تمرین اول درس یادگیری عمیق

داریوش تاکی 98109678

1| الف)

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\mathrm{Trace}(A^TA)} = \sqrt{\sum_{i=1}^{r} (\lambda_i(A^TA))^2} \qquad r = \mathrm{rank}(A)$$

$$\|A\|_2 = \sqrt{\lambda_{max}(A^TA)} \underset{\lambda_i \text{ها مثبت‌اند}}{\leq} \sqrt{\sum_{i=1}^{r} (\lambda_i(A^TA))^2} = \|A\|_F \underset{\lambda_i < \lambda_{max}}{\leq} \sqrt{r\,\lambda_{max}(A^TA)}$$

$$\Rightarrow \boxed{\|A\|_2 \leq \|A\|_F \leq \sqrt{\mathrm{rank}(A)}\,\|A\|_2}$$

ب) I

$$E(x) = \underbrace{\int_{x=0}^{\alpha} x\,P(X=x)\,dx}_{\text{عبارت مثبت}} + \int_{x=\alpha}^{\infty} x\,P(X=x)\,dx \geq \int_{x=\alpha}^{\infty} x\,P(X=x)\,dx$$

$$\underset{\alpha \leq x}{\geq} \alpha \int_{x=\alpha}^{\infty} P(X=x)\,dx = \alpha P(X \geq \alpha) \Rightarrow \boxed{\frac{E(x)}{\alpha} \geq P(X \geq \alpha)}$$

II

$$P(|z-\mu| \geq \varepsilon) = P(\underbrace{(z-\mu)^2}_{X} \geq \underbrace{\varepsilon^2}_{\alpha}) \leq \frac{E((z-\mu)^2)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

$$\Rightarrow \boxed{P(|z-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}}$$

III

نسبت تقریب زده شده را متغیر تصادفی Z تعریف می‌کنیم

$$Z = \frac{S_{(\text{دایره})}}{S_{(\text{مربع})}} \to \frac{\pi}{4}$$

انتظار داریم این مقدار به $\mu = \frac{\pi}{4}$ میل کند

که x_i ها iid اند (مستقل اند)

$$\Rightarrow Z = \frac{\sum_{i=1}^{N} x_i}{N}, \qquad x_i = \begin{cases} 0 & \text{با احتمال } 1-\frac{\pi}{4} \\ 1 & \text{با احتمال } \frac{\pi}{4} \end{cases}$$

$$Z = \frac{\sum_{i=1}^{N} x_i}{N} \Rightarrow \mu = E(Z) = \frac{E\left(\sum_{i=1}^{N} x_i\right)}{N} = \frac{\sum_{i=1}^{N} E(x_i)}{N} = \frac{N \times \frac{\pi}{4}}{N} = \frac{\pi}{4}$$

$$\sigma^2 = E(z^2) - \mu^2 = \frac{E\left[\left(\sum_{i=1}^{N} x_i\right)^2\right]}{N^2} - \left(\frac{\pi}{4}\right)^2 = \frac{E\left(\sum_{i=1}^{N} x_i^2 + 2\sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i x_j\right)}{N^2}$$

$$- \frac{\pi^2}{16} = \frac{1}{N^2}\left[\sum_{i=1}^{N} E(x_i^2) + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} E(x_i x_j)\right] - \frac{\pi^2}{16}$$

$$E(x_i x_j) = E(x_i) E(x_j) = \frac{\pi^2}{16}$$

x_i, x_j مستقلند (بنا iid)

$$= \frac{1}{N^2}\left[N \frac{\pi}{4} + 2\times\left(\frac{N(N-1)}{2} \frac{\pi^2}{16}\right)\right] - \frac{\pi^2}{16}$$

$$= \frac{1}{N}\left[\frac{\pi}{4} - \frac{\pi^2}{16}\right] = \sigma^2$$

$$P\left[|Z - \mu| \geq 0.01\mu\right] \leq \frac{\sigma^2}{(0.01\mu)^2} \leq 0.05$$

$$\Rightarrow \frac{1}{N}\left(1 - \frac{\pi}{4}\right) \times \frac{\pi}{4} \leq 5 \times 10^{-6} \frac{\pi^2}{16}$$

$$\Rightarrow \frac{1 - \frac{\pi}{4}}{\frac{\pi}{4}} \times \frac{1}{5 \times 10^{-6}} \leq N \Rightarrow 54647.9 \leq N \Rightarrow \underline{N_{min} = 54648}$$

$$\underline{a}^T = [a_1 \;\cdots\; a_n] \Rightarrow \underline{a}^T\underline{x} = [a_1\underline{x} \;\cdots\; a_n\underline{x}]$$

$$\Rightarrow \frac{\partial(\underline{a}^T\underline{x})}{\partial \underline{x}} = \left[\frac{\partial(a_1\underline{x})}{\partial\underline{x}} \;\cdots\; \frac{\partial(a_n\underline{x})}{\partial\underline{x}} \right] = [a_1 \;\cdots\; a_n] = \underline{a}^T$$

$$\Rightarrow \boxed{\frac{\partial(a^T x)}{\partial \underline{x}} = a^T}$$

$$\frac{\partial(a^T x)}{\partial x} = \frac{\partial(x^T a)}{\partial x} = a^T \Rightarrow \left(\frac{\partial(x^T a)}{\partial x}\right)^T = a \quad ⊛$$

اساساً طبق

برابر I است

زنجیری

$$\frac{\partial(x^T A x)}{\partial x} = \frac{\partial(x^T A)}{\partial x}(Ix) + x^T A \frac{\partial I x}{\partial x} = x^T A + x^T \underbrace{\left(\frac{\partial(x^T A)}{\partial x}\right)^T}$$

طبق اثبات ⊛ ترانهاده‌ی

برابر A^T است

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

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

II

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

$$A = V D U^T$$

$$A^{-1} = U D^{-1} V^{-T}$$

$$\frac{\partial A^{-1}}{\partial \beta} = \frac{\partial U}{\partial \beta} D^{-1} V^{-T} + U \frac{\partial(D^{-1})}{\partial \beta} V^{-T} + U D^{-1} \frac{\partial (V^{-T})}{\partial \beta}$$

$$= \frac{\partial U}{\partial \beta} U^T A^{-1} + U \begin{pmatrix} -\frac{d_1'}{d_1^2} & & 0 \\ & \ddots & \\ 0 & & -\frac{d_n'}{d_n^2} \end{pmatrix} V^{-T} + A^{-1} V \left(\frac{\partial V}{\partial \beta}\right)^T$$

$$\frac{\partial A}{\partial \beta} = \frac{\partial V}{\partial \beta} D U^T + V \begin{pmatrix} d_1' & & 0 \\ & \ddots & \\ 0 & & d_n' \end{pmatrix} U^T + V D \left(\frac{\partial U}{\partial \beta}\right)^T$$

$$U D^{-1} V^T \frac{\partial V}{\partial \beta} D U^T U D^{-1} V^T + U D^{-1} V^T V \begin{pmatrix} d_1' & & 0 \\ & \ddots & \\ 0 & & d_n' \end{pmatrix} U^T U D^{-1} V^T + U D^{-1} V^T V D \left(\frac{\partial U}{\partial \beta}\right)^T U D^{-1} V^T$$

$$= A^{-1} \left(\frac{\partial A}{\partial \beta}\right) A^{-1} = U D^{-1} \underbrace{\left(V^T \frac{\partial V}{\partial \beta} V^T\right)}_{-\frac{\partial V^T}{\partial \beta}} + U \begin{pmatrix} d_1'/d_1^2 & & 0 \\ & \ddots & \\ 0 & & d_n'/d_n^2 \end{pmatrix} V^T + \underbrace{U \left(\frac{\partial U}{\partial \beta}\right)^T U}_{-\frac{\partial U}{\partial \beta}} D^{-1} V^T$$

$$V V^T = I \Rightarrow \frac{\partial V}{\partial \beta} V^T + V \frac{\partial V^T}{\partial \beta} = 0$$

$$\Rightarrow -A^{-1} \left(\frac{\partial A}{\partial \beta}\right) A^{-1} = +\left(U D^{-1} \frac{\partial V^T}{\partial \beta} + U \frac{\partial (D^{-1})}{\partial \beta} V^T + \frac{\partial U}{\partial \beta} D^{-1} V^T\right)$$

$$\Rightarrow \boxed{-A^{-1} \left(\frac{\partial A}{\partial \beta}\right) A^{-1} = \frac{\partial (A^{-1})}{\partial \beta}}$$

III

$$\det(A) = \det(V D U^T) = \det(D) = \prod_{i=1}^{n} \lambda_i$$

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

$$A = [a_{ij}] \Rightarrow \frac{\partial |A|}{\partial a_{ij}} = \operatorname{adj}(A)_{ji}$$

طبق تعریف :

$$|A| I = \operatorname{adj}(A) A$$

$$\operatorname{adj}(A)_{ij} = (-1)^{i+j} M_{ij} \Rightarrow \sum_{j=1}^{n} a_{ij} \underbrace{\operatorname{adj}(A)_{ji}}_{\text{شامل زوجها نیست}} = |A|$$

$$\Rightarrow \frac{\partial |A|}{\partial a_{ij}} = \operatorname{adj}(A)_{ji} \Rightarrow \nabla_A |A| = \operatorname{adj}(A)^T = \boxed{|A| \, {A^{-1}}^T = \nabla_A |A|}$$

$$\frac{\partial}{\partial a_{ij}} \log|A| = \frac{1}{|A|} \frac{\partial}{\partial a_{ij}} |A| \Rightarrow \nabla_A \log|A| = \frac{1}{|A|} \nabla_A |A| \Rightarrow$$

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

Scanned with Cam

3

$$A = \sum_{i=1}^{n} \lambda_i u_i v_i^T \Rightarrow \quad \text{Trace}(A) = \sum_{i=1}^{n} \lambda_i \, \text{Trace}(u_i v_i^T)$$

$$= \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} u_{ij} v_{ij}$$

I $A = PJP^{-1}$ فرم جردان ماتریس A،   II $\text{Trace}(ABC) = \text{Trace}(CAB)$

$$\overset{\text{I},\text{II}}{\Rightarrow} \quad \boxed{\text{Trace}(A) = \sum_{i=1}^{n} \lambda_i} = \text{Trace}(J)$$

$$A = UDV^T \Rightarrow \quad \det(A) = \underbrace{\det(U)}_{1} \; \underbrace{\det(D)}_{\prod_i \lambda_i} \; \underbrace{\det(V^T)}_{1}$$

$$\Rightarrow \quad \boxed{\det(A) = \prod_{i=1}^{n} \lambda_i}$$

همین اثبات را با فرم جردان هم می‌توان نوشت،

$$A = PJP^{-1} \Rightarrow \quad \det(P) \times \underbrace{\det(J)}_{\prod_{i=1}^{n} \lambda_i} \times \det(P) = \boxed{\det(A) = \prod_{i=1}^{n} \lambda_i}$$

۴) A: $n\times m$ است، U و $n\times n$، V و $m\times m$ است
و Σ $n\times m$ است

$$AA^T = U\Sigma\Sigma^T U^T$$

$$A = \sum_{i=1}^{r} \lambda_i u_i v_i^T \Rightarrow A^{\dagger} = \sum_{i=1}^{r} \frac{1}{\lambda_i} u_i v_i^T$$

* رتبه کامل سطری باشد $r=n$، $\Sigma\Sigma^T$: $n\times n$ خواهد بود پس $\Sigma\Sigma^T$ معکوس پذیر است، و AA^T معکوس
پذیر است $\Leftarrow$

$$A^{\dagger} = A^T (AA^T)^{-1}, \quad AA^{\dagger} = I$$

* رتبه کامل ستونی باشد $r=m$، $\Sigma^T\Sigma$: $m\times m$ خواهد بود
پس $\Sigma^T\Sigma$ معکوس پذیر است و A^TA معکوس پذیر است

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

$$\rightarrow \quad A^{\dagger} = (A^TA)^{-1}A^T, \quad A^{\dagger}A = I$$

Scanned with Cam

15

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

$$= \underbrace{\begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}}_{M'} \underbrace{\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}}_{V}$$

$$\det(M) = \det(UM'V) = \det(U)\det(M')\det(V)$$

ماتریس بالامثلثی
پایین‌مثلثی $\Rightarrow$ دترمینان ضرب قطرها $\Rightarrow$ $\det(U) = \det(V) = 1$

$$\Rightarrow \det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) \quad \det(M') = \det(A)\det(D - CA^{-1}B)$$

⊛

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

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

شبیه قبل $\Rightarrow$

$$\det(M) = \det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A - BD^{-1}C)\det(D)$$

$$(A + BD^{-1}C)^{-1} = (I + BD^{-1}CA^{-1})^{-1} A^{-1}$$

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

$$= \begin{bmatrix} I & 0 \\ +D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A + BD^{-1}C)^{-1} & 0 \\ 0 & -D^{-1} \end{bmatrix} \begin{bmatrix} I & +BD^{-1} \\ 0 & I \end{bmatrix}$$

به همین ترتیب

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

$$\Rightarrow \begin{bmatrix} (A + BD^{-1}C)^{-1} & 0 \\ +D^{-1}C(A + BD^{-1}C)^{-1} & -D^{-1} \end{bmatrix} \begin{bmatrix} I & +BD^{-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} (A + BD^{-1}C)^{-1} & +(A + BD^{-1}C)^{-1}BD^{-1} \\ +D^{-1}C(A + BD^{-1}C)^{-1} & +D^{-1}C(A + BD^{-1}C)^{-1}BD^{-1} - D^{-1} \end{bmatrix}$$

به همین ترتیب

$$= \begin{bmatrix} A^{-1}B(-D - CA^{-1}B)^{-1}CA^{-1} + A^{-1} & -A^{-1}B(-D - CA^{-1}B)^{-1} \\ -(-D - CA^{-1}B)^{-1}CA^{-1} & (-D - CA^{-1}B)^{-1} \end{bmatrix}$$

تساوی درایه اول

$$\Rightarrow \boxed{(A + BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(-D - CA^{-1}B)^{-1}CA^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}}$$

Scanned with Cam

$A = \overline{A}^T$ (Roger + Matrix Analysis اثبات با کمک کتاب) ⊛

فرض کنید x_i ها، y_i ها، z_i ها یکه متعامد اند

$$Ax_i = \lambda_i(A)\, x_i \qquad By_i = \lambda_i(B)\, y_i$$

$$(A+B)\, z_i = \lambda_i(A+B)\, z_i$$

$$i \in \{1, \dots, n\} \quad , \quad j \in \{0, \dots, n-i\}$$

$$S_1 = \operatorname{span}\{x_1, \dots, x_{i+j}\} \quad , \quad S_2 = \operatorname{span}\{y_1, \dots, y_{n-j}\}$$

$$S_3 = \operatorname{span}\{z_i, \dots, z_n\} \qquad \underbrace{\underbrace{\dim S_1 + \dim S_2}_{\text{(I)}\ n+i} + \underbrace{\dim S_3}_{n-i+1}}_{\text{(II)}} = 2n+1$$

(I) فضای S_1 و S_2 حداقل i بردار مشترک دارند

⇓

(II) فضای $S_1 \cap S_2$ و S_3 حداقل یک بردار مشترک دارند

$$\Rightarrow \exists x : \quad x \in S_1 \cap S_2 \cap S_3 \ , \|x\| = 1$$

ترتیب صعودی λ ها

$$\lambda_i(A+B) \le x^*(A+B)x = x^*Ax + x^*Bx \le \lambda_{i+j}(A) + \lambda_{n-j}(B)$$

$$\Rightarrow \boxed{\lambda_i(A+B) \le \lambda_{i+j}(A) + \lambda_{n-j}(B)}$$

$$A = \begin{bmatrix} B & y \\ y^* & a \end{bmatrix} \Rightarrow Av_i = \begin{bmatrix} B & y \\ y^* & a \end{bmatrix} \begin{bmatrix} x_i \\ b_i \end{bmatrix}$$

$v_i = \begin{bmatrix} x_i \\ b_i \end{bmatrix}$ ها بردارهای ویژه‌ی متعامد A هستند ⊛

$$= \lambda_i(A)\, v_i = \begin{bmatrix} Bx_i + b_i y \\ y^* x_i + a b_i \end{bmatrix}$$

$$Bx_i + b_i y = \lambda_i(A)\, x_i$$
$$y^* x_i + a b_i = \lambda_i(A)\, b_i$$

$$A = \underbrace{\begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}}_{M} + \underbrace{\begin{bmatrix} 0 & y \\ y^* & a \end{bmatrix}}_{N}$$

$\circledast$

$\det(A - uv^T)$

$$\det\begin{bmatrix} A & u \\ v^T & I \end{bmatrix} = \det(A)\det(I - v^T A^{-1} u) = \det(I)\det(A - uv^T)$$

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

$\circledast$ 16

$$P_{A(t)} = |tI - A| = \det\begin{bmatrix} -B + tI & -x \\ -y^* & -a + t \end{bmatrix}$$

طبق سوال 5 ، اسکالر است

$$\overset{\circledast}{=} \det(-B + tI)\det\left(-a + t - (-y^*)(-B + tI)^{-1}(-x)\right)$$

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

$$(tI - B)^{-1} = \frac{\operatorname{adj}(tI - B)}{\det(tI - B)}$$

$$\overset{\circledast}{\Rightarrow} P_{A(t)} = P_{B(t)} \times \left[(-a + t) - y^*\left(\operatorname{adj}(tI - B)\, P_{B(t)}^{-1}\right) x\right]$$

$$\Rightarrow \boxed{P_{A(t)} = (t - a) P_{B(t)} - y^* \operatorname{adj}(tI - B)\, x}$$

$$\begin{bmatrix} 0 & y \\ y^* & a \end{bmatrix} = N \Rightarrow \quad P_N(t) = (t-a)\times t^n - y^*y \; t^{n-1}$$

$$= t^{n-1}\left(t^2 - at - y^*y\right)$$

$$= t^{n-1}\Big(t - \underbrace{\big(\tfrac{a}{2} + \sqrt{(\tfrac{a}{2})^2 + y^*y}\,\big)}_{\text{مثبت}}\Big)\Big(t - \underbrace{\big(+\tfrac{a}{2} - \sqrt{(\tfrac{a}{2})^2 + y^*y}\,\big)}_{\text{منفی}}\Big)$$

N ، دو مقدار ویژه‌ی ناصفر دارد و $n-1$ مقدار ویژه‌ی صفر دارد $\lambda_1(N) \leq 0 \;,\; \lambda_{n+1}(N) \geq 0$

قضیه ویل

$$\Rightarrow \quad \lambda_i(A) \leq \lambda_{i+j}(M) + \lambda_{n+1-j}(N)$$

$j=1$

$$\Rightarrow \quad \lambda_i(A) \leq \lambda_{i+1}(M) + \cancel{\lambda_n(N)}^{\,0}$$

$$\left\{\begin{array}{ll} \lambda_i(M) = \lambda_{i-1}(B) & i \geq 2 \\ \lambda_i(M) = 0 & i = 1 \end{array}\right\} \overset{i\geq 1}{\Rightarrow} \quad \lambda_{i+1}(M) = \lambda_i(B)$$

$$\Rightarrow \quad \lambda_i(A) \leq \lambda_i(B)$$

$$\lambda_{n+1}(A) \leq \lambda_{n+1}(M) + \lambda_{n+1}(N) \geq \lambda_n(B)$$

$$M = A + (-N)$$

قضیه ویل

$$\Rightarrow \quad \lambda_i(M) \leq \lambda_{i+j}(A) + \lambda_{n-j}(N)$$

$$\left.\begin{array}{l} i=n \\ j=1 \end{array}\right\} \Rightarrow \underbrace{\lambda_n(M)}_{\lambda_n(B)} \leq \lambda_{n+1}(A) + \cancel{\lambda_n(-N)}_{\,0} \Rightarrow \quad \lambda_n(B) \leq \lambda_{n+1}(A)$$

$$\Rightarrow \quad \lambda_1(A) \leq \lambda_1(B) \leq \cdots \leq \lambda_n(B) \leq \lambda_{n+1}(A)$$

17

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left(\frac{1}{\theta^{2i}} x_i e^{-\frac{x_i}{\theta}} \right)$$

$$= \frac{1}{\theta^{2N}} \left(\prod_{i=1}^{n} x_i\right) \times e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$$

$$\Rightarrow \left.\frac{\partial \mathcal{L}(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0 \Rightarrow \left(\prod_{i=1}^{n} x_i\right)\left(\frac{-2N}{\hat{\theta}^{2N+1}} e^{-\frac{\sum_{i=1}^{n} x_i}{\hat{\theta}}} + \frac{1}{\hat{\theta}^{2N}} \left(\frac{+\sum_{i=1}^{n} x_i}{\hat{\theta}^2}\right) e^{-\frac{\sum_{i=1}^{n} x_i}{\hat{\theta}}}\right) = 0$$

$$\Rightarrow 2N\hat{\theta} = \sum_{i=1}^{n} x_i \Rightarrow \boxed{\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{2N}}$$

18

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

$$\Rightarrow \mathcal{L}(\mu) = \prod_{i=1}^{n} f_X(x_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow \hat{\mu}_{MLE} : \left.\frac{\partial \mathcal{L}(\mu)}{\partial \mu}\right|_{\mu=\hat{\mu}_{MLE}} = 0 \Rightarrow \frac{e^{-\frac{\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE})^2}{2\sigma^2}}}{(2\pi\sigma^2)^{n/2}} \left[\frac{2}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\hat{\mu}_{MLE})\right)\right] = 0$$

$$\Rightarrow \boxed{\hat{\mu}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{N}}$$

$$\hat{\mu}_{MAP} = \arg\max f_X(x|\mu) f_\mu(\mu) = \arg\max \left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(\mu-\gamma)^2}{2\beta^2}}\right]$$

$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}_{MAP}) + \frac{1}{\beta^2}(\gamma-\hat{\mu}_{MAP}) = 0 \Rightarrow \boxed{\frac{\beta^2\sum_{i=1}^{n} x_i + \sigma^2\gamma}{N\beta^2+\sigma^2} = \hat{\mu}_{MAP}}$$

$$P(x_a, x_b) = \frac{1}{(2\pi)^{?}\det(\Sigma)} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

$$\Sigma^{-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \Rightarrow \begin{cases} \Sigma^{-1}\Sigma = \begin{bmatrix} \alpha_{11}\Sigma_{aa} + \alpha_{12}\Sigma_{ba} & \alpha_{11}\Sigma_{ab} + \alpha_{12}\Sigma_{bb} \\ \alpha_{21}\Sigma_{aa} + \alpha_{22}\Sigma_{ba} & \alpha_{21}\Sigma_{ab} + \alpha_{22}\Sigma_{bb} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ \\ \Sigma\Sigma^{-1} = \begin{bmatrix} \Sigma_{aa}\alpha_{11} + \Sigma_{ab}\alpha_{21} & \Sigma_{aa}\alpha_{12} + \Sigma_{ab}\alpha_{22} \\ \Sigma_{ba}\alpha_{11} + \Sigma_{bb}\alpha_{21} & \Sigma_{ba}\alpha_{12} + \Sigma_{bb}\alpha_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \end{cases}$$

$$\Rightarrow \quad \alpha_{12} = -\alpha_{11}\Sigma_{ab}\Sigma_{bb}^{-1}, \quad \alpha_{21} = \begin{cases} -\Sigma_{bb}^{-1}\Sigma_{ba}\alpha_{11} \Rightarrow \alpha_{22} = \Sigma_{bb}^{-1} - \Sigma_{bb}^{-1}\Sigma_{ba}(-\alpha_{11}\Sigma_{ab}\Sigma_{bb}^{-1}) \\ -\alpha_{22}\Sigma_{ba}\Sigma_{aa}^{-1} \end{cases}$$

$$\alpha_{11} = \left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right)^{-1}$$

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = \left((x_a-\mu_a)^T\alpha_{11} + (x_b-\mu_b)^T\alpha_{21}\right)(x_a-\mu_a) + \left((x_a-\mu_a)^T\alpha_{21} + (x_b-\mu_b)^T\alpha_{22}\right)(x_b-\mu_b)$$

$$= (x_a-\mu_a)^T\alpha_{11}(x_a-\mu_a) + (x_b-\mu_b)^T(-\Sigma_{bb}^{-1}\Sigma_{ba}\alpha_{11})(x_a-\mu_a) + (x_a-\mu_a)^T(-\alpha_{11}\Sigma_{ab}\Sigma_{bb}^{-1})(x_b-\mu_b)$$

$$+ (x_b-\mu_b)^T\Sigma_{bb}^{-1}(x_b-\mu_b) + (x_b-\mu_b)^T\Sigma_{bb}^{-1}\Sigma_{ba}\alpha_{11}\Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b)$$

$$= \left(x_a - \mu_a - \Sigma_{ba}^T\Sigma_{bb}^{T^{-1}}(x_b-\mu_b)\right)^T \alpha_{11}\left(x_a - \mu_a - \Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b)\right)$$

$$+ (x_b-\mu_b)^T\Sigma_{bb}^{-1}(x_b-\mu_b)$$

$$\xrightarrow[\Sigma_{bb}^T = \Sigma_{bb}]{\Sigma_{ba}^T = \Sigma_{ab}} \quad \begin{aligned} &x_a - (\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b-\mu_b)) = u_a \\ &x_b - \mu_b = u_b \end{aligned}$$

$$\Rightarrow (x-\mu)^T\Sigma^{-1}(x-\mu) = u_a^T\alpha_{11}u_a + u_b^T\Sigma_{22}^{-1}u_b$$

$$\int_{x_a} P(x_a,x_b)\,dx_a = P(x_b) = \underset{\text{ضریب نرمالیزیشن}}{C_{a.b}}\, e^{-\frac{1}{2}u_b^T\Sigma_{22}^{-1}u_b} \times \underbrace{\int_{u_a} e^{-\frac{1}{2}u_a^T\alpha_{11}u_a}\,du_a}_{\text{ضریب نرمالیزیشن}} = C_b\, e^{-\frac{1}{2}u_b^T\Sigma_{22}^{-1}u_b}$$

$$\Rightarrow x_b \sim \mathcal{N}(\mu_b, \Sigma_{22})$$

$$P_{(x_a|x_b)} = \frac{P_{(x_a, x_b)}}{P_{(x_b)}} = C_{a|b}\, e^{-\frac{1}{2} u_a^T \alpha_{11} u_a} \Rightarrow [x_a | x_b] \sim \mathcal{N}(\mu_{a|b}, \Sigma_{a|b})$$

که ضریب نمایی پیش

$$\Rightarrow \begin{cases} \mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b) \\ \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba} \end{cases}$$

مشابه روند اثبات توزیع x_b

$$\Rightarrow x_a \sim \mathcal{N}(\mu_a, \Sigma_{11}) \Rightarrow \begin{cases} \mathbb{E}[x_a] = \mu_a \\ \mathrm{Cov}(x_a) = \Sigma_{aa} \end{cases}$$

10

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

سوال 2

$$\Rightarrow \nabla_x \|Ax - b\|^2 = x^T (A^T A + (A^T A)^T) - 2 b^T A$$

$$= 2(x^T A^T - b^T) A$$

در حالت بهینه

$$\Rightarrow (x^T A^T - b^T) A = 0 \Rightarrow b^T A = x^T A^T A$$

$$\Rightarrow A^T A x = A^T b$$

Z_{A^TA} : فضای پوچی راست ماتریس A^TA

$$\Rightarrow x = x^* + v \qquad x^* = \underbrace{(A^T A)^{-1} A^T}_{A^\dagger} b$$

$$v \in Z_{A^TA}$$

$\|x^*\| \le \|x^* + v\| = \|x\|$ $\Leftarrow$ فضای پوچی ماتریس در فضای بردارهای سطری A^TA عمود است

$$\Rightarrow \boxed{x^* = A^\dagger b}$$

کوچکترین پاسخ مینیمم کننده‌ی عبارت است

$$x^{(1)} = +\nu A^T b \ , \quad x^{(2)} = \nu A^T b + (I - \nu A^T A)\nu A^T b$$

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

$$= \nu A^T b + (I - \nu A^T A)\left(\nu A^T b + (I - \nu A^T A)\, x^{(t-1)}\right)$$

$$\overset{*}{=} \sum_{k=0}^{t-1} (I - \nu A^T A)^k \nu A^T b \Rightarrow (I - \nu A^T A)\, x^{(t+1)} - x^{(t+1)} = \left((I - \nu A^T A)^{t+1} - I\right)\nu A^T b$$

$$\Rightarrow \quad -\nu A^T A\, x^{(t+1)} = \left((I - \nu A^T A)^{t+1} - I\right)\nu A^T b$$

$$\Rightarrow \quad x^{(t+1)} = (A^T A)^{-1}\left[I - (I - \nu A^T A)^{t+1}\right] A^T b$$

$$t \to \infty, \ \overset{*}{\Rightarrow} \quad (I - \nu A^T A)\, x^{(\infty)} = x^{(\infty)} - I\,\nu A^T b \Rightarrow x^{(\infty)} = (A^T A)^{-1} A^T b$$

$$\Rightarrow \boxed{t \to \infty \Rightarrow x \to (A^T A)^{-1} A^T b = A^\dagger b}$$

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

$$= (I - \nu A A^T)(A x^{(t)} - b)$$

$$\| A x^{(t+1)} - b \|^2 = (A x^{(t)} - b)^T \underbrace{(I - \nu A A^T)^T (I - \nu A A^T)}\, (A x^{(t)} - b)$$

$$= (A x^{(t)} - b)^T (I - \nu A A^T)^2 (A x^{(t)} - b)$$

$$\| A x^{(t+1)} - b \|^2 - \| A x^{(t)} - b \|^2 = (A x^{(t)} - b)^T \underbrace{\left[(I - \nu A A^T)^2 - I\right]}_{-2AA^T\nu + \nu^2 (AA^T)^2} (A x^{(t)} - b)$$

t بزرگ

$\Rightarrow$

برقرار باشه

$$2I > \nu (A A^T) \Rightarrow \forall i: \ \lambda_i(AA^T) = \lambda_i^2(A) \le \frac{2}{\nu}$$

$$\Rightarrow \frac{2}{\max \lambda_i^2(A)} \ge \nu \Rightarrow \boxed{\nu \le \frac{2}{\lambda_{\max}^2(A)}}$$