Subject: 400109843 Date: / /

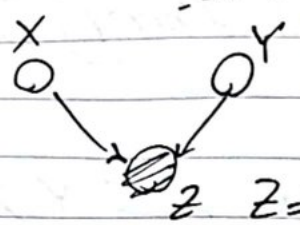
نیما کهن

سوال ① : الف)

1) $I(X;Y|Z) < I(X;Y)$

فرض کنید که X و Y متغیرهای Deterministic باشند به جز Z. در این صورت $I(X;Y)$ یک اطلاعات برای ما دارد اما given Z هر دو متغیر به صورت Deterministic بدست می آیند و هیچ اطلاعاتی برای ما ندارد.

2) $I(X;Y|Z) > I(X;Y)$ Bayes-net روبرو را در نظر بگیرید



در این شرایط X و Y دو متغیر کاملاً مستقل از هم هستند، $I(X;Y)=0$

اما با داشتن Z، X و Y به هم مربوط می شوند به عنوان مثال: $Z=X \text{ or } Y$

~~پس~~ در نتیجه در این صورت با داشتن Z، اگر X یا Y را داشته باشیم با داشتن یکی می توانیم فرض های راجب دیگری کنیم.

* X و Y از هم مستقل باشند: $I(X;Y) = H(X) - H(X|Y) =$

$$= H(X) - \sum_y \sum_x P(x,y) \log P(X|Y) = H(X) - \sum_y \sum_x P(x,y) \log P(x$$

$$= H(X) - \sum_x \sum_y P(x,y) \log P(x) = H(x) - \sum_x P(x) \log P(x) = 0$$

s.a.m

Subject: Date: / /

Chain Rule: $H(X_1, \dots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, \dots, X_1)$ * (1 (ج

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

$= H(Z|Y,X)$

$H(X,Z) - H(X) = H(Z|X) + H(X) - H(X) = H(Z|X)$

$H(Z|X,Y) \leq H(Z|X)$ *** $\rightarrow$ Conditioning Reduces Entropy ** $\Rightarrow$

$\Rightarrow H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$

*** $\rightarrow$ اگر Y نسبت به Z|X مستقل باشد، شرایط تساوی این نامساوی بوجود می‌آید.

* Proof of Chain Rule

$P(x_1, \dots, x_n) = P(x_1) P(x_2|x_1) \cdots P(x_n | x_{n-1} \cdots x_1) = \cancel{\prod_{i=1}^{n} P(x_i|x_{i-1})}$

$= \prod_{i=1}^{n} P(x_i | x_{i-1}, \dots, x_1)$

$H(X) = \mathbb{E}[-\log P(x)] = \mathbb{E}\left\{ -\sum_{i=1}^{n} \log P(x_i | x_{i-1}, \dots, x_1) \right\} =$

$= \sum_{i=1}^{n} \left\{ \mathbb{E}\left\{ \log P(x_i | x_{i-1}, \dots, x_1) \right\} \right\} = \sum_{i=1}^{n} H(X_i | X_{i-1}, \dots, X_1)$

** Proof of Conditioning Reduces Entropy:

$f(x) = -x \log x \rightarrow$ is a Convex function $\Rightarrow -\Sigma x \log x \rightarrow$ Convex

$f\left(\sum_y P(y) P(x|y)\right) \geq \sum_y P(y) f(P(x|y)) \Rightarrow H(X) \geq H(X|Y)$

s.a.m

Subject: Date: / /

سؤال ۱-ب-2) نامساوی داده شده در تمامی شرایط مساوی است.

برای Mutual Information داریم که: $I(X;Y)=I(Y;X)$

و $I(X;Y|Z)=I(Y;X|Z)$

$$I(X;Z|Y)=I(Z;X|Y)=H(Z|Y)-H(Z|Y,X) \quad \text{I}$$

$$I(Z;Y|X)-I(Z;Y)+I(X;Z)=I(Z;Y|X)-I(Z;Y)+I(Z;X)=$$

$$\cancel{=H(Z|X)-H(Z|X,Y)-H(Z|X)+H}$$

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

$$=H(Z|Y)-H(Z|X,Y) \quad \text{II}$$

$$\text{I},\text{II} \Rightarrow I(X;Z|Y)=I(Z;Y|X)-I(Z;Y)+I(X;Z)$$

سؤال ۱-ج-1) توزیع y را به صورت Discrete Uniform بین 4,5,6 در نظر می‌گیریم.

$$\frac{1}{3}=P(y=4)=P(y=5)=P(y=6)$$

$$P(Z=A)=\sum_y P(Z=A|Y=y)P(Y=y)=\frac{1}{3}\times\sum_y P(Z=A|Y=y)=$$

$$=\frac{1}{3}\times\left\{\sum_y \frac{1}{1+e^{-0.5(y-5)}}\right\}=\frac{1}{3}\times\left\{\frac{1}{1+e^{-0.5}}+\frac{1}{1+e^{0}}+\frac{1}{1+e^{0.5}}\right\}=$$

$$=\frac{1}{3}\times\left\{0.622+0.5+0.378\right\}=0.5$$

$$=\frac{1}{3}\times\left\{\frac{1}{2}+\frac{2+e^{-0.5}+e^{0.5}}{2+e^{-0.5}+e^{0.5}}\right\}=0.5$$

s.a.m

CamScanner

Subject: Date: / /

$$P(Z=B) = \sum_y P(Z=B|Y=y)\,P(Y=y) = \frac{1}{2} - \frac{1}{2}P(Z=A) = \frac{1}{2} - \frac{1}{4} =$$

$$P(Z=B) = \frac{1}{4}$$

$$P(Z=C) = \sum_y P(Z=C|Y=y)\,P(Y=y) = \frac{1}{6}\left\{1 - P(Z=A)\right\} = \frac{1}{12}$$

$$P(Z=D) = \sum_y P(Z=D|Y=y)\,P(Y=y) = \frac{1}{3}\left\{1 - P(Z=A)\right\} = \frac{1}{6}$$

$$H(Z) = -\sum_z P(z)\log P(z) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{12}\log 12 + \frac{1}{6}\log 6 =$$

$$\Rightarrow H(Z) = 1{,}73$$

$$\cancel{\sum_y P(Z=A|Y=y)\frac{1}{1+e^{-0{,}5}}\log(1+e^{-0{,}5}) + \frac{1}{2}\log 2 +}$$

$$\cancel{H(Z|Y) = \sum_{z,y} P(Z=z|Y=y)}$$

$$H(Z|Y) = \sum_{z,y} P(Z=z|Y=y)\,P(Y=y)\log\left(P(Z=z|Y=y)\right) =$$

$$= \frac{1}{3}\left\{\frac{1}{1+e^{-0{,}5}}\log(1+e^{-0{,}5}) + \frac{1}{2}\log 2 + \frac{1}{1+e^{0{,}5}}\log(1+e^{0{,}5}) + \right.$$

$$+ \frac{1}{2}\left\{\frac{e^{-0{,}5}}{1+e^{-0{,}5}}\log\left(\frac{2(1+e^{-0{,}5})}{e^{-0{,}5}}\right) + \frac{1}{2}\log 4 + \frac{e^{+0{,}5}}{1+e^{0{,}5}}\log\left(\frac{2(1+e^{+0{,}5})}{e^{0{,}5}}\right)\right\}$$

$$+ \frac{1}{6}\left\{\frac{e^{-0{,}5}}{1+e^{-0{,}5}}\log\left(\frac{6(1+e^{-0{,}5})}{e^{-0{,}5}}\right) + \frac{1}{2}\log 12 + \frac{e^{0{,}5}}{1+e^{0{,}5}}\log\left(\frac{6(1+e^{0{,}5})}{e^{0{,}5}}\right)\right\}$$

$$\left.+ \frac{1}{3}\left\{\frac{e^{-0{,}5}}{1+e^{-0{,}5}}\log\left(\frac{3(1+e^{-0{,}5})}{e^{-0{,}5}}\right) + \frac{1}{2}\log 6 + \frac{e^{0{,}5}}{1+e^{0{,}5}}\log\left(\frac{3(1+e^{0{,}5})}{e^{0{,}5}}\right)\right\}\right\}$$

s.a.m

Subject: Date: / /

$$\Rightarrow H(Z|Y) = \frac{1}{3}\left\{\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{12}\log 12 + \frac{1}{6}\log 6\right\} +$$

$$+ \frac{1}{3}\left\{\frac{1}{2}\log 2 + \frac{1}{6}\log 6 + \frac{1}{3}\log 3\right\} +$$

$$+ \frac{1}{3}\left\{\left(1+\frac{1}{2}+\frac{1}{6}+\frac{1}{3}\right)\left(\frac{1}{1+e^{-0,5}}\log(1+e^{-0,5}) + \frac{1}{1+e^{0,5}}\log(1+e^{0,5})\right)\right\}$$

$$= \frac{2}{3}\left\{\frac{\log(1+e^{-0,5})}{1+e^{-0,5}} + \frac{\log(1+e^{0,5})}{1+e^{0,5}}\right\} + \frac{1}{3}\log 2 +$$

$$+ \frac{1}{3}\times\left\{\frac{1}{2}\log 2 + \frac{1}{6}\log 6 + \frac{1}{3}\log 3\right\} = 1,46$$

$$I(Z;Y) = H(Z) - H(Z|Y) = 1,73 - 1,46 = 0,27$$

سوال ۱-ج-2) مقدار اطلاعاتی است که Z به ما راجع Y می دهد.

در صورتی که مقدار به 0 نزدیک تر باشد، به این معناست که Z و Y به یکدیگر وابسته تر هستند. در غیر این صورت، هرچه قدر مقدار بزرگتری داشته باشد، به این معناست که Y و Z مستقل تر از یکدیگر هستند.

s.a.m

Subject: Date / /

$$D_{KL}\left(P_{(x,y,z)} \,\|\, P_{(x)}P_{(y)}P_{(z)}\right) = \sum_{x,y,z} P_{(x,y,z)} \log\left(\frac{P_{(x,y,z)}}{P_{(x)}P_{(y)}P_{(z)}}\right)$$ سؤال ۱-د)

$$= \sum_{x,y,z} P_{(x,y,z)} \log\left(P_{(x,y,z)}\right) - \sum_{x,y,z} P_{(x,y,z)} \log\left(P_{(x)}\right)$$

$$- \sum_{x,y,z} P_{(x,y,z)} \log\left(P_{(y)}\right) - \sum_{x,y,z} P_{(x,y,z)} \log\left(P_{(z)}\right) =$$

$$= \sum_{x,y,z} P_{(x,y,z)} \log\left(P_{(x,y,z)}\right) - \sum_{x} P_{(x)} \log P_{(x)} - \sum_{y} P_{(y)} \log P_{(y)}$$

$$- \sum_{z} P_{(z)} \log P_{(z)} =$$

$$\Rightarrow \cancel{D_{KL}} = -H_{(x,y,z)} + H_{(x)} + H_{(y)} + H_{(z)} \Rightarrow$$

$$\Rightarrow D_{KL}\left(P_{(x,y,z)} \,\|\, P_{(x)}P_{(y)}P_{(z)}\right) = -H_{(x,y,z)} + H_{(x)} + H_{(y)} + H_{(z)}$$

سؤال ۱-هـ)

$$\mathcal{I}(X_1;X_3) + \mathcal{I}(X_2;X_4) \leq \mathcal{I}(X_1;X_4) + \mathcal{I}(X_2;X_3) \Leftrightarrow$$

$$\Leftrightarrow \mathcal{I}(X_3;X_1) + \mathcal{I}(X_4;X_2) \leq \mathcal{I}(X_4;X_1) + \mathcal{I}(X_3;X_2) \Leftrightarrow$$

$$\Leftrightarrow H(X_3) - H(X_3|X_1) + H(X_4) - H(X_4|X_2) \leq$$

$$\leq H(X_4) - H(X_4|X_1) + H(X_3) - H(X_3|X_2) \Leftrightarrow$$

$$\Leftrightarrow H(X_4|X_1) + H(X_3|X_2) \leq H(X_3|X_1) + H(X_4|X_2) \quad (I)$$

s.a.m

CamScanner

Subject: Date: / /

$$H(X_4|X_2) = H(X_4, X_3|X_2) - H(X_3|X_2, X_4)$$

$$H(X_3|X_1) = H(X_3, X_4|X_1) - H(X_4|X_3, X_1)$$

$$H(X_4|X_1) = H(X_3, X_4|X_1) - H(X_3|X_4, X_1)$$

$$H(X_3|X_2) = H(X_3, X_4|X_2) - H(X_4|X_3, X_2)$$

$$\text{I} \Leftrightarrow H(X_3|X_2, X_4) + H(X_3|X_4, X_1) \leq H(X_4|X_3, X_2) + H(X_4|X_3, X_1)$$

$\Leftrightarrow$ Using Markov Chain Properties:

$$H(X_3|X_2) - H(X_3|X_1) \leq H(X_4|X_3) - H(X_4|X_3) \Leftrightarrow$$

$$\Leftrightarrow H(X_3|X_2, X_1) \leq H(X_3|X_1)$$

↳ Conditioning Reduces Entropy

s.a.m

Object: Date: / /

t-Order Condition: $f(y) \geq \nabla f(x)^T (y-x) + f(x)$ سؤال 2-آ)

$\cancel{\forall y \; f(y) \geq f}$ در نتیجه اگر $\nabla f(x) = 0$ باشد.

$\forall y \in dom(f) \quad f(y) \geq f(x)$

of of first-Order Condition:

$f(\alpha y + (1-\alpha)x) \leq (1-\alpha) f(x) + \alpha f(y) \rightsquigarrow$ Convexity

$\rightsquigarrow f(y) \geq \dfrac{f(\alpha y + (1-\alpha)x) - f(x)}{t} + f(x) \rightarrow \lim_{t \to 0}$

$f(y) \geq f(x) + f'(x) \times (y-x)$

به همین صورت می‌توانیم برای چند متغیره آن را داشته باشیم.

$\cancel{f(\alpha y + (1-\alpha)x)} \quad f(y) \geq f(x) + \nabla f(x)^T (y-x)$

ب)

$z^2 = x^2 + y^2 \rightarrow f(x,y,z) = g(x,y) = 2\{x^2 + y^2\}$

$\left.\begin{array}{l} z^2 = x^2 + y^2 \\ z = (x+y+1) \end{array}\right\} \Rightarrow z^2 = x^2 + y^2 = (x+y+1)^2 = x^2 + y^2 + 1 + 2x + 2y + 2xy \Rightarrow$

$\Rightarrow 1 + 2x + 2y + 2xy = 0 \Rightarrow y = -\dfrac{1+2x}{2x+2}$

$g(x,y) = h(x) = 2\left\{x^2 + \left(\dfrac{1+2x}{2x+2}\right)^2\right\}$

s.a.m

CamScanner

Subject: Date: / /

مسئله اولیه ۶ برابر است با

$$\begin{aligned} \min\ \text{max} \quad & x^2+y^2+z^2 \\ \text{s.t.} \quad & z^2 = x^2+y^2 \\ & z = x+y+1 \end{aligned}$$

و می‌توان این مسئله را حل کرد با پیدا کردن کمینه و بیشینه $h_{(x)} = 2\left\{x^2+\left(\frac{1+2x}{2+2x}\right)^2\right\}$

$$\cancel{h'_{(x)} = 0 \rightsquigarrow h'_{(x)} = 2x+2\left\{\frac{1+2x}{2+x}\right\}\left\{\frac{2(2+x)-1+2x}{(2+x)^2}\right\} = 0}$$

$$\cancel{= 2x + \frac{6(1+2x)}{(2+x)^3} = 0 \Rightarrow}$$

$$\cancel{\Rightarrow x + \frac{3(1+2x)}{(2+x)^3} = 0 \rightsquigarrow x(2+x)^3+3(1+2x) = 0 \rightsquigarrow}$$

$$h_{(x)} = 2\left\{x^2+\frac{1}{4}\left\{\frac{1+2x}{x+1}\right\}^2\right\} \rightsquigarrow h'_{(x)} = 0 \rightsquigarrow$$

$$\rightsquigarrow h'_{(x)} = 2\left\{2x+\frac{1}{2}\left(\frac{1+2x}{x+1}\right)\times\left\{\frac{2(x+1)-1+2x}{(x+1)^2}\right\}\right\} = 0 \rightsquigarrow$$

$$\rightsquigarrow 2x+\frac{1}{2}\frac{(1+2x)}{(x+1)^3} = 0 \rightsquigarrow \cancel{4(x+1)}\ 4x(x+1)^3+(1+2x) = 0$$

$$= 4x^4+12x^3+12x^2+6x+1 = 0 \rightsquigarrow$$

$$\rightsquigarrow \cancel{x^4+3x^3} = (2x^2+2x+1)(2x^2+4x+3) = 0 \Rightarrow$$

$$\Rightarrow x = -\frac{1}{2}-\frac{i}{2} \ \text{یا}\ x = -\frac{1}{2}+\frac{i}{2} \ \text{یا}\ x = -1-\frac{1}{\sqrt{2}} \ \text{یا}\ x = \frac{1}{\sqrt{2}}-1$$

در نظر بگیرید که $h_{(x)} = 2\left\{x^2+\left(\frac{1+2x}{2+2x}\right)^2\right\}$ یک تابع است که هیچ حد بالایی ندارد و از بالا unbounded است. در نتیجه بیشینه‌ای برابر با بی‌نهایت دارد.

s.a.m

برای به دست آوردن کمینه نیز با استفاده از جواب‌هایی که در R وجود دارند می‌توان به دست آورد.

$$x = -1+\frac{\sqrt{2}}{2} \rightsquigarrow h(x) = 2\left\{\left(-1+\frac{\sqrt{2}}{2}\right)^2 + \frac{1}{4}\left(\frac{1+2\left(-1+\frac{\sqrt{2}}{2}\right)}{1+-1+\frac{\sqrt{2}}{2}}\right)^2\right\}$$

$$\hookrightarrow h(x) = 6-4\sqrt{2}$$

$$x = -1-\frac{\sqrt{2}}{2} \rightsquigarrow h(x) = 2\left\{\left(-1-\frac{\sqrt{2}}{2}\right)^2 + \frac{1}{4}\left(\frac{1+2\left(-1-\frac{\sqrt{2}}{2}\right)}{1+1-\frac{\sqrt{2}}{2}}\right)^2\right\}$$

$$\hookrightarrow h(x) = +6+4\sqrt{2}$$

$$h(x) \rightarrow \min: 6-4\sqrt{2} \qquad \max: +\infty$$

Dual Problem: $\max\limits_{\mu,\lambda} L(\mu,\lambda)$ s.t. $\mu \geq 0$ (سوال 2-ج)

$$L(\mu,\lambda) = \inf_x \{ h(x) + \mu^T g(x) + \lambda^T h(x) \}$$

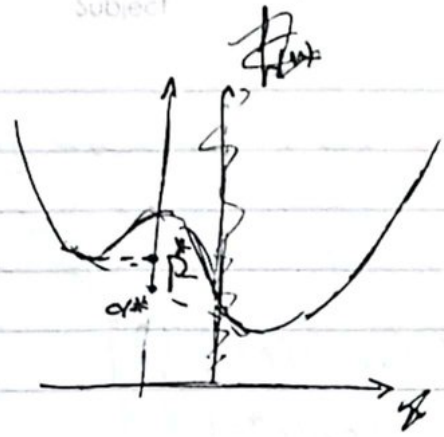
در نظر بگیرید که $\tilde{x}$ یک نقطه feasible در مسئله اصلی باشد.

$$g_i(x) \leq 0,\ h_j(x) = 0,\ \mu \geq 0 \Rightarrow \mu^T g(x) + \lambda^T h(x) \leq 0$$

$$L(\tilde{x},\mu,\lambda) = h(\tilde{x}) + \mu^T g(\tilde{x}) + \lambda^T h(\tilde{x}) \leq h(\tilde{x})$$

$$L(\mu,\lambda) = \min_x \{ L(x,\mu,\lambda) \} \leq L(\tilde{x},\mu,\lambda) \leq h(\tilde{x}) \Rightarrow$$

$$\Rightarrow L(\mu,\lambda) \leq \min_x h(x)$$



→ به عنوان مثال این مسئله رسم شده، duality gap غیرصفر دارد و در این صورت تساوی عبارت هیچ موقع رخ نمی‌دهد.

به عنوان مثال x^* مثالی از این عبارت است.

Primal: minimize $-\sum_{i=1}^{n} \log(x_i+\alpha_i)$ (سؤال 2 - د)

s.t. $x \geq 0$, $\mathbb{1}^T x = 1$

Dual: $L(x,\lambda,\nu) = -\sum_i \log(x_i+\alpha_i) - \lambda^T x + \nu(\mathbb{1}^T x - 1)$

KKT: ① $x \geq 0$, $\mathbb{1}^T x = 1$ ② $\lambda \geq 0$

③ $\lambda_i x_i = 0$ for $i=1,\dots,n$ ④ $\nabla_x L = 0$

$$\frac{\partial L}{\partial x_i} = \frac{1}{x_i+\alpha_i} + \lambda_i = \nu \quad \begin{cases} \nu \leq 1/\alpha_i : \lambda_i = 0 \Rightarrow x_i = 1/\nu - \alpha_i \\ \nu \geq 1/\alpha_i : x_i = 0 \quad \lambda_i = \nu - 1/\alpha_i \end{cases}$$

$$\mathbb{1}^T x = 1 \rightsquigarrow \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$$

s.a.m

سؤال 2-ه)

$$\max_P \left\{ -\int P_{(x)} \log P_{(x)} dx \right\} \qquad \rightsquigarrow P_{(x)} \geq 0$$

$$\int P_{(x)} dx = 1 \quad , \quad \int x P_{(x)} dx = \mu \quad , \quad \int \left(x - \int x' P_{(x')} dx'\right)^2 P_{(x)} = \sigma^2$$

برای حل این مسئله می‌توان تابع زیر را بایستی کمینه کرد:

$$\int P_{(x)} \log P_{(x)} dx + \lambda_0 \left\{ \int P_{(x)} dx - 1 \right\} + \lambda_1 \left\{ \int (x-\mu)^2 P_{(x)} dx - \sigma^2 \right\}$$

با استفاده از برابری لاگرانژ اویلر داریم که:

$$\frac{\partial F}{\partial P} = 0$$

$$\frac{\partial F}{\partial P_{(x)}} = 1 + \log P_{(x)} + \lambda_0 + \lambda_1 (x-\mu)^2 = 0 \rightsquigarrow$$

$$\rightsquigarrow P_{(x)} = \exp\left\{ -1 - \lambda_0 - \lambda_1 (x-\mu)^2 \right\}$$

$$1: \int_{-\infty}^{+\infty} P_{(x)} dx = 1 \Rightarrow \int_{-\infty}^{+\infty} e^{-1-\lambda_0-\lambda_1(x-\mu)^2} dx \rightsquigarrow u := x-\mu \quad du = dx$$

$$= \int_{-\infty}^{+\infty} e^{-1-\lambda_0} \times e^{-\lambda_1 u^2} du = e^{-1-\lambda_0} \times \sqrt{\frac{\pi}{\lambda_1}} = 1 \Rightarrow e^{-1-\lambda_0} = \sqrt{\frac{\lambda_1}{\pi}} \quad \text{I}$$

$$2: \int_{-\infty}^{+\infty} P_{(x)} (x-\mu)^2 dx = \sigma^2 = \int_{-\infty}^{+\infty} e^{-1-\lambda_0-\lambda_1 u^2} u^2 du =$$

$$= -\frac{d}{d\lambda_1} \left\{ \int_{-\infty}^{+\infty} e^{-1-\lambda_0-\lambda_1 u^2} du \right\} = -\frac{d}{d\lambda_1} \left\{ e^{-1-\lambda_0} \times \sqrt{\frac{\pi}{\lambda_1}} \right\} =$$

$$= \frac{1}{2} \sqrt{\frac{\pi}{\lambda_1^3}} e^{-1-\lambda_0} = \sigma^2 \quad \text{I} \quad \Rightarrow \frac{1}{2\lambda_1} = \sigma^2 \Rightarrow \lambda_1 = \frac{1}{2\sigma^2}$$

$$e^{-1-\lambda_0} = \frac{1}{\sqrt{2\pi\sigma^2}} \Longrightarrow P_{(x)} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = N(\mu, \sigma)$$

s.a.m

سؤال 3-الف)

Transition Function

Stationary Distribution: $\overset{\nearrow}{T}v = v \Rightarrow (T-I)v = 0$

$$\Rightarrow \forall x : -\frac{|N(x)|}{M} v_x + \sum_{y \in N(x)} \frac{v_y}{M} = 0 \rightsquigarrow |N(x)| v_x = \sum_{y \in N(x)} v_y \Rightarrow$$

$$\Rightarrow v_x = \frac{1}{|N(x)|} \sum_{y \in N(x)} v_y$$

(یا به عبارت دیگر، احتمال هر state میانگین همسایه‌های خود می‌باشد.)

حال state x_1 را در نظر بگیرید که $\forall j : v_{x_1} < v_j$

با توجه به اینکه v_{x_1} مینیمم حالت بین بقیه است و برابر میانگین همسایه‌ها است.

می‌توان گفت که $\forall y \in N(x) \cdot v_{x_1} = v_y$ حال همین صحبت را به صورت iterative، برای بقیه state ها گفت و با توجه به اینکه این زنجیره مارکوف غیرقابل کاهش است

می‌توان گفت که همه ایت‌ها v_i یکسانی دارند. $\sum_i v_i = 1 \Rightarrow v_i = \frac{1}{n}$

توزیع ساناً این زنجیره مارکوف یک توزیع یکنواخت است.

سؤال 3-ب) برای تخمین جواب نهایی از بردارهای موجود نمونه برداری می‌کنیم.

فرض کنید جواب اصلی مسئله برابر با k باشد. $\sum_{i=1}^{n} a_i x_i < b$

$$\text{bias} = \mathbb{E}\left\{\frac{|P| \times 2^n}{N}\right\} = \frac{2^n}{N} \times \mathbb{E}\{|P|\} = \frac{2^n}{N} \times \sum_{i=1}^{N} i \times P(\cancel{\ldots})$$

$$= \frac{2^n}{N} \times \frac{k \times N}{2^n} = k \longrightarrow \text{bias} = 0$$

s.a.m

$$\mathbb{E}\left\{\left(\frac{|P|\times 2^n}{N}\right)^2\right\} = \frac{4^n}{N^2}\times\sum_{i=1}^{N} i^2\times\binom{N}{i}\times\left(\frac{k}{2^n}\right)^i\left(1-\frac{k}{2^n}\right)^{N-i} =$$

$$= \frac{k}{N}\left\{(N-1)k+2^n\right\} = k^2-\frac{k^2}{N}+\frac{2^n k}{N} =$$

$$\Rightarrow Var\left\{\frac{|P|\times 2^n}{N}\right\} = \frac{k}{N}(2^n-k)$$ → در نتیجه با توجه به اینکه 2^n از k خیلی بیشتر است. نیاز داریم که خیلی داده بگیریم تا Variance کم شود.

می‌توان گفت مشکل این روش، مانند Monte Carlo First-Visit است.

زیرا اگر یک جواب داشته باشیم، تمامی زیر جواب‌ها نیز باید شمرده شوند که در این روش شمرده نمی‌شوند.

سوال 3- ج-1) با توجه به اینکه $\sum_{i=1}^{n} a_i \neq 1$ است. شرط کولمگروف نقض می‌شود.

قسمت الف) سوال را در نظر بگیرید این قسمت را مانند کمترین قسمت حل می‌کنیم با عوض کردن هر بیت به یک بیت دیگر می‌رویم. هر بیت n همسایه دارد. که با توجه به اینکه متغیرها را از توزیع یکنواخت انتخاب می‌کنیم، به هر همسایه با احتمال $1/n$ می‌رویم. همانطور که در قسمت الف اثبات کردیم، این زنجیره مارکوف یک توزیع مانا به صورت توزیع یکنواخت دارد.

سؤال 3-ج-2)

با توجه به اینکه اگر با یک کردن x_i شرط کوله پشتی نقض ~~می‌شود~~ شود، در همان ایتیت X_i می‌مانیم. می‌توان گفت که در نهایت تنها ایتیت‌هایی موجود هستند که شرط ~~کوله~~ کوله پشتی را رعایت کرده‌اند. در نتیجه می‌توان از تعداد ایتیت‌های موجود برای تخمین ~~تعداد~~ تعداد حالات برای مسئله کوله پشتی استفاده کرد.

در بی‌نهایت نیز می‌توان گفت که این زنجیره در واقع تمامی ایتیت‌های مورد قبول را برای ما ~~مشخص~~ مشخص می‌کند.

از ایتیت x به احتمال $\frac{1}{n}$ به ایتیت همسایه می‌رویم و با احتمال $1-\frac{|N(x)|}{m}$ که در آن $N(x)$ ایتیت‌هایی هستند که ~~شرط کوله پشتی~~ با تغییر یک x_i در ~~ایتیت~~ ایتیت اولیه به حالتی برویم که همچنان شرط کوله پشتی برقرار باشد.

با توجه به اینکه از هر ایتیت با تغییر 0و1ها می‌توان به بقیه ایتیت‌ها رسید. این زنجیره مارکوف غیرقابل کاهش است.

s.a.m

سوال ۴-

$$P(s|same) = \alpha_{ss} \exp(\lambda_s s) = \alpha_{ss} e^{\lambda_s s}$$

$$P(s|different) = \alpha_{ds} \exp(-\lambda_d s) = \alpha_{ds} e^{-\lambda_d s}$$

الف) در واقع هدف ما این است که خاطرات درست را به این افراد نشان دهد.

در نتیجه باید امتیاز بیشتری به same و امتیاز کمتر برای different بودن باشد.

که این شرایط در صورتی اتفاق می افتد که λ_d و λ_s بزرگتر از صفر باشند.

احتمال نمره بالاتر برای same بیشتر و احتمال نمره پایین برای different بیشتر است.

همچنین برای نرمال سازی باید دو شرط زیر را داشته باشیم.

$$① : \int_0^1 P(s|same)\, ds = 1 \rightsquigarrow \int_0^1 \alpha_{ss} e^{\lambda_s s} = 1 \Rightarrow \alpha_{ss} = \frac{\lambda_s}{e^{\lambda_s} - 1}$$

$$② : \int_0^1 P(s|different)\, ds = 1 \rightsquigarrow \int_0^1 \alpha_{ds} e^{-\lambda_d s} = 1 \rightsquigarrow \alpha_{ds} = \frac{\lambda_d}{1 - e^{-\lambda_d}}$$

در نتیجه ضرایب α_{ss} و α_{ds} تنها به خاطر نرمال سازی توابع احتمال ما هستند.

ب) در این سوال می خواهیم $P(same_j | s_1 \dots s_n)$ را حساب کنیم

$$P(same_j | s_1, \dots, s_n) = \frac{P(s_1 \dots s_n | j)}{P(s_1 \dots s_n)} \times P(j) =$$

$$① = \frac{P(s_1 \dots s_n | j)}{\sum P(s_1 \dots s_n | i) P(i)} \times P(j)$$

داریم که احتمال prior همه با هم برابر است

$$\Rightarrow (I) = \frac{P(s_1,\dots,s_n|j)}{\sum P(s_1,\dots,s_n|i)}$$

$$P(s_1,\dots,s_n|i) = P(\cancel{same}\ s_i|same) \times \prod_{i\neq j} P(s_j|different) =$$

$$= \alpha_{ss} \times \alpha_{ds}^{n-1} \times e^{\lambda_s s_i} \times \prod_{i\neq j} e^{-\lambda_d s_j} = \alpha_{ss} \times \alpha_{ds}^{n-1} e^{(\lambda_s+\lambda_d)s_i} \prod_j e^{-\lambda_d s_j}$$

$$P(same_j|s_1,\dots,s_n) = \frac{P(s_1,\dots,s_n|j)}{\sum_k P(s_1,\dots,s_n|k)} = \frac{\alpha_{ss} \times \alpha_{ds}^{n-1} e^{(\lambda_s+\lambda_d)s_i} \prod_j e^{-\lambda_d s_j}}{\sum_k \{\alpha_{ss} \times \alpha_{ds}^{n-1} e^{(\lambda_s+\lambda_d)s_k} \prod_j e^{-\lambda_d s_j}\}}$$

$$\Rightarrow P(same_j|s_1,\dots,s_n) = \frac{e^{(\lambda_s+\lambda_d)s_i}}{\sum_k e^{(\lambda_s+\lambda_d)s_k}}$$

سوال ۹ ب) داریم که

$$P(s_1,\dots,s_n) = \sum_i P(s_1,\dots,s_n|i) \times P(i) =$$

$$= \sum_i \alpha_{ss} \cancel{\alpha_{ds}^{n-1}} \times e^{(\lambda_s+\lambda_d)s_i} \times \frac{1}{n} \rightsquigarrow P(i): \text{prior}$$

از قسمت قبل نیز داریم که

$$P(same_i|s_1,\dots,s_n) = \frac{e^{(\lambda_s+\lambda_d)s_i}}{\sum_k e^{(\lambda_s+\lambda_d)s_k}}$$

$$P(same_{s_{max}}) = \int_0^1 P(same_{max}|s_{max}) P(s_{max})\, ds$$

$$P(same_{max}|s_{max}) = \int P(same_{max}|s_1,\dots,s_n) P(s_1,\dots,s_{n-1}|s_n)\, ds$$

$$P(same_{max}|s) = \frac{e^{(\lambda_s+\lambda_d)s_{max}}}{\sum_k e^{(\lambda_s+\lambda_d)s_k}} \Rightarrow$$

s.a.m

در نتیجه داریم که:

$$P(s_1 \dots s_{n-1} | s_n) = \frac{P(s_1 \dots s_n)}{P(s_n)} = \frac{\sum_k e^{(\lambda_s + \lambda_d) s_k}}{\int_s e^{(\lambda_s + \lambda_d) s} \, ds}$$

که به این صورت $P(same_{max} | s_{max})$ محاسبه می‌شود

و $P(same) = \int P(same_{max} | s_{max}) P(s_{max}) \, ds_{max}$

را نیز می‌توان حساب کرد.

s.a.m

سوال ۵)

آ)

$$P(x_1, \dots, x_n \mid \theta) = \prod_i P(x_i \mid \theta) = \prod_i \cancel{e^{-x}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\hookrightarrow \log P(X \mid \theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum (x_i-\mu)^2}{2\sigma^2} = \log L(\theta; X)$$

$$\frac{\partial L}{\partial \theta} = 0 \rightsquigarrow \frac{\partial \log L}{\partial \theta} = 0$$

$$\frac{\partial \log L}{\partial \hat{\mu}} = 0 \rightsquigarrow \frac{1}{\hat{\sigma}^2} \sum_i (x_i - \hat{\mu}) = 0 \rightsquigarrow \hat{\mu} = \frac{\sum x_i}{n}$$

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \rightsquigarrow -\frac{n}{2\hat{\sigma}^2} + \frac{\sum (x_i - \hat{\mu})^2}{2(\hat{\sigma}^2)^2} \Rightarrow \hat{\sigma}^2 = \frac{\sum (x_i - \hat{\mu})^2}{n}$$

$$\hat{\theta}_{MLE} = \left(\frac{\sum x_i}{n}, \frac{\sum \left(x_i - \frac{\sum x_j}{n}\right)^2}{n} \right)$$

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}\left(\frac{\sum x_i}{n}\right) = \frac{\sum \mathbb{E}(x_i)}{n} = \frac{n\mu}{n} = \mu \rightarrow \text{Unbiased}$$

$$\cancel{\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}\left\{\frac{\sum (x_i - \mu)^2}{n}\right\} = \frac{1}{n} \sum_i \left\{\mathbb{E}(x_i^2) - 2\mu \mathbb{E}(x_i) + \mu^2\right\} =}$$

$$\cancel{= \frac{1}{n} \sum_i \left\{\mathbb{E}(x_i^2) - \mu^2\right\}}$$

$$\cancel{Var(x_i) = \mathbb{E}(x_i^2) - \mathbb{E}(x_i)^2 = \sigma^2 \rightsquigarrow \mathbb{E}(x_i^2) = \sigma^2 + \mu^2} \Rightarrow$$

$$\cancel{\Rightarrow \mathbb{E}(\hat{\sigma}^2) =}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \hat{\mu})^2 = \frac{1}{n} \sum_i \left(x_i - \frac{\sum x_j}{n}\right)^2 =$$

$$= \frac{1}{n} \sum_i \left\{ x_i^2 - 2x_i \frac{\sum x_j}{n} + \left(\frac{\sum x_j}{n}\right)^2 \right\} =$$

s.a.m

$$= \frac{\sum x_i^2}{n} - \frac{2\left(\sum x_i\right)^2}{n^2} + \frac{\left(\sum x_i\right)^2}{n^2} = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$$

$$\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}\left\{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2\right\} = \frac{\sum \mathbb{E}(x_i^2)}{n} - \frac{\mathbb{E}\left(\left(\sum x_i\right)^2\right)}{n^2} =$$

$$= \frac{\sum \mathbb{E}(x_i^2)}{n} - \frac{1}{n^2}\left\{\mathbb{E}\left(\sum_i x_i^2\right) + \mathbb{E}\left\{\sum_i \sum_{j\neq i} x_i x_j\right\}\right\} =$$

$$= \frac{\sum \mathbb{E}(x_i^2)}{n} - \frac{\sum \mathbb{E}(x_i^2)}{n^2} - \frac{\sum_i \sum_{j\neq i} \mathbb{E}(x_i)\mathbb{E}(x_j)}{n^2}$$

$$\mathbb{E}(x_i^2) - \{\mathbb{E}(x_i)\}^2 = \sigma^2 \Rightarrow \mathbb{E}(x_i^2) = \sigma^2 + \mu^2$$

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2 + \mu^2 - \frac{\sigma^2 + \mu^2}{n} - \frac{n(n-1)\mu^2}{n^2} = (\sigma^2 + \mu^2)\left(1 - \frac{1}{n}\right) - \left(1 - \frac{1}{n}\right)\sigma^2 =$$

$$= \sigma^2 \times \left(1 - \frac{1}{n}\right)$$

ب) برای بدست آوردن $\tau(\theta)_{mom}$ در واقع باید

$\underset{\tau(\theta)}{\operatorname{argmax}}\ L^*(\eta; x)$ را حساب کنیم

که در واقع دو آن داریم.

$$L^*(\eta, x) = \sup_{\{\theta : \tau(\theta) = \eta\}} L(\theta; x)$$

$$L^*(\hat{\eta}; x) = \sup_{\eta} L^*(\eta, x) = \sup_{\eta} \sup_{\{\theta : \tau(\theta) = \eta\}} L(\theta; x) = \sup_{\theta} L(\theta, x)$$

$$= \sup L(\hat{\theta}, x)$$

در نتیجه $\hat{\tau}(\theta)$ همان $\tau(\hat{\theta})$ است.

s.a.m

$$\frac{1}{\hat{\mu}^2+1} = \frac{1}{\left(\frac{\sum x_i}{n}\right)^2+1} \quad , \quad 1$$

$$\sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum\left(x_i - \left(\frac{\sum x_i}{n}\right)\right)^2}$$

(1-2-5

$$\mathbb{E}_q(\tilde{\mu}_{IS}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_q\left(h(x_i)\frac{P(x_i)}{q(x_i)}\right) = \mathbb{E}_q\left(h(x_i)\frac{P(x_i)}{q(x_i)}\right) =$$

$$= \int h(x)\frac{P(x)}{q(x)}q(x)\,dx = \int h(x)P(x)\,dx = E_p(h(x)) \rightarrow \text{Unbiased}$$

(2-2-5 * برای این قسمت فرض می کنیم $h(x)$ باید نامنفی باشد)

$$Var(\tilde{\mu}_{IS}) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left\{h(x_i)\frac{P(x_i)}{q(x_i)}\right\} = \frac{1}{n}Var\left\{\frac{h(x)P(x)}{q(x)}\right\}$$

$$Var\left\{\frac{h(x)P(x)}{q(x)}\right\} = \mathbb{E}\left\{\left(\frac{h(x)P(x)}{q(x)}\right)^2\right\} - \mu^2 = \int\frac{h(x)^2P(x)^2}{q(x)}dx - \mu^2$$

اگر $q = \frac{hP}{\mu}$ باشد. در نتیجه برای $Var(\tilde{\mu}_{IS})$ داریم که:

$$Var(\tilde{\mu}_{IS}) = \frac{1}{n}\left\{\mu\int h(x)P(x)\,dx - \mu^2\right\} = 0$$

اما با توجه به این که مقدار حقیقی μ را نداریم، فرض کنید که $q = \frac{hP}{c}$ باشد

که در آن c یک مقدار ثابت است.

$$Var\left\{\frac{h(x)P(x)}{q(x)}\right\} = \frac{1}{n}\left\{c\int h(x)P(x)\,dx - \mu^2\right\} \rightarrow = 0$$

$$c = \int h(x)P(x)\,dx$$

در نتیجه مقدار کمینه واریانس برابر خواهد شد با 0.

$$Var\left(\tilde{\mu}_{IS}\right) = \frac{1}{n}\left\{\int \frac{f^2 p^2}{q} dx - \left\{\int f p \, dx\right\}^2\right\}$$

نشان دادیم که Expected این تخمینگر دقیقاً برابر است با μ.

در نظر بگیرید که q در جاهایی که p مقدار دارد مقدار کمی و در جاهایی که p مقداری ندارد مقدار به شدت زیادی داشته باشد. در این شرایط می‌توان مقدار Variance را به مقدار مورد نیاز کم یا زیاد کرد.

$$\hat{\mu}_{N\text{-}IS} = \frac{\sum h(x_i) w(x_i)}{\sum w(x_i)} = \cancel{\frac{1}{n}\sum h(x_i) P(x_i)}$$

(3-2-5

$$\hookrightarrow \hat{\mu}_{N\text{-}IS} = \frac{\frac{1}{n}\sum h(x_i) \frac{P(x_i)}{q(x_i)}}{\frac{1}{n}\sum P(x_i)/q(x_i)}$$

$$\mathbb{E}_q\left\{\frac{1}{n}\sum h(x_i) \frac{P(x_i)}{q(x_i)}\right\} = \mathbb{E}_q\left\{h(x) \frac{P(x)}{q(x)}\right\} = \mu$$

$$\mathbb{E}_q\left\{\frac{1}{n}\sum \frac{P(x_i)}{q(x_i)}\right\} = \mathbb{E}_q\left\{\frac{P(x)}{q(x)}\right\} = \int \frac{P(x)}{q(x)} q(x) \, dx = 1$$

بدیهی است که

$$\mathbb{E}_q\left(\tilde{\mu}_{NIS}\right) = \mathbb{E}_q\left\{\frac{\frac{1}{n}\sum h(x_i) \times \frac{P(x_i)}{q(x_i)}}{\frac{1}{n}\sum \frac{P(x_i)}{q(x_i)}}\right\} \neq \frac{E_q\left\{\frac{1}{n}\sum h(x_i) w(x_i)\right\}}{\mathbb{E}_q\left\{\frac{1}{n}\sum w(x_i)\right\}} = \mu$$

در نتیجه این تخمینگر biased است.

s.a.m

با توجه به قانون اعداد بزرگ داریم که

$$\lim_{n\to\infty} \hat{\mu}_{N-IS} = \lim_{n\to\infty} \frac{\frac{1}{n}\sum h(x_i)\, w(x_i)}{\frac{1}{n}\sum w(x_i)} = \frac{\mathbb{E}_q\left(h(x)\, w(x)\right)}{\mathbb{E}_q\left(w(x)\right)} = \mu$$

سوال 5-2-4)

$m = \inf h(x)$, $M = \sup h(x)$ $\qquad \mu = \mathbb{E}(\hat{\mu}_{N-IS})$

$$\left.\begin{aligned} \tilde{\mu}_{N-IS} &= \frac{\sum w(x_i)\, h(x_i)}{\sum w(x_i)} \leq \frac{M \times \sum (w(x_i))}{\sum (w(x_i))} = M \\ \hat{\mu}_{N-IS} &= \frac{\sum w(x_i)\, h(x_i)}{\sum (w(x_i))} \geq \frac{m \times \sum (w(x_i))}{\sum w(x_i)} = m \end{aligned}\right\} \rightsquigarrow m \leq \hat{\mu}_{N-IS} \leq M$$

$$\hookrightarrow 0 \leq \mathbb{E}\left\{(M - \hat{\mu}_{N-IS})(\hat{\mu}_{N-IS} - m)\right\} = -\mathbb{E}(\hat{\mu}_{N-IS}^2) - mM + (m+M)\,\mathbb{E}(\hat{\mu}_{N-IS})$$

$$\sigma^2 = \mathbb{E}(\hat{\mu}_{N-IS}^2) - \mu^2 \leq -mM + (m+M)\mu - \mu^2 = (M-\mu)(\mu - m) \rightarrow$$

با توجه به نامساوی حسابی - هندسی داریم که:

$$\sigma^2 \leq (M-\mu)(\mu-m) \leq \left(\frac{M-\mu+\mu-m}{2}\right)^2 = \frac{(M-m)^2}{4}$$

سوال ۵-ج-۵-آ)

هیچ کدام از نمونه هایی که داریم در بازهٔ مشخص شده نیستند در نتیجه با استفاده از Monte-Carlo Estimation، مقدار تخمین ما برابر با 0 خواهد بود.

۵-ج-2-ب)

$p = N(0,1)$ و $q = N(0,1)$ ← بنابراین داده های سوال را باید یک شیفت قرار بدهیم.

$S' = \{3,126, 2,73, 1,44, 3,41, 3,40, 2,98, 3,10, 1,31, 2,72, 4,47\}$

که نمونه هایی هستند که از توزیع q داده گرفتیم ← داده هایی را در نظر باید گرفت که از 2 بزرگتر هستند.

$$\hat{H}_{IS} = \frac{1}{N} \sum_{\substack{x \in S' \\ x \geq 2}} \frac{e^{-x^2/2}}{\frac{1}{\sqrt{2\pi}} e^{\frac{-(x-3)^2}{2}}} = \frac{\sqrt{2\pi}}{N} \sum e^{4,5-3x}$$

$\hookrightarrow \hat{H}_{IS} = 0,02951$

مقدار اصلی H برابر است با $H = \sqrt{\frac{\pi}{2}}\, erfc(\sqrt{2}) = 0,057$

همانطور که بدست آوردیم با استفاده از روش دوم بهتر است و دلیل

دلیل آن می تواند این باشد که چون می خواهیم انتگرال را در بازهٔ 2 تا $+\infty$ حساب کنیم بهتر است که از توزیعی نمونه گیری کنیم که میانگین و واریانس حول این بازه داشته باشد. که $N(3,1)$ دقیقاً آخر یک قسمت آن 2 است. به همین دلیل تخمین بهتری به ما داده است.

سوال 6-آ)

$$-D_{KL}(P\|q) = -\int P(x) \log \frac{P(x)}{q(x)}\, dx = \int P(x) \log \frac{q(x)}{P(x)}\, dx$$

Using Jensen Inequality

$$\int P(x) \log \frac{q(x)}{P(x)}\, dx \leq \log\left\{\int P(x) \frac{q(x)}{P(x)}\, dx\right\} = 0 \Rightarrow$$

$$\Rightarrow D_{KL}(P\|q) \geq 0$$

با توجه به معادله بالا و استفاده از Jensen داریم که تساوی زمانی برقرار می‌شود که نسبت $\frac{P(x)}{q(x)}$ برابر با یک مقدار ثابت باشد. با توجه به اینکه $\int P(x)\, dx = \int q(x)\, dx = 1$ در نتیجه این مقدار ثابت برابر با 1 است و کمینه فاصله KL زمانی اتفاق می‌افتد که $P(x) = q(x)$ باشد $\Leftrightarrow D_{KL}(P\|q) = 0$

سوال 6-ب)

$$D_{KL}(q(z)\|P(z|X)) = \int q(z) \log \frac{q(z)}{P(z|X)}\, dz = \mathbb{E}_q\left(\log \frac{q(z)}{P(z|X)}\right)$$

$$= \mathbb{E}_q(\log q(z)) - \mathbb{E}_q\{\log P(z|X)\} = \mathbb{E}_q(\log q(z)) - \mathbb{E}_q\left\{\frac{\log P(X,z)}{P(X)}\right\}$$

$$= \mathbb{E}_q\{\log q(z)\} - \mathbb{E}_q \log(P(X,z)) + \mathbb{E}_q(\log P(X))$$

با توجه به اینکه X به توزیع $q(z)$ مربوط نیست داریم که $\mathbb{E}_q(\log P(X)) = \log P(X)$

Subject Date / /

$$= \log P_{(x)} - \underbrace{\left\{ \mathbb{E}_q\left(\log P_{(x,z)}\right) - \mathbb{E}_q\left(\log q_{(z)}\right) \right\}}_{L(q)} \Rightarrow$$

$$\Rightarrow D_{KL}\left\{ q(z) \| P(z|x) \right\} = \log P_{(x)} - L_{(q)} = \log P_{(x)} - \left\{ \mathbb{E}_q(\log P(x,z)) - E_q(\log q_z) \right\}$$

سؤال 6 - ج) (کمک از کتاب Bishop)

در نظر داشته باشیم که نا- به معنای همه چیز نا است. $q(z) = \prod q_i(z_i)$

هدف Variational Inference در واقع بیشینه کردن $L_{(q)}$ است که

$$L(q) = \mathbb{E}_q\left\{ \log \frac{P(x,z)}{q(z)} \right\} = \int q(z) \log\left\{ \frac{P(x,z)}{q(z)} \right\} dz$$

حال داریم که $q(z) = \prod q_i(z_i)$ و $dz = \prod_i dz_i$

$$L(q) = \int_z \prod_i \left\{ q_i(z_i) \right\} \left\{ \log P(x,z) - \sum_j \log q_j(z_j) \right\} \prod_i dz_i$$

$$= \int_z \prod_i \left\{ q_i(z_i) \right\} \left\{ \log P(x,z) - \sum_{i \neq j} \log q_j(z_j) - \log q_i(z_i) \right\} dz$$

$$= \int_{z_i} q_i(z_i) \left\{ \int_{z_{-i}} \prod_{i \neq j} \left\{ q_j(z_j) \right\} \left(\log P(x,z) - \log q_i(z_i) - \sum_{i \neq j} \log(q_j(z_j)) \right) \prod dz_j \right\} dz_i$$

$$= \int_{z_i} q_i(z_i) \left\{ \int_{z_{-i}} \prod_{i \neq j} \left(q_j(z_j) \right) \left(\log P(x,z) - \log q_i(z_i) \right) \prod dz_j \right\} dz_i$$

$$= \int_{z_i} q_i(z_i) \left\{ \int_{z_{-i}} \prod_{i \neq j} \left(q_j(z_j) \right) \left(\sum_{i \neq j} \log q_j(z_j) \prod dz_j \right) \right\} dz_i \quad \int q_i \, dz_i = 1$$

$$\int_{z_{-i}} \prod_{i \neq j} \left(q_j(z_j) \sum_{i \neq j} \log q_j(z_j) dz_j \right)$$ ← به q_i مربوط نیست

s.a.m

CamScanner

برای قسمت اول نیز داریم که

$$\int_{z_i} q_i(z_i) \left\{ \int_{z_{-i}} \prod_{i\neq j} \left(q_j(z_j)\right) \left(\log(P(x,z)) - \log q_i(z_i) \prod_{i\neq j} dz_j \right) \right\} dz_i =$$

$$\cancel{= \int_{z_i} \prod_{i\neq j} q_j(z_j) \left(\log(P(x,z)) - \log q_i(z_i) \right.}$$

$$\cancel{= \int_{z_i} \prod} = \int_{z_i} q_i(z_i) \left\{ \int_{z_{-i}} \prod_{i\neq j} \left(q_j(z_j)\right) \left(\log P(x,z)\right) \prod_{i\neq j} dz_j \right\} dz_i$$

$$- \int q_i(z_i) \log q_i(z_i)\, dz_i =$$

$$= \int_{z_i} q_i(z_i) \times \mathbb{E}_{q_{-i}} \left\{ \log(P(x,z)) \right\} dz_i - \int q_i(z_i) \log q_i(z_i)\, dz_i$$

حال با جای گذاری در رابطه اصلی داریم که:

$$\cancel{L(q) = \int q_i(z_i) \, dz}$$

$$L(q) = \int_{z_i} q_i(z_i)\, \mathbb{E}_{q_{-i}} \left\{ \log(P(x,z)) \right\} dz_i - \int q_i(z_i) \log q_i(z_i)\, dz_i$$

$$- \underbrace{\int_{z_{-i}} \prod_{j\neq i} q_j(z_j) \times \sum_{j\neq i} \log q_j \times \prod_{j\neq i} dz_j}_{H(\prod_{j\neq i} q_j)}$$

حال باید این تابع را ماکسیمم کنیم

به شرط اینکه $\int q_i \, dz_i = 1$ باشد

برای اینکار از ضرایب لاگرانژ استفاده خواهیم کرد.

s.a.m

maximize $L(q)$ s.t. $\forall i: \int q_i(z_i)\, dz_i = 1$

$\hookrightarrow$ Langrangian $= L(q) + \sum \lambda_i \left\{ \int q_i(z_i)\, dz_i - 1 \right\}$

با استفاده از Euler-Lagrangian Equation

$$\hookrightarrow \frac{\partial L}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \left(\frac{\partial L}{\partial q_{i,j}} \right) = 0$$

این معادله بهینه سازی را حل می کنیم.

داریم که $\frac{\partial L}{\partial q_i} = 0$ باید باشد، زیرا هیچ مشتقی در رابطه ما ظاهر نشده است.

$$\hookrightarrow \frac{\partial L}{\partial q_i} = \frac{\partial}{\partial q_i} \left\{ q_i \mathbb{E}_{q_{-i}} \left\{ \log P(x,z) \right\} - q_i \log q_i + \lambda_i q_i \right\} = 0$$

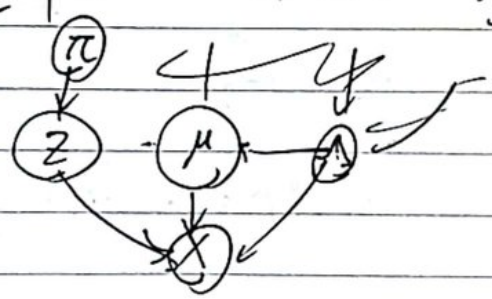
$$\hookrightarrow \mathbb{E}_{q_{-i}} \left\{ \log P(x,z) \right\} - \log q_i - 1 + \lambda_i = 0 \rightsquigarrow$$

$$\rightsquigarrow \log q_i(z_i) = \mathbb{E}_{q_{-i}} \left\{ \log P(x,z) \right\} + \text{Const.}$$

سؤال 6-د) (کمک از کتاب Bishop)

$$P(x, z, \mu, \pi, \Lambda) = P(\pi) P(\mu, \Lambda) P(z|\pi) P(x|z, \mu, \Lambda)$$

داریم که $\ln q^*(z) = \mathbb{E}_{\pi, \mu, \Lambda} \left\{ \log P(x, z, \pi, \mu, \Lambda) \right\} + \text{Const.}$



$$\ln q^*(z) = \mathbb{E}_{\pi,\mu,\Lambda}\left\{ \log P(x|z,\mu,\Lambda) + \log P(z|\pi) + \log P(\mu,\Lambda) + \log P(\pi) \right\} + Const$$

$$= \mathbb{E}_{\pi,\mu,\Lambda}\left\{ \log P(x|z,\mu,\Lambda) \right\} + \mathbb{E}_{\pi,\mu,\Lambda}\left\{ \log P(z|\pi) \right\} + \mathbb{E}_{\pi,\mu,\Lambda}\left\{ \log P(\mu,\Lambda) \right\} + \mathbb{E}_{\pi,\mu,\Lambda}\left\{ P(\pi) \right\} + Const \Rightarrow$$

$$\Rightarrow q^*(z) = \mathbb{E}_{\mu,\Lambda}\left\{ \log P(x|z,\mu,\Lambda) \right\} + \mathbb{E}_{\pi}\left\{ \log P(z|\pi) \right\} + Const$$

$$\mathbb{E}_{\pi}\left\{ \log P(z|\pi) \right\} = \mathbb{E}_{\pi}\left\{ \log \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \right\} = \sum_n \sum_k z_{nk} \mathbb{E}_{\pi}\left\{ \log \pi_k \right\}$$

$$\mathbb{E}_{\mu,\Lambda}\left\{ \log P(x|z,\mu,\Lambda) \right\} = \mathbb{E}_{\mu,\Lambda}\left\{ \log \left\{ \prod_n \prod_k \mathcal{N}(x_n|\mu_k,\Lambda_k^{-1})^{z_{nk}} \right\} \right\} =$$

$$= \sum_k \sum_n \mathbb{E}_{\mu,\Lambda}\, z_{nk} \left\{ \log \left(\frac{1}{\sqrt{(2\pi)^D |\Lambda_k^{-1}|}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Lambda_k (x_n-\mu_k)} \right) \right\} =$$

$$= \sum_k \sum_n \cancel{\mathbb{E}_{\mu,\Lambda}}\, \frac{1}{2} \left\{ \frac{1}{2} \mathbb{E}_{\Lambda_k}\left(\ln |\Lambda_k^{-1}| \right) - \frac{1}{2} D \ln(2\pi) + \frac{1}{2} \mathbb{E}_{\mu_k,\Lambda_k}\left((x_n-\mu_k)^T \Lambda_k (x_n-\mu_k) \right) \right.$$

تعریف می کنیم

$$\ln \rho_{nk} = \mathbb{E}_{\pi}\left\{ \log \pi_k \right\} + \frac{1}{2} \mathbb{E}\left\{ \ln |\Lambda_k| \right\} - \frac{D}{2} \ln(2\pi) - \frac{1}{2} \mathbb{E}_{\mu_k,\Lambda_k}\left\{ (x_n-\mu_k)^T \Lambda_k (x_n-\mu_k) \right\}$$

s.a.m

در نتیجه:

$$\ln q^*(z) = \sum_n \sum_k z_{nk} \ln \rho_{nk} + Const$$

چون داریم که $\sum_k z_{nk} = 1$

$$q^*(z) = \prod_n \prod_k \left(\frac{\rho_{nk}}{\sum_j \rho_{nj}}\right)^{z_{nk}} \rightsquigarrow \sum_k \left(\frac{\rho_{nk}}{\sum_j \rho_{nj}}\right) = 1$$

تخمین Variational برای $q(z)$ دقیقاً مانند $P(z|\pi)$ که همان prior است، می‌باشد. مانند بخش قبل

$$\ln q^*(\pi, \mu, \Lambda) = \mathbb{E}_z\{\ln P(X|Z, \mu, \Lambda)\} + \mathbb{E}_z\{\ln P(Z|\pi)\} + \mathbb{E}_z\{\ln P(\pi)\} + \mathbb{E}_z\{\ln P(\mu, \Lambda)\} + Const =$$

$$= \ln P(\pi) + \mathbb{E}_z\{\ln P(Z|\pi)\} + \mathbb{E}_z \ldots\ldots$$

در نتیجه می‌توان نوشت که

$$q(\pi, \mu, \Lambda) = q(\pi) \times \prod_k q(\mu_k, \Lambda_k)$$

برای بدست آوردن $q(\pi)$

$$\ln q^*(\pi) = \ln P(\pi) + \mathbb{E}_z\{\ln P(Z|\pi)\} =$$

$$= (\alpha_0 - 1)\sum_k \ln \pi_k + \sum_n \sum_k \left(\frac{\rho_{nk}}{\sum_j \rho_{nj}}\right) \ln \pi_k + Const$$

$$q^*(\pi) = Dir(\pi | \alpha)$$

که در آن

$$N_k = \sum_n \left(\frac{\rho_{nk}}{\sum_j \rho_{nj}}\right), \quad \alpha_k = \alpha + N_k$$

s.a.m

~~برای~~ لطفا عبارت روبرو را به این شکل تعریف می کنیم

$$r_{nk} = \frac{\rho_{nk}}{\sum_j \rho_{nj}} \qquad \bar{X}_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} X_n$$

$$S_{nk} = \frac{1}{N_k} \sum^{N} r_{nk} (X_n - \bar{X}_k)(X_n - \bar{X}_k)^T$$

که مانند mean و variance هستند حال به ادامه حل می‌پردازیم.

$$\ln q^*(\mu_k, \Lambda_k) = \ln P(\mu_k, \Lambda_k) + N_k \ln \left\{ \frac{1}{\sqrt{(2\pi)^D |\Lambda_k^{-1}|}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Lambda_k (x_n-\mu_k)} \right\} + Const =$$

$$= \ln P(\mu_k, \Lambda_k) + N_k \ln N(x_n | \mu_k, \Lambda_k^{-1}) + Const =$$

$$= \ln N(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) + \ln W(\Lambda_k | W_0, \nu_0)$$

$$+ N_k \ln N(X_n | \mu_k, \Lambda_k^{-1}) + Const$$

$$\hookrightarrow q^*(\mu_k, \Lambda_k) = N(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) W(\Lambda_k | W_k, \nu_k)$$

$$\beta_k = \beta_0 + N_k \quad , \quad m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k \bar{X}_k)$$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{X}_k - m_0)(\bar{X}_k - m_0)^T$$

$$\nu_k = \nu_0 + N_k$$

برای حل این سوال از کتاب Bishop کمک گرفته شده است

s.a.m