}$$

$$\Rightarrow c = 1$$

قضیه: $D(p\|q)$ نسبت به (p,q) محدب است. فرض کنید که توزیع های (p_1, q_1) و (p_2, q_2)

$$p(x) = \lambda p_1(x) + (1-\lambda) p_2(x)$$

$$q(x) = \lambda q_1(x) + (1-\lambda) q_2(x)$$

اگر $\lambda \geq 0$

$$D(p\|q) = D\big(\lambda p_1 + (1-\lambda)p_2 \,\|\, \lambda q_1 + (1-\lambda)q_2\big)$$

$$\le \lambda D(p_1\|q_1) + (1-\lambda) D(p_2\|q_2)$$

اثبات :

$$\Big(\overbrace{\lambda p_1(x)}^{a_1 \ge 0} + \overbrace{(1-\lambda)p_2(x)}^{a_2 \ge 0}\Big)\log\frac{\lambda p_1(x) + (1-\lambda)p_2(x)}{\underbrace{\lambda q_1(x)}_{b_1 \ge 0} + \underbrace{(1-\lambda)q_2(x)}_{b_2 \ge 0}}$$

$$\overset{\text{log-sum}}{\le} \lambda p_1(x)\log\frac{\cancel{\lambda}p_1(x)}{\cancel{\lambda}q_1(x)} + (1-\lambda)p_2(x)\log\frac{\cancel{(1-\lambda)}p_2(x)}{\cancel{(1-\lambda)}q_2(x)}$$

$$\xrightarrow{\sum_x} \quad D(p\|q) \le \lambda\cdot D(p_1\|q_1) + (1-\lambda)D(p_2\|q_2)$$

تعریف: محدب بودن یک تابع n متغیره: $f(x_1,\dots,x_n): \mathbb{R}^n \longrightarrow \mathbb{R}$

$f(\underline{x})$

می‌گوییم f محدب است اگر که

$\forall\, \underline{x}, \underline{y} \in D_f$

$$f(\lambda \underline{x} + (1-\lambda)\underline{y}) \leq \lambda f(\underline{x}) + (1-\lambda) f(\underline{y})$$

نتیجه: آنتروپی $H(p)$ تابعی مقعر از p است.

می‌دانیم که: $D(P \| \underset{\substack{\uparrow \\ \text{uniform dist.}}}{U}) = \log|\mathcal{X}| - H(P)$

$$\Rightarrow \quad H(P) = \log|\mathcal{X}| - \underbrace{D(P\|U)}_{\text{نسبت به } p \text{ محدب است.}}$$

نسبت به p مقعر است

$\Downarrow$

Whiteboard 2231407-8211-0005

$$\Rightarrow \quad H(p) = \log|\mathcal{X}| - D(p\|u)$$

نسبت به p مقعر است

نسبت به p محدب است.

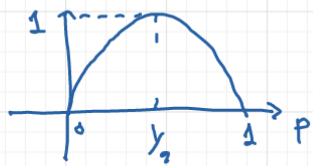
$\Downarrow$

H(p) نسبت به p مقعر است.

مثال: آنتروپی یک متغیر تصادفی برنولی

$$X = \begin{cases} 0 & : \ 1-p \\ 1 & : \ p \end{cases}$$

$$\longrightarrow H(X) = H(p) = -p\log p - (1-p)\log(1-p)$$



با توجه به حرف هایی که بالا زدیم: اگر دو توزیع برنولی با پارامترهای

Whiteboard 2231407-8211-0005

$$X_1 \sim \text{Bernoulli}(q) \qquad X_2 \sim \text{Bernoulli}(r)$$

$$X \sim \text{Bernoulli}(p) \;:\quad p = \lambda q + (1-\lambda) r$$

$$H(p) \geq \lambda H(q) + (1-\lambda) H(r)$$

---

* نامساوی پردازش اطلاعات (Data processing Ineq) :

مرور زنجیره مارکف :

$$X_1, X_2, X_3, \ldots, X_n, \ldots$$

$$\mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_1 = x_1] = \mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}]$$

تعریف : می‌گوییم X و Y و Z یک زنجیره مارکف به شکل $X \to Y \to Z$ تشکیل می‌دهند اگر که

$$P(Z \mid Y, X) = P(Z \mid Y)$$

در حالت کلی برای هر توزیع 3 متغیره داریم :

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

ولی اگر زنجیره مارکف $X \to Y \to Z$ آنگاه :

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

در نتیجه :

$$P(x,z|y) = \frac{P(x,y,z)}{P(y)}$$

$$= \frac{P(x)\cdot P(y|x)\cdot P(z|y)}{P(y)} = \frac{P(x,y)}{P(y)}\cdot P(z|y)$$

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

$$\Rightarrow \boxed{X \to Y \to Z \iff X \perp Z \mid Y}$$

نکته ①: اگر داشته باشیم $X \to Y \to Z \;\Leftarrow\; Z \to Y \to X$

پس خیلی وقت‌ها می‌نویسیم: $X - Y - Z$

یا $X \leftrightarrow Y \leftrightarrow Z$

② فرض کنید $(X,Y) \sim P_{XY}$ و $Z = f(Y)$

قطعی
غیر قطعی

$$\Rightarrow \quad X - Y - Z$$

---

* نامساوی پردازش اطلاعات:

قضیه: اگر زنجیره مارکف $X \leftrightarrow Y \leftrightarrow Z$ داشته باشیم:

$$I(X;Y) \geq I(X;Z)$$

Whiteboard 2231407-8211-0005

اثبات:

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0}$$

$$X \perp Z \mid Y$$

$$\to I(X;Y) + \boxed{I(X;Z|Y)} = 0$$

$$\Rightarrow \quad I(X;Y) \geq I(X;Z)$$

شرط تساوی: $I(X;Y|Z) = 0 \iff X \leftrightarrow Z \leftrightarrow Y$

زنجیره اول: $P(x,y,z) = p(x) \cdot \boxed{p(y|x) \cdot p(z|y)}$

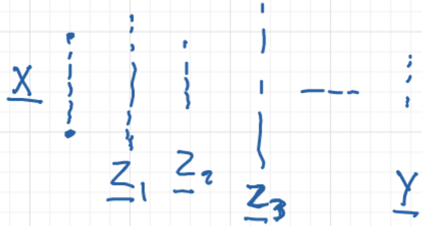
زنجیره دوم: $= p(x) \cdot \boxed{p(z|x) \cdot p(y|z)}$

$$p(y|x) \cdot p(z|y) = p(z|x) \cdot p(y|z)$$

$$P(z|x,y) = P(z|y)$$

$$P(y|x,z) = P(y|z)$$

مثال: شبکه های عصبی



$$\Rightarrow \underline{X} \longleftrightarrow \underline{Z}_1 \longleftrightarrow \underline{Z}_2 \cdots \longleftrightarrow \underline{Y}$$

$$I(\underline{X};\underline{Z}_1) \geq I(\underline{X};\underline{Z}_2) \geq \cdots \geq I(\underline{X};\underline{Y})$$