تمرین سری اول "یادگیری عمیق" - علیرضا ملکی ۴۰۱۲۱۰۹۵۶ [redacted]

سوال ۱)

آ - حکم : $\|A\|_2 \leqslant \|A\|_F \leqslant \sqrt{\mathrm{rank}(A)}\ \|A\|_2$

$$\Rightarrow \underbrace{\|A\|_2^2}_{\lambda_{max}(AA^T)} \leqslant \underbrace{\|A\|_F^2}_{\sum_{i=1}^{N} \lambda_i(AA^T)} \leqslant \underbrace{\mathrm{rank}(A)}_{\leqslant N} \|A\|_2$$

$$\Rightarrow \lambda_{max}(AA^T) \leqslant \sum_{i=1}^{N} \lambda_i(AA^T) \leqslant N\, \lambda_{max}(AA^T)$$ حکم اثبات شد

ب)

۱- $$E[X] = \int_{-\infty}^{+\infty} x\, P(x)\, dx = \int_{\alpha}^{+\infty} x\, P(x)\, dx + \int_{-\infty}^{\alpha} x\, P(x)\, dx$$

$$\Rightarrow E[X] \geqslant \int_{\alpha}^{+\infty} x\, P(x)\, dx \geqslant \alpha \int_{\alpha}^{+\infty} P(x)\, dx = \alpha P(X \geqslant \alpha)$$

$$\Rightarrow E[X] \geqslant \alpha P(X \geqslant \alpha)$$

۲- حکم: $P(|z-\mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$

قسمت قبل $\Rightarrow P(|z-\mu|^2 \geqslant \varepsilon^2) \leqslant \frac{E((z-\mu)^2)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$

$$\Rightarrow P((z-\mu) \geqslant \varepsilon^2) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

۳. $$P\left(\left|Z-\frac{\pi}{٤}\right| \geq \frac{۰/۰۱}{٤}\right) \leq \frac{\sigma^{۲}}{\left(\frac{۰/۰۱}{٤}\right)^{۲}} = ۰/۰۵$$

$$\Rightarrow \sigma^{۲} = \frac{۰/۰۱^{۲}}{۱۶} \times ۰/۰۵ = ۳/۱۲۵ \times ۱۰^{-۷}$$

$$Z = \frac{N_{in}}{N_{out}} \qquad N_{in} \rightarrow Bin\left(N_{tot}, \frac{\pi}{٤}\right)$$

$$\sigma_Z^{۲} = \frac{۱}{N_{out}^{۲}} \sigma_{N_{in}}^{۲} = \frac{۱}{N_{out}^{۲}} \times N_{out}\left(۱-\frac{\pi}{٤}\right)\frac{\pi}{٤}$$

$$\rightsquigarrow N_{out} \simeq ۵٤۰,۰۰۰$$

سوال ۲)

1) $a^T x = a_1 x_1 + \dots + a_n x_n$

$$\rightsquigarrow \nabla_x a^T x = [a_1, \dots, a_n] = a^T$$

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

$$= x^T (A + A^T)$$

۲) $$AA^{-1} = I \rightarrow A \frac{\partial A^{-1}}{\partial \beta} + \frac{\partial A}{\partial \beta} A^{-1} = 0$$

$$\rightsquigarrow A \frac{\partial A^{-1}}{\partial \beta} = -\frac{\partial A}{\partial \beta} A^{-1} \rightsquigarrow \frac{\partial A^{-1}}{\partial \beta} = -A^{-1} \frac{\partial A}{\partial \beta} A^{-1}$$

$$\nabla_A \log |A| = \frac{1}{|A|} \nabla_A |A| = A^{-T}$$

سوال ۳) $A = U \Lambda U^T$ $\qquad$ $trace(A) = \sum_n A_{nn}$ $\qquad$ $a_i^T a_j = 1$

الف)

$$\sum A_{nn} = \sum_n \sum_i \sum_j U_{nj} U^T_{in} \Lambda_{ij} = \sum_j \Lambda_{jj} \sum_n U_{in} U^T_{ni} = \sum_j \Lambda$$

$$\leadsto trace(A) = \sum_{i=1}^{N} \lambda_i$$

ب) characteristic polynomial : $\det(A - \lambda I) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$

$$= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$$

$$\leadsto \det A = \prod_{i=1}^{n} \lambda_i$$

سوال ۴) حکم : $A^+ = (A^T A)^{-1} A^T \leadsto (V \Sigma U^T U \Sigma V^T)^{-1} V \Sigma V^T$

الف) $\Leftarrow$ $\qquad = (V \Sigma^T V^T)^{-1} V \Sigma U^T$

در این حالت : $UU^T = I$ $\qquad = V \Sigma^{-T} V^T V \Sigma U^T = V \Sigma^{-1} U^T = A^+$

که حکم اثبات شد

~~$A A^T \ldots$~~ ~~برای ...~~

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

ب) در این حالت ، $\underline{VV^T = I}$

حکم : $A^+ = A^T(AA^T)^{-1}$

$$\Rightarrow V\Sigma U^T(U\Sigma V^T V\Sigma U^T)^{-1} = V\Sigma U^T U\Sigma^{-2}U^T = V\Sigma^{-1}U^T = A^+$$

سوال ۵)

الف)

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix}\begin{pmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{pmatrix}\begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}$$

$$\Rightarrow \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \underbrace{\det\begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix}}_{1}\underbrace{\det\begin{pmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{pmatrix}}_{\det(A)\det(D-CA^{-1}B)}\underbrace{\det\begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}}_{1}$$

$$\Rightarrow \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D-CA^{-1}B)$$

ب)

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{pmatrix}\begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}$$

$$\Rightarrow \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \underbrace{\det\begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}}_{1}\underbrace{\det\begin{pmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{pmatrix}}_{\det(A-BD^{-1}C)\det(D)}\underbrace{\det\begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}}_{1}$$

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

ج) حکم: $(A+BDC)^{-1}=A^{-1}-A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}$

$$\rightsquigarrow (A+BDC)\left(A^{-1}-A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}\right)$$

$$= I+BDCA^{-1}-B(D+CA^{-1}B)^{-1}CA^{-1}-BDCA^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}$$

$$= I+BDCA^{-1}-(B+BDCA^{-1}B)(D+CA^{-1}B)^{-1}CA^{-1}$$

$$= I+BDCA^{-1}-BD\underbrace{(D^{-1}+CA^{-1}B)(D+CA^{-1}B)^{-1}}CA^{-1}$$

? این دو جمله باید یکدیگر را خنثی کنند

در صورت سوال حکم باید به صورت زیر باشد: $(A+BDC)^{-1}=A^{-1}-A^{-1}B(D^{-1}+CA^{-1}B)^{-1}CA^{-1}$

$$\rightsquigarrow = I+BDCA^{-1}-BDCA^{-1}=I$$

حکم درست است

>)

$$\underbrace{\begin{pmatrix} A & u \\ v^T & 1 \end{pmatrix}}_{\det(A)(1-v^TA^{-1}B)} = \underbrace{\begin{pmatrix} I & u \\ 0 & 1 \end{pmatrix}}_{\det 1} \underbrace{\begin{pmatrix} A-uv^T & 0 \\ 0 & 1 \end{pmatrix}}_{\det(A-uv^T)} \underbrace{\begin{pmatrix} I & 0 \\ v^T & 1 \end{pmatrix}}_{\det 1}$$

$$\rightsquigarrow \underline{\det(A-uv^T) = \det(A)(1-v^TA^{-1}B)}$$

$E\ XZ + E$

سوال ۴) $A = \begin{bmatrix} B & x \\ y^* & a \end{bmatrix}$

الف) $\leadsto \tilde{A} = tI - A = \begin{bmatrix} tI-B & x \\ y^* & t-a \end{bmatrix} = \begin{bmatrix} \tilde{B} & x \\ y^* & \tilde{a} \end{bmatrix}$

$\rightarrow$ حکم: $\det(\tilde{A}) = \tilde{a} \det(\tilde{B}) - y^* \operatorname{adj}(\tilde{B}) x$

$$= \det(\tilde{B}) \left(\tilde{a} - y^* \tilde{B}^{-1} x \right)$$

$$= \det(\tilde{B}) \det\left(\tilde{a} - y^* \tilde{B}^{-1} x \right)$$

با توجه به سوال ۵ حکم صادق است.

ب)

$$A = \begin{pmatrix} B & y^* \\ y^* & a \end{pmatrix} \quad ; \quad B = U_B \Lambda_B U_B^* \quad ; \quad \lambda_1^B < \lambda_2^B < \dots < \lambda_n^B$$

2.) $$U_A = \begin{pmatrix} U_B & 0 \\ 0 & 1 \end{pmatrix} \rightsquigarrow \tilde{A} = U_A^* A U_A = \begin{pmatrix} \Lambda_B & y' \\ y'^* & 1 \end{pmatrix}$$

$$\rightsquigarrow s' = (s_1', s_2', \dots, s_{i}', 0, \dots, 0)$$

$$\rightsquigarrow s'^* \tilde{A} s' = \sum \lambda_j^B |s_j'|^2 \leq \lambda_i^B \sum |s_j'|^2 = \lambda_i^B$$

$$\lambda_i^A = \min_{\dim V = i} \max_{\substack{s \in V \\ |s| = 1}} s'^* \tilde{A} s' \leq \max_{s \in V} s^* \tilde{A} s \leq \lambda_i^B$$

$$\cancel{s^2 = (s_1^2, s_2^2, \dots}$$

$$s^2 = (0, \dots, 0, s_{i-1}^2, s_i^2, \dots, s_n^2, 0)$$

$$\rightsquigarrow s^{2*} \tilde{A} s^2 = \sum \lambda_j^B |s_j|^2 \geq \lambda_{i-1}^B \sum |s_j|^2 = \lambda_{i-1}^B$$

$$\lambda_i^A = \max_{\dim V = n-i} \min_{\substack{s \in V \\ |s| = 1}} s^* \tilde{A} s \geq \min_{\substack{s \in V \\ |s| = 1}} s^* \tilde{A} s \geq \lambda_{i-1}^B$$

$$\rightsquigarrow \underline{\lambda_{i-1}^B \leq \lambda_i^A \leq \lambda_i^B}$$

CS CamScanner

سوال ۷)

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta}$$

$$L(\theta) = \prod \frac{1}{\theta^2} X_i e^{\frac{-X_i}{\theta}} = \left(\frac{1}{\theta^2}\right)^n \prod_i X_i \, e^{-\sum_i X_i/\theta}$$

$$\rightsquigarrow \log L(\theta) = -2n\theta + \sum \log X_i - \sum \frac{x_i}{\theta}$$

$$\rightsquigarrow \frac{d}{d\theta} \log L(\theta) = -2n + \frac{\sum x_i}{\theta^2} \rightsquigarrow \tilde{\theta} = \sqrt{\frac{\sum x_i}{2n}}$$

سوال ۸)

MLE:

$$L(x) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\rightsquigarrow \log L(x) = C - \sum \frac{(x_i-\mu)^2}{2\sigma^2} \rightsquigarrow \frac{d}{d\mu} L(x) = \sum \frac{(x_i-\mu)}{\cancel{2}\sigma^2} = 0$$

$$\rightsquigarrow \tilde{\mu} = \frac{\sum x_i}{n}$$

MAP:

$$L_{\mu} = \left(\prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \times \frac{1}{\sqrt{2\pi}\beta} e^{\frac{-(\mu-\gamma)^2}{2\beta^2}}$$

$$\rightsquigarrow \log L_{\mu} = C - \frac{1}{2\sigma^2} \sum_i (x_i-\mu)^2 - \frac{(\mu-\gamma)^2}{2\beta^2}$$

$$\rightsquigarrow \frac{\partial}{\partial \mu} \log L_{\mu} = \frac{+1}{\sigma^2} \sum_i (x_i-\mu) - \frac{\mu-\gamma}{\beta^2}$$

$$\rightsquigarrow \tilde{\mu} = \frac{\beta^2 \sum_i x_i + \sigma^2 \gamma}{\beta^2 + n\sigma^2}$$

- سوال ۹

الف)
$$z = x_a - \Sigma_{ab}\Sigma_{bb}^{-1} x_b$$

$$\mathrm{Cov}(z, x_b) = \underbrace{\mathrm{Cov}(\cancel{X_a}, X_b)}_{\Sigma_{ab}} - \Sigma_{ab}\cancel{\Sigma_{bb}^{-1}}\,\mathrm{Cov}(\cancel{X_b, X_b}) = 0$$

$$E[z] = \mu_1 - \mu_2 \Sigma_{ab}\Sigma_{bb}^{-1}$$

$$E[X_a | X_b] = E(z) + \Sigma_{ab}\Sigma_{bb}^{-1} X_b = \mu_a - \mu_b \Sigma_{ab}\Sigma_{bb}^{-1} + X_b \Sigma_{ab}\Sigma_{bb}^{-1}$$

$$\Rightarrow \mu_{a|b} = \mu_a - \Sigma_{ab}\Sigma_{bb}^{-1}(\mu_b - X_b)$$

ب)
$$\Sigma_{a|b} = \mathrm{Var}(z - \Sigma_{ab}\Sigma_{bb}^{-1} X_b \,|\, X_b)$$
$$= \mathrm{Var}(z)$$
$$= \mathrm{Var}(X_a - \Sigma_{ab}\Sigma_{bb}^{-1} X_b)$$
$$= \mathrm{Var}(X_a) + (\Sigma_{ab}\Sigma_{bb}^{-1})\,\mathrm{Var}(X_b)\,(\Sigma_{bb}^{-1}\Sigma_{ba}) - 2\Sigma_{ab}\Sigma_{bb}^{-1}\,\mathrm{Cov}(X_a, X_b)$$
$$= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

* X_b های مشخص صرفا واریانس و میانگین را تغییر می‌دهند و فرم کلی به صورت گاوسی باقی می‌ماند.

$$E[X_a] = \langle \mu_{a|b} \rangle_b = \langle \mu_a \rangle_b + \Sigma_{aa} \Sigma_{bb}^{-1} \cancel{\langle X_b - \mu_b \rangle}^{\,0}$$

$$\rightsquigarrow E[X_a] = \mu_a$$

$$X_a = AX \quad ; \quad A = \begin{pmatrix} \overbrace{1 \quad 1 \quad \cdots \quad 1}^{n_a} & & \\ & 0 & \\ & & \ddots \quad 0 \end{pmatrix}$$

$$\rightsquigarrow \mathrm{Cov}(X_a) = A \Sigma A := \Sigma_a$$

راه حل ـ
$$L = (Ax-b)^T (Ax-b)$$

الف)
$$= x^T A^T A x - x^T A^T b - b^T A x + b^T b$$

$$\frac{\partial L}{\partial x} = ۲ A^T A x - ۲ A^T b \Rightarrow x^* = (A^T A)^{-1} A^T b$$
$$= A^{+} b$$

ب)
$$x^{t+1} = x^t - \nu A^T (A x^t - b)$$

converge $\rightsquigarrow x^{t+1} = x^t \longrightarrow A^T A x^* - A^T b = 0 \rightsquigarrow x^* = \underbrace{(A^T A)^{-1} A^T}_{A^{+}} b$

ج)
$$A x^{t+1} - b = A x^t - b - \nu A A^T (A x^t - b)$$

$$\rightsquigarrow L^{t+1} = L^t - \nu A A^T L^t = (I - \nu A A^T) L^t$$

CamScanner

برای برقراری حکم : $\det(I - \upsilon AA^T) < ۱$

$$\rightsquigarrow \; I - \upsilon AA^T = \cancel{A}U(I - \upsilon\Lambda)U\cancel{B}^T$$

$$= U \begin{pmatrix} ۱ - \upsilon\lambda_۱ & & \\ & ۱ - \upsilon\lambda_۲ & \\ & & \ddots \end{pmatrix}$$

$$\rightsquigarrow \; |۱ - \upsilon\lambda_i| < ۱ \;\rightsquigarrow\; \upsilon < \frac{۲}{\lambda_i} \;\rightsquigarrow\; \underline{\upsilon < \frac{۲}{\lambda_{max}}}$$