

به نام خدا

متغیرهای نامساوی اطمینان $D(P||Q) \geq 0$ درست است زیرا است اگر

$$\forall x; P(x) = Q(x)$$

- قضیه بازسازی

اگر $x_1, x_2, \dots, x_n$ ، متغیرهای حاضری با تابع احتمال شرم $P(x_1, x_2, \dots, x_n)$ باشد، نامساوی زیر همواره برقرار است

$$H(x_1, x_2, \dots, x_n) \leq \sum_{i=1}^n H(x_i)$$

ست است برقرار است اگر x_i ها تواتر مستقل باشند

اثبات :

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n \underbrace{H(X_i | X_{i-1}, \dots, X_1)}_{\text{رابطه زنجیره‌ای}} \leq \sum_{i=1}^n H(X_i)$$

رابطه زنجیره‌ای

ستادی برقرار است اگر

X_i ها تدریجاً مستقل باشند

$\forall i$

$$H(X_i | X_{i-1}, \dots, X_1) \leq H(X_i)$$

↑
شرط کردن از ابهام من

ستادی برقرار است اگر X_i مستقل از $X_{i-1}, \dots, X_1$ باشد

خاصیت های Log-Sum

خاصیت های Log-Sum برای اعداد مثبت $a_1, a_2, \dots, a_n$ و $b_1, b_2, \dots, b_n$ به صورت زیر تقریب می شود.

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

اثبات: به کمک خاصیت Jensen

تابع $f(t) = t \log t$ را در نظر بگیرید. این تابع برای $t > 0$ تابعی مقعر (Concave) است. $f'(t) = \frac{t \log t}{t} > 0$ (البته Convex است).

نامساوی Jensen برای یک تابع مقعر $f(t)$ به صورت زیر بیان می‌شود:

$$f(\epsilon t) \leq \epsilon f(t)$$

$$\Rightarrow f\left(\sum \alpha_i t_i\right) \leq \sum \alpha_i f(t_i)$$

$$\sum_{i=1}^n \alpha_i = 1 \quad \text{و} \quad 0 \leq \alpha_i \leq 1$$

برای رسیدن به نامساوی log-sum فراموش نکنیم،

$$t_i = \frac{a_i}{b_i}, \quad \alpha_i = \frac{b_i}{\sum_{j=1}^n b_j}$$

$$f\left(\sum \alpha_i t_i\right) \leq \sum_{i=1}^n \alpha_i f(t_i) \quad \therefore \text{jensen}$$

$$\sum \alpha_i t_i = \sum \frac{\cancel{b_i}}{\sum b_i} \frac{a_i}{\cancel{b_i}} = \frac{\sum a_i}{\sum b_i}$$

$$\begin{aligned} f(t) &= t \log t \\ \implies \frac{\sum a_i}{\cancel{\sum b_i}} \log \frac{\sum a_i}{\sum b_i} &\leq \sum_{i=1}^n \frac{\cancel{b_i}}{\cancel{\sum b_i}} \frac{a_i}{\cancel{b_i}} \log \frac{a_i}{b_i} \end{aligned}$$

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

نسبتی برقرار است اگر $\frac{a_i}{b_i} = c$ ✓

باید ثابت کرد $\log$ -sum برای اطلاعات را محدودیت ثابت کرد.

قضیه: برای اطلاعات $D(P(x) \parallel q(x)) \geq 0$ ، و تساوی برقرار است اگر $P(x) = q(x)$
 $\forall x \in \mathcal{X}$

اثبات: باید $\log$ -sum

$$D(P(x) \parallel q(x)) = \sum_x \underbrace{P(x)}_{a_i} \log \underbrace{\frac{P(x)}{q(x)}}_{b_i} \geq \underbrace{(\sum a_i)}_{\underbrace{\sum P(x)}_1} \log \underbrace{\frac{\sum a_i}{\sum b_i}}_{\underbrace{\frac{\sum P(x)}{\sum q(x)}}_1} \geq 0$$

تست دی برقرار است اگر $\forall x \in X; \frac{p(x)}{q(x)} = c$

اگر چون $p(x)$, $q(x)$ توابع احتمال هستند ، باید $c=1$ باشد

$$p(x) = c q(x) ; \forall x \in X$$

$$\Rightarrow \underbrace{\sum_x p(x)}_1 = c \underbrace{\sum_x q(x)}_1 \Rightarrow c=1$$

پس تست دی برقرار است اگر $\forall x \in X; p(x) = q(x)$

قضیه: آنتروپی نسبی $D(P||Q)$ نسبت به زوج (P, Q) تابعی Convex است.

یعنی اگر (P_1, Q_1) و (P_2, Q_2) زوج‌های مجزا باشند داریم

$$D(\lambda P_1 + (1-\lambda)P_2 \parallel \lambda Q_1 + (1-\lambda)Q_2) \leq \lambda D(P_1 \parallel Q_1) + (1-\lambda)D(P_2 \parallel Q_2)$$

$$D(\lambda P_1 + (1-\lambda)P_2 \parallel \lambda Q_1 + (1-\lambda)Q_2) = \sum \underbrace{\frac{a_i}{\lambda P_1 + (1-\lambda)P_2}}_{\sum a_i} \log \frac{\lambda P_1 + (1-\lambda)P_2}{\underbrace{\lambda Q_1 + (1-\lambda)Q_2}_{\substack{b_1 \quad b_2}}}$$

log-sum $\rightarrow (\sum a_i) \log \frac{\sum a_i}{\sum b_i} \leq \sum a_i \log \frac{a_i}{b_i}$

با یک log-sum

$$\Rightarrow \left(\underbrace{\lambda P_1}_{a_1} + \underbrace{(1-\lambda)P_2}_{a_2} \right) \log \frac{\lambda P_1 + (1-\lambda)P_2}{\underbrace{\lambda q_1}_{b_1} + \underbrace{(1-\lambda)q_2}_{b_2}} \leq \lambda P_1 \log \frac{\lambda P_1}{\lambda q_1} + (1-\lambda)P_2 \log \frac{(1-\lambda)P_2}{(1-\lambda)q_2}$$

$$\Rightarrow (\lambda P_1 + (1-\lambda)P_2) \log \frac{\lambda P_1 + (1-\lambda)P_2}{\lambda q_1 + (1-\lambda)q_2} \leq \lambda P_1 \log \frac{P_1}{q_1} + (1-\lambda)P_2 \log \frac{P_2}{q_2}$$

$$\xrightarrow[\text{از طرفین}]{\sum_n} \sum_n (\lambda P_1 + (1-\lambda)P_2) \log \frac{\lambda P_1 + (1-\lambda)P_2}{\lambda q_1 + (1-\lambda)q_2} \leq \lambda \sum_n P_1 \log \frac{P_1}{q_1} + (1-\lambda) \sum_n P_2 \log \frac{P_2}{q_2}$$

$$\Rightarrow D(\lambda P_1 + (1-\lambda)P_2 \parallel \lambda q_1 + (1-\lambda)q_2) \leq \lambda D(P_1 \parallel q_1) + (1-\lambda)D(P_2 \parallel q_2)$$

قضیه: تابع $H(P)$ تابعی Concave نسبت به P است.

$$H(P) = -\sum P(x) \log P(x) = H(X), \quad X \sim P(x)$$

اثبات ۱:

$$H(P) = \log |X| - \underbrace{D(P \| U)}_{\text{Concave}}$$

قضیه قبل / نسبت به P
Concave

Concave

Concave نسبت به P

اثبات ۲: باینشون حجم که $H(P)$ نسبت به P Concave است یعنی

$$H(\lambda P_1 + (1-\lambda)P_2) \geq \lambda H(P_1) + (1-\lambda)H(P_2)$$

برای اثبات این قضیه در متغیر تصادفی X_1, X_2 را به ترتیب با احتمال P_1 و P_2

در نظر می گیریم $\Rightarrow$

$$X = \begin{cases} X_1 & \lambda \\ X_2 & 1-\lambda \end{cases} \Rightarrow X = \begin{cases} X_1 \sim P_1 & \lambda \\ X_2 \sim P_2 & 1-\lambda \end{cases}$$

متغیر تصادفی $\Rightarrow$ $Z = X_\theta = \begin{cases} X_1 & \lambda (\theta=1) \\ X_2 & 1-\lambda (\theta=2) \end{cases}, \theta = \begin{cases} 1 & \lambda \\ 2 & 1-\lambda \end{cases}$

لیست در متغیر تصادفی

$$\Rightarrow P_z(z) = \lambda P_1 + (1-\lambda) P_2$$

تاج احوال متنہ تصادفی لکس 2

انطرت ریسر می دانیم کہ شروط کردن از اہام می کاہد

$$\Rightarrow H(z) \geq H(z|\theta)$$

$$\Rightarrow H(\underbrace{\lambda P_1 + (1-\lambda) P_2}_{\text{تاج احوال 2}}) \geq \underbrace{E_{\theta} H(z|\theta=\theta.)}_{P(\theta=1) H(z|\theta=1) + P(\theta=2) H(z|\theta=2)}$$

$\lambda H(P_1) + (1-\lambda) H(P_2)$

$$\Rightarrow H(\lambda P_1 + (1-\lambda)P_2) \geq \lambda H(P_1) + (1-\lambda)H(P_2)$$

بنابراین $H(P)$ تابعی Concave / محدب است.

قضیه: اگر x, y ، در مشخّصه داخلی تابع احتمال تراز $P(x, y)$ باشند، آنگاه

تابع $L(x; y)$ متقابل $I(x; y)$ تابعی Concave نسبت به $P(x)$ است، اگر $P(y|x)$ نسبت به y باشد و تابعی Convex نسبت به $P(y|x)$ است، اگر $P(x)$ ثابت باشد.

اثبات صفت اول:

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

$$L(x; y) = H(x) - H(x|y) = H(y) - H(y|x)$$

$$I(x; y) = \underbrace{H(y)} - H(y|x)$$

$$= H(p_y) - H(y|x)$$

از رابطه متقابل استفاده می‌کنیم

از طرف دیگر می‌دانیم که

$$P(y) = \sum_x P(x, y) = \sum_x P(x) \underbrace{P(y|x)}_{c_i} = \sum c_i P(x)$$

کمی رابطه خطی بین $P(y)$ و $P(x)$ برقرار است.

$$I(x; y) = \underbrace{H(y)}_{\text{ConCave}} - H(y|x) = \underbrace{H(y)}_{\text{ConCave}} - \underbrace{\sum P(x) H(y|x=x)}_{\text{خطی}}$$

$H(y)$ مقعر است، $P(y)$ هم مقعر است و ضریب مثبت به $P(x)$ هم مقعر است

$I(x, y)$ نسبت به $P(x)$ تابعی G_{NCAve} است. اگر $P(y|x)$ ثابت باشد

(اثبات سمت دوم: تمرین)

• نامساوی پرایزش داده ها

این نامساوی در رابطه با دنباله های مارکوف تعریف می شود. برای بیان این نامساوی ابتدا مطالبی را در مورد دنباله های مارکوف بیان خواهیم کرد.

سه متغیر تصادفی x, y, z ، یک دنباله مارکوف تشکیل می دهند که به صورت
 $x \rightarrow y \rightarrow z$ نمایش داده می شود اگر دنباله ای

به شرط داشتن y ، x ، z مستقل از هم باشند (استقلال شرطی)

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

به عبارت دیگر دانسته باشیم

۶

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

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

↑
مارک

$x \rightarrow y \rightarrow z$

دنباله‌ای از متغیرهای تصادفی $x_1, x_2, \dots, x_n$ تشکیل می‌دهد دنباله مارکوف در هر

$x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_n$ اگر تنها اثر برای هر سه متغیر تصادفی زیر مجموعه این دنباله

با ترتیب زمانی صعودی یا نزولی، تشکیل یک دنباله مارکن سه تایی به حد.

* اگر $x \rightarrow y \rightarrow z$ دنباله مارکن باشد، $x \rightarrow y \rightarrow z$ هم دنباله مارکن

است. (خاصیت $x \leftrightarrow y \leftrightarrow z$)

* اگر $x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_n$ دنباله مارکن باشد آنگاه

$x_n \rightarrow x_{n-1} \rightarrow \dots \rightarrow x_1$ هم دنباله مارکن است.

* همواره $x \rightarrow y \rightarrow z = f(y)$ یک دنباله مارکن تشکیل می دهد.

مقدار پردازش داده‌ها (راه‌اندازی) از $X \rightarrow Y \rightarrow Z$ یک رابطه مارکوفی شکل

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

دسته‌بندی

- هر چه در رابطه بیش‌تر داریم به‌صورت استدلال ترکیبی می‌شود.

اثبات:

$$I(X; \overbrace{Y}^{k_1}, \overbrace{Z}^{k_2}) = I(X; Z) + I(X; Y|Z)$$

$$\uparrow \text{ رابطه ترکیبی } = I(X; Y) + \underbrace{I(X; Z|Y)}_0$$

چون به شرط Y ،
 X و Z مستقل هستند.

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_1, \dots, X_{i-1})$$

$$\Rightarrow I(x; y) = \underbrace{I(x; z)}_{\geq 0} + \underbrace{I(x; y|z)}_{\geq 0}$$

$$\Rightarrow I(x; y) \geq I(x; z)$$

قضیه پیرایش اطلاعات
 $x \rightarrow y \rightarrow z$

$$x \rightarrow y \rightarrow z \quad I(x; y) \geq I(x; y|z) \quad * \text{نسیب 1}$$

* نسیب 2 : اگر $I(x; y|z) = 0$ باشد آنگاه ستادی برقرار است یعنی

$$I(x; y) = I(x; z)$$

$I(x; y | z) = 0$ اگر به شرط داشتن z ، x و y مستقل (احتمالاً)

به عبارت دیگر $x \rightarrow z \rightarrow y$ رابطه مارکوف تشکیل دهند.

