(1)

تمرین سری ۱ - درس یادگیری عمیق   نیما المسیری - ۹۸۱۰۸۱۲۴

۱)

الف - بنابه تعریف نرم F داریم:

$$\|A\|_F^2 = \sum_{i,j \in m \times n} a_{ij}^2 \longrightarrow \|UA\|_F^2 = Tr\,(UA)^T(UA)$$

$$= Tr\, A^T \cancel{U^T U} A \overset{I}{=} \|A\|_F^2$$

به طور مشابه برای $\|AV\|_F$ نیز داریم یعنی:

$$\|AV\|_F^2 = \|A\|_F$$

$$\sqrt{\sigma_1^2+\dots+\sigma_r^2} \le \sqrt{\sigma_{max}^2+\dots+\sigma_{max}^2} \Rightarrow \sqrt{r\sigma_{max}^2} = \sigma\sqrt{r}$$

$$\|A\|_F \xrightarrow{SVD} \|U^T \Sigma V\|_F = \|\Sigma V\|_F = \sqrt{\sum_{i \le r} \sigma_i^2} \rightarrow \sigma_{max}^2 \ge \sigma_i^2,\ i \le r$$

$$\boxed{\sqrt{\sum_{i \le r} \sigma_i^2} \le \sigma_{max}\sqrt{r}} \ *$$

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} \Longrightarrow \sigma_i \ge 0,\ i \le r \rightarrow \sigma_{max}^2 \le \sum_{i \le r} \sigma_i^2 \Rightarrow \boxed{\sigma_{max} \le \sqrt{\sigma_1^2+\dots+\sigma_r^2}} \ **$$

$$\left.\begin{matrix} * \\ ** \end{matrix}\right\} \|A\|_2 \le \|A\|_F \le \sqrt{rank(A)}\,\|A\|_2$$

ب - i.

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \xrightarrow{X>0} E(X) = \int_0^{\infty} x f(x)\,dx$$

$$X = [0, a] \cup [a, \infty) \longrightarrow E(X) = \int_0^a x f(x)\,dx + \int_a^{\infty} x f(x)\,dx$$

$$E(X) \ge \int_a^{\infty} x f(x)\,dx \ge \int_a^{\infty} a f(x)\,dx \ge a \underbrace{\int_a^{\infty} f(x)\,dx}_{P(X \ge a)}$$

$$\Rightarrow E(X) \ge a\,P(X \ge a)$$

$$\longrightarrow P(X \ge a) \le \frac{E(X)}{a} \ \blacksquare$$

ii.

نامساوی مارکوف $\longrightarrow P(Y \ge a) \le \frac{E(Y)}{a}$, $a > 0$

$\cancel{Y = (X-\mu)^2}$, $\cancel{a = (k\sigma)^2}$

$$\left.\begin{matrix} Y = (X - EX)^2 \\ a = \varepsilon^2 \end{matrix}\right\} \longrightarrow$$ Y متغیر تصادفی نامنفی می شود

$$P(Y \ge \varepsilon^2) \le \frac{E(Y)}{\varepsilon^2}$$

$$E(Y) = E(X - EX)^2 = \overset{\text{Variance}(X)}{\sigma^2}$$

$$P(Y \ge \varepsilon^2) = P((X - E(X))^2 \ge \varepsilon^2) = P(|X - E(X)| \ge b) \quad , \quad E(X) = \mu$$

$$\Rightarrow P(|X - \mu| \ge \varepsilon) \le \frac{\sigma^2}{\varepsilon^2} \ \blacksquare$$

SOBHAN

تمرین سری ۱ - یادگیری عمیق

۱۷. احتمال اینکه نقطه تصادفی ما در محل درست واقع شود برابر مساحت دایره تقسیم بر مساحت کل است ، پس :

$$P = \frac{4\pi}{16} = \boxed{\frac{\pi}{4}}$$

برای هر i

$$E(X_i) = \frac{\pi}{4} \cdot 1 + \left(1 - \frac{\pi}{4}\right) \cdot 0 = \frac{\pi}{4}$$

$$V(X_i) = \frac{\pi}{4} \cdot \left(1 - \frac{\pi}{4}\right) \longrightarrow$$ با داشتن $\hat{\pi}(n)$ تا اینجا داریم :

$$E(\hat{\pi}) = E\left(\frac{4}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\sum_{i=1}^{n} E(X_i) = \pi$$

$$V(\hat{\pi}) = V\left(\frac{4}{n}\sum_{i=1}^{n} X_i\right) = \frac{16}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\pi(4-\pi)}{n}$$

$$P(|\hat{\pi}(n) - \pi| < \delta) > \varepsilon$$

$$P(|\hat{\pi}(n) - \pi| \geq \delta) \leq 1 - \varepsilon$$

$$\left.\begin{aligned} \varepsilon &= 0.98 \\ \delta &= 0.01 \end{aligned}\right\} \Rightarrow P(|\hat{\pi}(n) - \pi| \geq 0.01) \leq 1 - 0.95 = 0.05$$

از چبیشف می‌دانیم $V(\hat{\pi})/a^2 \leq 0.05$ برقرار است . پس خواهیم داشت :

$$\frac{V(\hat{\pi})}{a^2} = \frac{\pi(4-\pi)}{n(0.01)^2} \leq 0.05$$

$$\Rightarrow n \geq 539353.24 \Rightarrow \boxed{n \geq 539,354}$$

یعنی حداقل باید حدود ۵۴۰ هزار نقطه انتخاب کنیم .

یادگیری عمیق - تمرین سری اول

نیما السیدی ۹۸۱۰۸۱۲۴۰

۲) الف -

$$a^T x = a_1 x_1 + \cdots + a_n x_n$$

$$\frac{\partial a^T x}{\partial x} = \left(\frac{\partial a^T x}{\partial x_1}, \dots, \frac{\partial a^T x}{\partial x_n}\right) = (a_1, \dots, a_n) = \boxed{a^T}$$

ب -

$$x^T A x = \left([a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n], \dots, [a_{1n}x_n + \cdots + a_{nn}x_n]\right)(x_1, \dots, x_n)^T$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j \longrightarrow S$$

$$\frac{\partial x^T A x}{\partial x} = \frac{\partial S}{\partial x} = \left(\frac{\partial S}{\partial x_1}, \dots, \frac{\partial S}{\partial x_n}\right) = \sum_{j=1}^{n} (a_{ij} + a_{ji}) x_j$$

$$\Rightarrow \frac{\partial x^T A x}{\partial x} = \boxed{x^T (A + A^T)}$$

پ -

$$A^{-1}A \rightarrow I \Rightarrow \text{ثابت}$$

$$\frac{\partial A^{-1}A}{\partial \beta} = \frac{\partial I}{\partial \beta} = \left(\frac{\partial A^{-1}}{\partial \beta} A + A^{-1}\frac{\partial A}{\partial \beta}\right) = 0$$

$$\Rightarrow \frac{\partial A^{-1}}{\partial \beta} A = -A^{-1}\frac{\partial A}{\partial \beta} \xrightarrow[\text{از راست}]{A^{-1}\text{ضرب}} \frac{\partial A^{-1}}{\partial \beta} = \boxed{-A^{-1}\frac{\partial A}{\partial \beta} A^{-1}}$$

ت - K را مجموعه اعداد صحیح مثبت (ساده) به جز n در نظر می‌گیریم:

$$|A| = \sum_{j=1}^{n} a_{kj} C_{kj} \rightarrow \frac{\partial |A|}{\partial a_{ij}} = C_{ij}$$

$$C = \frac{\partial |A|}{\partial A} \longrightarrow A^{-1} = |A^{-1}| C \Rightarrow \nabla_A |A| = \boxed{|A| A^{-1}}$$

$$A^{-1} \xrightarrow{\text{طبق تعریف}} \nabla_A^{-1} |A| A^{-1} = |A|^{-1} \nabla_A |A| \xrightarrow{\text{خاصیت لگاریتم}} \nabla_A \log |A|$$

$$\Rightarrow \nabla_A \log |A| = \boxed{A^{-1}}$$

SOBHAN

تمرین سری ۱ - یادگیری عمیق — نیما لمیدی - ۹۸۱۰۸۱۲۳

۳) الف) میدانیم معادله مشخصه ماتریس A که $n\times n$ است به صورت زیر می‌باشد:

$$P(t) = \det(A - tI) = (-1)^n (t^n - (tr(A))\, t^{n-1} + \cdots + (-1)^n \det(A) \quad *$$

از طرفی داریم:

$$P(t) = (-1)^n (t - \lambda_1) \cdots (t - \lambda_n) \rightarrow \boxed{\lambda_j \rightarrow \text{مقادیر ویژه}} \quad **$$

ملذا با تناظر ضرایب این دو برابری، خواهیم داشت:

$$trace(A) = \lambda_1 + \cdots + \lambda_n$$

ب)

$$\det(A - \lambda I) = P(\lambda) \longleftarrow \text{بنابر تعریف}$$

$$P(\lambda) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

$$= (-1)(\lambda - \lambda_1) \cdots (-1)(\lambda - \lambda_n) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda)$$

$$\lambda = 0 \longleftarrow \text{حال فرض می‌کنیم}$$

پس خواهیم داشت:

$$\det(A - \lambda I) = (\lambda_1 - 0)(\lambda_2 - 0) \cdots (\lambda_n - 0)$$

$$\Rightarrow \det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$$

که حکم را اثبات می‌کند.

تمرین ۱ – یادگیری عمیق

۴) الف – میدانیم که وقتی A^TA وارون پذیر است که بردارهای ستون ها مستقل باشند (ماتریس رنک ستونی کامل داشته باشد.

بنابراین $(A^TA)^{-1}$ وجود دارد و ماتریس وارون آن را در دو طرف ضرب کنیم:

$$A^+ = (A^TA)^{-1}A^T \Rightarrow A^TAA^+ = A^T$$

* از آنجایی که رابطه های استفاده شده در ادامه برابر هستند، می‌توانیم بگوییم اگر A و A^+ را تجزیه کرده و ساده کنیم و باز هم رابطه برقرار بماند، پس فرض اولیه درستی داشته ایم. داریم:

$$A = U\Sigma V^T$$

$$A^+ = V\Sigma^{-1}U^T$$

$$A^T = (U\Sigma V^T)^T = V\Sigma^TU^T$$

خواص ماتریس های U و V: $\begin{cases} VV^T = I \\ U^TU = I \end{cases}$

$$\rightarrow A^TAA^+ = A^T \Rightarrow (V\Sigma^T\underbrace{U^TU}_{I}\Sigma \underbrace{V^T)V}_{I}\Sigma^{-1}U^T = V\Sigma^TU^T$$

$$\Rightarrow V\Sigma^T\underbrace{\Sigma\Sigma^{-1}}_{I}U^T = V\Sigma^TU^T$$

که این برابری درستی عبارت را نتیجه می دهد.

ب – مشابه قبل، هرگاه AA^T وارون پذیر باشد، لازم است بردارهای سطرها مستقل باشد (ماتریس رنک سطری کامل داشته باشد.

پس $(AA^T)^{-1}$ وجود دارد و باز هم از سمت راست وارون آن را با A در دو طرف ضرب می کنیم:

$$A^+ = A^T(AA^T)^{-1} \rightarrow A^+AA^T = A^T$$

$$\Rightarrow V\Sigma^{-1}\overbrace{U^T(U}^{I}\Sigma \overbrace{V^TV}^{I}\Sigma^TU^T) = V\Sigma^TU^T$$

$$\rightarrow V\underbrace{\Sigma^{-1}\Sigma}_{I}\Sigma^TU^T = V\Sigma^TU^T$$

که باز هم برابری صحیح درآمده و حکم اثبات است.

۵) الف – طبق تجزیه ی مکمل شور داریم: (هرگاه A^{-1} وجود داشته باشد)

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad M/A = D - CA^{-1}B$$

که M بصورت زیر تجزیه می شود:

$$M = \begin{bmatrix} I & 0 \\ CA & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & M/A \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

و با جایگذاری M/A خواهیم داشت:

$$M = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

درستی این تجزیه مشخص است، طبق قانون دترمینان ماتریس بلوکی به یاد داریم که

$$\det(M) = \det(AD - BC)$$

پس خواهیم داشت:

$$\det(M) = \det(II-0)\det(A(D-CA^{-1}B))\det(II-0) = 1 \times \det(A(D-CA^{-1}B)) \times 1$$

و از قوانین دترمینان:

$$\det(M) = \det(A)\det(D-CA^{-1}B)$$

تمرین ۱ – یادگیری عمیق

۵) ب – مانند بخش قبل با این بار با وارون پذیر در نظر گرفتن D، ماتریس M را به شکل متفاوت تجزیه می‌کنیم:

$$M = \underbrace{\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}}_{det \to 1} \begin{bmatrix} \overbrace{A - BD^{-1}C}^{M/D} & 0 \\ 0 & D \end{bmatrix} \underbrace{\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}}_{det \to 1} \rightarrow det(M) = 1 \times det((A - BD^{-1}C)D - 0 \times 0) \times 1$$

$$\Rightarrow det(M) = det(D)\, det(A - BD^{-1}C)$$

پس این حکم نیز اثبات است.

پ – برای این بخش، از وارون M استفاده می‌کنیم و در نهایت آن از ملکل سئوال استفاده خواهد شد.

$$M = \begin{bmatrix} A & -B \\ C & D \end{bmatrix}$$

$$M^{-1} = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

$$M^{*}M^{-1} = I$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE - BG & AF - BH \\ CE + DG & CF + DH \end{bmatrix} = \begin{bmatrix} I_A & 0 \\ 0 & I_D \end{bmatrix}$$

پس ۴ معادله داریم:

$$\begin{cases} AE - BG = I_A & (1) \\ AF - BH = 0 & (2) \\ CE + DG = 0 & (3) \\ CF + DH = I_D & (4) \end{cases}$$

$E = A^{-1} + A^{-1}BG$ (5) $\rightarrow$ با فرض وارون پذیر بودن A

(5)→(3) $\rightarrow CA^{-1} + CA^{-1}BG + DG \rightarrow (D - CA^{-1}B)G = -CA^{-1}$

$\rightarrow G = -(D + CA^{-1}B)^{-1}CA^{-1}$ (6)

(5)+(6) $\longrightarrow \boxed{E = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}} \longleftarrow$ قسمت اول *

یک جواب برای E بدست آوردیم. حال جواب دیگر با فرض وارون پذیر بودن D بصورت زیر بدست می‌آید:

$H = D^{-1} - D^{-1}CF$ (7)

(7)→(2) $\rightarrow AF - BD^{-1} + BD^{-1}CF = 0 \rightarrow \overbrace{(A + BD^{-1}C)}^{M/D}F = BD^{-1}$

$F = +(M/D)^{-1}BD^{-1}$ $\qquad H = D^{-1} - D^{-1}C(M/D)^{-1}BD^{-1}$

$G = -D^{-1}CE$ (8) $\qquad$ (8) (1) $\rightarrow AE + BD^{-1}CE = I_A \Rightarrow (A + BD^{-1}C)E = I_A$

با فرض وارون پذیر بودن $\longrightarrow \boxed{E = (A + BD^{-1}C)^{-1}} \longleftarrow$ قسمت دوم * *

و از این دو برابری، حکم سوال اثبات می‌شود.

تمرین سری اول - یادگیری عمیق

/ /

۵) ت - برای اثبات حکم ابتدا در حالت خاص فرض را برابر میگذاریم که $A = I$ برقرار است. داریم:

$$\begin{bmatrix} I & 0 \\ v^T & 1 \end{bmatrix} \begin{bmatrix} I + uv^T & u \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ -v^T & 1 \end{bmatrix} = \begin{bmatrix} I & u \\ 0 & 1+v^Tu \end{bmatrix}$$

با توجه به قوانین دترمینان به اشاره شده داریم:

$$\det(I + uv^T) = \det(I + v^Tu) = 1 + v^Tu$$

~~det(A+~~

پس برای دترمینان هر ماتریس به شکل $(I + uv^T)$ داریم:

$$\det(I + uv^T) = 1 + v^Tu \quad (1)$$

$$-p = A^{-1}u$$

$$A - uv^T = A(I - A^{-1}uv^T) \qquad \det(I + pv^T) \to 1 + v^Tp$$

$$\Rightarrow \det(A - uv^T) = \det(A)\det(I - \underbrace{A^{-1}u}_{-p}v^T) \xrightarrow{(1)} (1 - v^T\underbrace{A^{-1}u}_{-p})$$

$$\Rightarrow \det(A - uv^T) = \det(A)(1 - v^TA^{-1}u)$$

که این درستی حکم سوال را به ما نتیجه می‌دهد.

۶) الف) از بسط کوشی به صورت زیر استفاده می‌کنیم:

$$p_A(t) = \det\begin{pmatrix} tI_n - B & -x \\ -y^* & t-a \end{pmatrix}$$

$$= (t-a)\det(tI_n - B) - y^*(\operatorname{adj}(tI_n - B))x$$

پس می‌توان اینطور نوشت:

$$p_A(t) = (t-a)\,p_B(t) - y^*(\operatorname{adj}(tI_n - B))\,x$$

SOBHAN

تمرین سری ۱ - یادگیری عمیق

(۶

ب - بنابر قضیه کورانت - فیشر، زیرفضاهای S_A و S_B از F^n وجود دارند که

$$\begin{cases} \dim S_A = k+j \\ \dim S_B = n-j \end{cases}$$

$$\begin{cases} \lambda_{k+j}(A) = \max\limits_{\substack{x \in S_A \\ \|x\|=1}} \langle Ax, x\rangle \\ \lambda_{n-j}(B) = \max\limits_{\substack{x \in S_B \\ \|x\|=1}} \langle Bx, x\rangle \end{cases}$$

میدانیم $\dim(S_A \cap S_B) \ge k$

اگر این را به روی کورانت - فیشر اعمال کنیم، خواهیم داشت:

$$\lambda_k(A+B) = \min_{\dim S=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle (A+B)x, x\rangle \le \min_{\substack{\dim S=k \\ S \subseteq S_A \cap S_B}} \max_{\substack{x \in S \\ \|x\|=1}} \left(\langle Ax, x\rangle + \langle Bx, x\rangle\right)$$

$$\le \min_{\substack{\dim S=k \\ S \subseteq S_A \cap S_B}} \left(\max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle + \max_{\substack{x \in S \\ \|x\|=1}} \langle Bx, x\rangle\right) = \lambda_{k+j}(A) + \lambda_{n-j}(B)$$ و $j=0,1,\dots,n-1$

که این درستی حکم را نتیجه می‌دهد.

ب - فرض می‌کنیم $x=(y_1,\dots,y_m,0,\dots,0)\in F^n, y \in F^m$ و همچنین $\langle Bx, x\rangle = \langle Ax, x\rangle$ برقرار است.

و همچنین $\|x\|=\|y\|$. بنابر بخش الف و ب داریم:

$$\lambda_k(A) = \min_{\substack{S \subseteq F^n \\ \dim S=k}} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle \le \min_{\substack{S \subseteq \operatorname{span}(e_1,\dots,e_m) \subseteq F^n \\ \dim S=k}} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle$$

$$= \min_{\substack{S \subseteq F^m \\ \dim S=m-k+1}} \max_{\substack{y \in S \\ \|y\|=1}} \langle By, y\rangle = \lambda_k(B)$$

حال وقتی $m=n-1$ قرار دهیم، به طور خاص به ما می‌دهد:

$$\lambda_1(A) \le \lambda_1(B) \le \lambda_2(A) \le \lambda_2(B) \le \dots \le \lambda_{n-1}(B) \le \lambda_n(A)$$

که حکم نهایی سوال را اثبات می‌کند.

/ /

تمرین سری اول - یادگیری عمیق

۷)

$$f(x) = \frac{1}{\theta^2} x e^{\frac{-x}{\theta}} \quad , \quad 0 < \theta < \infty$$

$$L(\theta) = \frac{1}{\theta^{2n}} \cdot e^{\sum_{i=1}^{n} -\frac{x_i}{\theta}} \quad (x_1 \cdots x_n) \quad , \quad l(\theta) = -2n \log\theta - \frac{1}{\theta} n\bar{x} + \underbrace{\sum_{i=1}^{n} \log(x_i)}_{\text{به } \theta \text{ وابسته نیست}}$$

$$l'(\theta) = \frac{-2n}{\theta} + \frac{n}{\theta^2}\bar{x} \longrightarrow l'(\hat{\theta}) = 0 \to \frac{2n}{\theta} = \frac{n}{\theta^2}\bar{x} \Rightarrow \hat{\theta} = \frac{\bar{x}}{2}$$

که این به تخمین گر بهینه برای تابع $f(x)$ است.

۸) در ابتدا تعریف می‌کنیم $\{x_i\}_{i=1}^{N} = S$ ، $x_i \sim N(\mu, \sigma^2)$

$P(S|\mu) \xrightarrow{\text{max}} MLE$

$$P(S|\mu) = \prod_{i=1}^{N} P(x_i|\mu) = \log\{P(S|\mu)\} \propto \sum_{i=1}^{N} \frac{-1}{\sigma^2} \cancel{\frac{}{2}} (\mu - x_i)^2$$

برای MLE داریم:

$$\to \frac{\partial \log(P(S|\mu))}{\partial\mu} = 0 \Rightarrow \boxed{\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i}$$

$$P(\mu|S) \propto P(S|\mu) \cdot P(\mu)$$

$$\begin{cases} P(\mu) \sim N(\gamma, \beta^2) \Rightarrow \log\{P(\mu|S)\} \propto \sum_i^N \frac{1}{2\sigma^2}(\mu - x_i)^2 + \frac{1}{2\beta^2}(\mu - \gamma) \\ \Rightarrow \frac{\partial \log\{P(\mu|D)\}}{\partial\mu} = 0 \Rightarrow \boxed{\mu_{MAP} = \frac{\beta^2}{N\beta^2+\sigma^2}\sum_i x_i + \frac{\sigma^2}{N\beta^2+\sigma^2}\gamma} \end{cases}$$

برای MAP داریم:

حال رفتارشان را در بینهایت بررسی می‌کنیم:

$$\mu_{N\to\infty} \to \mu_{MAP} = \frac{1}{N}\sum_i^N x_i + 0 = \frac{\sum^N x_i}{N} = \mu_{ML}$$

پس با هم برابر می‌شوند.

تمرین سری ۱ - یادگیری عمیق

۱) الف)

$z = a + Ab$

$A = -\Sigma_{ab}\Sigma_{bb}^{-1}$

$Cov(z, b) = Cov(a, b) + Cov(Ab + b) = \Sigma_{ab} + A\,Var(b) = \Sigma_{ab} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{bb} = 0$

پس z و b مستقلند. پس واضحا داریم:

$E(z) = \mu_a + A\mu_b$

$\mu_{a|b} = E(x_a|x_b) = E(z - Ax_b|x_b)$

$= E(z|x_b) - E(Ax_b|x_b) = E(z) - Ab = \mu_a + A(\mu_b - x_b) = \boxed{\mu_1 + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)}$ بخش اول

$Var(x_a|x_b) = Var(z - Ax_b|x_b) = Var(z|x_b) + Var(-Ax_b|x_b) - A\,cov(z, -x_b)$

$- cov(z, -x_b)A' = Var(z|x_b) = var(z)$

پس داریم:

$var(x_a|x_b) = var(z) = Var(x_a + Ax_b)$

$= Var(x_a) + A\,var(x_b)A' + A\,cov(x_a, x_b) + cov(x_b, x_a)A'$

$= \Sigma_{aa} + \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{bb}\Sigma_{bb}^{-1}\Sigma_{ba} - 2\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \Sigma_{aa} + \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} - 2\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$

$= \boxed{\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}}$

پس و بدین صورت دو حکم مسئله اثبات می‌شود.

ب) از امید ریاضی و تعریف آن استفاده می‌کنیم:

$E(x_a) = \int_{-\infty}^{+\infty} E(x_a|x_b = t)\,P(x_b = t)\,dx$

$= \int_{-\infty}^{\infty} \mu_a P(x_b = t)\,dx - \int_{-\infty}^{+\infty} \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}(t - \mu_b)P(x_b = t)\,dx$

$= \mu_1 - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\underbrace{\left(\int_{-\infty}^{+\infty} t\,P(x_b = t)\,dx - \mu_b\int_{-\infty}^{+\infty} P(x_b = t)\,dx\right)}_{0} = \mu_1$

بنابراین امید ریاضی x_a برابر μ_1 می‌شود ①

$Var(x_a) = \underbrace{E[var(x_a|x_b)]}_{①} + \underbrace{var(E[x_a|x_b])}_{②}$ ← تعریف

① $E[Var(x_a|x_b)] = E(\Sigma_{aa} - \Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba}) = \Sigma_{aa} - \Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba}$ از بخش قبل

$var(E[x_a|x_b]) = var(\mu_a + \Sigma_{aa}\Sigma_{bb}^{-1}(x_b - \mu_b)) = var(\Sigma_{aa}\Sigma_{bb}^{-1}x_b) = \Sigma_{aa}\Sigma_{bb}^{-1}Var(x)\Sigma_{bb}^{-1}\Sigma_{aa} = \Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{bb}\Sigma_{bb}^{-1}\Sigma_{ba}$

② $= \Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba} \longrightarrow Var(x_a) = ① + ② = \Sigma_{aa} - \cancel{\Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba}} + \cancel{\Sigma_{aa}\Sigma_{bb}^{-1}\Sigma_{ba}} = \Sigma_{aa} \longrightarrow \boxed{x_a = N(\mu_a, \Sigma_{aa})}$

از بخش قبل

SOBHAN

تمرین سری ۱ - یادگیری عمیق

/ /

۱) الف - ابتدا $\|Ax-b\|^2$ را به شکل زیر می‌نویسیم:

$$\|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$

$$= (x^TA^T-b^T)(Ax-b) = x^TA^TAx - x^TA^Tb - b^TAx - b^Tb$$

پس خواهیم داشت:

$$\|Ax-b\|^2 = x^TA^TAx - 2x^TA^Tb - b^2$$

با مشتق‌گیری از آن خواهیم داشت:

$$\frac{\partial\|Ax-b\|^2}{\partial x} = 2A^TAx^* - 2A^Tb = 0$$

بنا به فرض، می‌توانیم بگوییم:

$$A^TAx^* = A^Tb \Rightarrow x^* = (A^TA)^{-1}A^Tb$$

که در ~~تمرین قبلی~~ سوال‌های قبل، دیدیم بخش اول معادله بالا برابر با A^+ است. پس:

$$\boxed{x^* = A^+b}$$

ب -

$$x^{(t+1)} = x^{(t)} - VA^T(Ax^{(t)}-b)$$

با فرض همگرا بودن، می‌توان گفت $x^* = x^{(t)} = x^{(t+1)}$، پاسخ آن را بر اساس راه زیر بدست می‌آوریم:

$$VA^TAx^* + \alpha A^Tb = 0$$

$$XA^TAx^* = XA^Tb \Rightarrow \boxed{x^* = (A^TA)^{-1}A^Tb} \rightarrow x^* = A^+b$$

← نشان دهنده میل کردن x^* به A^+b در مقادیر بسیار زیاد می‌باشد.