99102189 - محمدمهدی [illegible]

۱.
$$P(X \mid Y=i) = P(\underset{\text{ویژگی اول}}{X_1}, \underset{\text{ویژگی دوم}}{X_2} \mid Y=i) = \frac{1}{2\pi\sqrt{\det C}} \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)$$

$$\det(C_1) = 0.7 \times 0.7 = 0.49 \qquad C_1^{-1} = \begin{bmatrix} \frac{10}{7} & 0 \\ 0 & \frac{10}{7} \end{bmatrix}$$

$$\det(C_2) = 0.8\times 0.2 - 0.3\times 0.3 = 0.07 \qquad C_2^{-1} = \begin{bmatrix} \frac{20}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{80}{7} \end{bmatrix}$$

$$\det(C_3) = 0.7\times 0.8 - 0.2\times 0.2 = 0.52 \qquad C_3^{-1} = \begin{bmatrix} \frac{20}{13} & -\frac{5}{13} \\ -\frac{5}{13} & \frac{35}{26} \end{bmatrix}$$

الف)

لیبل اول $= P(Y=1)\,P(X \mid Y=1) = \frac{1}{3} \times \frac{1}{2\pi\sqrt{0.49}} \exp\left(-\frac{1}{2}[50, 0.5]\begin{bmatrix} \frac{10}{7} & 0 \\ 0 & \frac{10}{7} \end{bmatrix}\begin{bmatrix} 50 \\ 0.5 \end{bmatrix}\right)$

احتمال پسین اینکه X عضو لیبل اول باشد $= 1.875 \times 10^{-777} \Rightarrow \boxed{\ln(P_1) = -1788.5}$

لیبل دوم $= P(Y=2)\,P(X \mid Y=2) = \frac{1}{3} \times \frac{1}{2\pi\sqrt{0.07}} \exp\left(-\frac{1}{2}[49, -0.5]\begin{bmatrix} \frac{20}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{80}{7} \end{bmatrix}\begin{bmatrix} 49 \\ -0.5 \end{bmatrix}\right)$

احتمال پسین اینکه X عضو لیبل دوم باشد ← ماشین حساب صفر می‌دهد ← پس لاگ می‌گیریم $\Rightarrow \ln(P_2) = \ln\left(\frac{1}{6\pi\sqrt{0.07}}\right) - 3536.4$

$$\Rightarrow \boxed{\ln(P_2) = -3538.0}$$

لیبل سوم $= P(Y=3)\,P(X \mid Y=3) = \frac{1}{3} \times \frac{1}{2\pi\sqrt{0.52}} \exp\left(-\frac{1}{2}[49, -0.5]\begin{bmatrix} \frac{20}{13} & -\frac{5}{13} \\ -\frac{5}{13} & \frac{35}{26} \end{bmatrix}\begin{bmatrix} 49 \\ -0.5 \end{bmatrix}\right)$

احتمال پسین اینکه X عضو لیبل سوم باشد $= 3.972 \times 10^{-808} \rightarrow \boxed{\ln(P_3) = -1859.1}$

بنابراین $x = [50, 0.5]$ عضو کلاس و لیبل اول است.

ب)

$$\ln(P_1) = \ln\left(\frac{1}{6\pi\sqrt{0.49}}\right) - \frac{1}{2} \times [0.5, 0.5]\begin{bmatrix} \frac{10}{7} & 0 \\ 0 & \frac{10}{7} \end{bmatrix}\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = -2.937$$

$$\ln(P_2) = \ln\left(\frac{1}{6\pi\sqrt{0.07}}\right) - \frac{1}{2}[-0.5 \ \ -0.5]\begin{bmatrix} \frac{20}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{80}{7} \end{bmatrix}\begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = -0.7143$$

$$\ln(P_3) = \ln\left(\frac{1}{6\pi\sqrt{0.52}}\right) - \frac{1}{2}[-0.5 \ \ -0.5]\begin{bmatrix} \frac{20}{13} & -\frac{5}{13} \\ -\frac{5}{13} & \frac{35}{26} \end{bmatrix}\begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = -2.874$$

بنابراین $x = [0.5\ 0.5]$ عضو کلاس و لیبل دوم است.

Scanned with CamScanner

$$E\left\{\frac{1}{2}\sum_{n=1}^{N}\left(w_0+\sum_{i=1}^{D}(w_i x_{ni}+w_i \varepsilon_{ni})-y_n\right)^2\right\}=\frac{1}{2}E\left\{\sum_{n=1}^{N}\left(w_0+\sum_{i=1}^{D}w_i x_{ni}+\sum_{i=1}^{D}w_i\varepsilon_{ni}-y_n\right)^2\right\}$$

$$=\frac{1}{2}E\left\{\sum_{n=1}^{N}\left(w_0+\sum_{i=1}^{D}w_i x_{ni}-y_n+\sum_{i=1}^{D}w_i\varepsilon_{ni}\right)^2\right\}=\frac{1}{2}E\left\{\sum_{n=1}^{N}\left(w_0+\sum_{i=1}^{D}w_i x_{ni}-y_n\right)^2\right\}+\frac{1}{2}E\left\{\left(\sum_{i=1}^{D}w_i\varepsilon_{ni}\right)^2\right\}$$

$$+E\left\{\left(\sum_{i=1}^{D}w_i\varepsilon_{ni}\right)\left(w_0+\sum_{i=1}^{D}w_i x_{ni}-y_n\right)\right\}$$

چون ε_{ni} ها از همه‌ی x_{ni} ها مستقل اند و y_n و w_i ها عدد ثابت اند. پس:

$$\tilde{E}_D(w)=E_D(w)+\frac{1}{2}E\left\{\left(\sum_{i=1}^{D}w_i\varepsilon_{ni}\right)^2\right\}\underbrace{-\frac{1}{2}\left(E\left\{\sum_{i=1}^{D}w_i\varepsilon_{ni}\right\}\right)^2}+\underbrace{E\left\{\sum_{i=1}^{D}w_i\varepsilon_{ni}\right\}}_{0}E\left\{w_0+\sum_{i=1}^{D}w_i x_{ni}-y_n\right\}$$

این جمله برابر با صفر است.
آن را اضافه کردیم تا با جمله‌ی
قبلی واریانس را تشکیل دهد.

$$E\left\{\sum_{i=1}^{D}w_i\varepsilon_{ni}\right\}=\sum_{i=1}^{D}E\{w_i\varepsilon_{ni}\}=\sum_{i=1}^{D}w_i\underbrace{E\{\varepsilon_{ni}\}}_{0}=0$$

$$\Rightarrow \tilde{E}_D(w)=E_D(w)+\frac{1}{2}\mathrm{Var}\left(\left(\sum_{i=1}^{D}w_i\varepsilon_{ni}\right)\right)\xlongequal{\text{استقلال } \varepsilon_{ni} \text{ ها}}E_D(w)+\frac{1}{2}\sum_{i=1}^{D}w_i^2\,\mathrm{Var}(\varepsilon_{ni})$$

$$\Rightarrow \boxed{\tilde{E}_D(w)=E_D(w)+\frac{1}{2}\sigma^2\sum_{i=1}^{D}w_i^2}$$

Scanned with CamScanner

3 - الف)

$$\begin{bmatrix} p(y=1 \mid x,w) \\ p(y=2 \mid x,w) \\ \vdots \\ p(y=k \mid x,w) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{w_j^T x}} \begin{bmatrix} e^{(w_1^T x)} \\ e^{(w_2^T x)} \\ \vdots \\ e^{(w_k^T x)} \end{bmatrix}$$

ب)

$$L(w_1, \dots, w_{k-1}) = \sum_{i=1}^{n} \ln\left(P(y=y_i \mid X=x_i)\right) = \sum_{i=1}^{n} \ln\left(\frac{e^{w_{y_i}^T x_i}}{\sum_{j=1}^{k} e^{w_j^T x_i}}\right)$$

$$= \sum_{i=1}^{n} \ln\left(e^{w_{y_i}^T x_i}\right) - \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{k-1} e^{w_j^T x_i}\right)$$

$$= \boxed{\sum_{i=1}^{n} w_{y_i}^T x_i - \ln\left(\sum_{j=1}^{k-1} e^{w_j^T x_i}\right)}$$

ج) در ب از notation k استفاده کردیم. گرادیان را برای هرکدام نسبت به w_m محاسبه می‌کنیم:

$$\frac{\partial L(w_1, \dots, w_{k-1})}{\partial w_m} = \frac{\partial \left(\sum_{i=1}^{n} w_{y_i}^T x_i - \ln\left(\sum_{j=1}^{k-1} e^{w_j^T x_i}\right)\right)}{\partial w_m}$$

قاعده زنجیره‌ای:

$$= \underbrace{\frac{\partial}{\partial w_m} \sum_{i=1}^{n} w_{y_i}^T x_i}_{\sum_i x_i (y_i == m)} - \sum_{i=1}^{n} \underbrace{\frac{\partial}{\partial w_m} \ln\left(\sum_{j=1}^{k-1} e^{w_j^T x_i}\right)}_{\sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{k-1} e^{w_j^T x_i}} \times x_i e^{w_j^T x_i}}$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial w_m} = \underbrace{\sum_i x_i (y_i == m)}_{\text{هایی که } y_i = m \text{ است } x_i} - \sum_{i=1}^{n} \frac{x_i e^{w_j^T x_i}}{\sum_{j=1}^{k-1} e^{w_j^T x_i}}}$$

د)

$$\frac{\lambda}{2} \sum_{j=1}^{k-1} w_j^2 = \frac{\lambda}{2} \sum_{j=1}^{k-1} w_j^T w_j \Rightarrow \frac{\partial}{\partial w_m} \frac{\lambda}{2} \sum_{j=1}^{k-1} w_j^T w_j = \frac{\lambda}{2} \times 2 w_m = \lambda w_m$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial w_m} = \sum_i x_i (y_i == m) - \sum_{i=1}^{n} \frac{x_i e^{w_j^T x_i}}{\sum_{j=1}^{k-1} e^{w_j^T x_i}} - \lambda w_m}$$

4- الف) مثل این است که دیتای ما تنها x_j باشد. یعنی با باقی x_i ها $i \neq j$ کاری نداریم؛ طبق فرمول جواب مسئله‌ی رگرسیون خطی داریم:

$$w_j = (\underbrace{x_j^T x_j}_{\text{اسکالر}})^{-1} x_j y = \boxed{\frac{x_j y}{x_j^T x_j}}$$

ب) از آنجا که ستون های ماتریس X متعامد هستند، ضرب داخلی آن ها در هم صفر است.

$$\Rightarrow X^T X = \operatorname{diag}(x_1^T x_1, x_2^T x_2, \dots, x_m^T x_m)$$

$$\Rightarrow (X^T X)^{-1} = \operatorname{diag}\left(\frac{1}{x_1^T x_1}, \frac{1}{x_2^T x_2}, \dots, \frac{1}{x_m^T x_m}\right)$$

$$\Rightarrow w = (X^T X)^{-1} X^T y = \operatorname{diag}\left(\frac{1}{x_1^T x_1}, \dots, \frac{1}{x_m^T x_m}\right) X^T y$$

$$\Rightarrow w_j = \left(\operatorname{diag}\left(\frac{1}{x_1^T x_1}, \dots, \frac{1}{x_m^T x_m}\right) X^T y\right)_j = \left(\operatorname{diag}\left(\frac{1}{x_1^T x_1}, \dots, \frac{1}{x_m^T x_m}\right)\right)_j (X^T y)_j$$

$$\Rightarrow w_j = \frac{1}{x_j^T x_j} (X^T y)_j = \frac{(X^T)_j y}{x_j^T x_j} \Rightarrow \boxed{w_j = \frac{x_j^T y}{x_j^T x_j}}$$

که همان جوابی است که در الف بدست آوردیم. پس اثبات کامل شد.

پ)

$$\text{داده‌ها} = \Big[\underbrace{x_j}_{\text{برای } x_j} \;\Big|\; \underbrace{\begin{matrix}1\\ \vdots \\ 1\end{matrix}}_{\text{برای بایاس}}\Big] = X \Rightarrow X^T X = \begin{bmatrix} x_j^T x_j & \operatorname{sum}(x_j) \\ \operatorname{sum}(x_j) & n \end{bmatrix}$$

$$\begin{cases} \Rightarrow (X^T X)^{-1} = \dfrac{1}{n\|x_j\|^2 - (\operatorname{sum}(x_j))^2} \begin{bmatrix} n & -\operatorname{sum}(x_j) \\ -\operatorname{sum}(x_j) & x_j^T x_j \end{bmatrix} \\ \Rightarrow X^T y = \begin{bmatrix} x_j^T y \\ \operatorname{sum}(y) \end{bmatrix} \end{cases}$$

$$[w_j, w_0] = (X^T X)^{-1} X^T y = \frac{1}{n\|x_j\|^2 - (\mathrm{sum}(x_j))^2} \begin{bmatrix} n & -\mathrm{sum}(x_j) \\ -\mathrm{sum}(x_j) & x_j^T x_j \end{bmatrix} \begin{bmatrix} x_j^T y \\ \mathrm{sum}(y) \end{bmatrix}$$

$$\Rightarrow w_j = \frac{n x_j^T y - \mathrm{sum}(x_j)\,\mathrm{sum}(y)}{n\|x_j\|^2 - (\mathrm{sum}(x_j))^2} = \frac{\frac{x_j^T y}{n} - \frac{\mathrm{sum}(x_j)}{n}\,\frac{\mathrm{sum}(y)}{n}}{\frac{\|x_j\|^2}{n} - \frac{(\mathrm{sum}(x_j))^2}{n^2}}$$

$$= \frac{E\{x_j y\} - E\{x_j\}E\{y\}}{E\{x_j^2\} - (E\{x_j\})^2} = \boxed{\frac{\mathrm{cov}(x_j, y)}{\mathrm{Var}(x_j)}}$$

$$w_0 = \frac{\mathrm{sum}(y)\, x_j^T x_j - \mathrm{sum}(x_j)\, x_j^T y}{n\|x_j\|^2 - (\mathrm{sum}(x_j))^2} = \frac{n^2 E\{y\}E\{x_j^2\} - n^2 E\{x_j\}E\{x_j y\}}{n^2\,\mathrm{Var}(x_j)}$$

$$= \frac{E\{y\}E\{x_j^2\} - E\{x_j\}E\{x_j y\}}{\mathrm{var}(x_j)} = \frac{E\{y\}\underbrace{(E\{x_j^2\} - (E\{x_j\})^2)}_{\mathrm{var}(x_j)} + (E\{x_j\})\underbrace{(E\{x_j\}E\{y\} - E\{x_j y\})}_{-\mathrm{cov}(x_j, y)}}{\mathrm{var}(x_j)}$$

$$= E\{y\} - E\{x_j\}\frac{\mathrm{cov}(x_j, y)}{\mathrm{var}(x_j)} = \boxed{E\{y\} - w\, E\{x_j\}}$$

5- الف)

$X \to$ positive random variable

$$E\{X\} = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^{+\infty} x f_X(x)\,dx \geq \int_a^{+\infty} x f_X(x)\,dx$$

$$\geq \int_a^{+\infty} a f_X(x)\,dx = a\int_a^{+\infty} f_X(x)\,dx = a\,P(X \geq a)$$

$$\Rightarrow E\{X\} \geq a\,P(X \geq a) \Rightarrow \boxed{\frac{E\{X\}}{a} \geq P(X \geq a)}$$

ب)

$$P(Y \geq \varepsilon^2) \leq \frac{E\{Y\}}{\varepsilon^2} \qquad Y = (X - E\{X\})^2$$

$$P((X - E\{X\})^2 \geq \varepsilon^2) \leq \frac{E\{(X - E\{X\})^2\}}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

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

✓

ج) X : متغیر تصادفی ۱ در صورت افتادن نقطه‌ی رندوم داخل دایره
و 0 در صورت افتادن خارج دایره

$$X \sim \text{bernouli}\left(\frac{\pi}{4}\right)$$

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{\pi}{4}\right| < 0.01\pi\right] \geq 0.95$$

$$\Rightarrow P\left[\left|\sum_{i=1}^{n} X_i - \frac{n\pi}{4}\right| < 0.01\pi n\right] \geq 0.95$$

$$\varepsilon = 0.01\pi n \qquad \frac{\sigma^2}{\varepsilon^2} = 0.95 \qquad \sigma^2 = n\frac{\pi}{4}\left(1-\frac{\pi}{4}\right)$$

واریانس Y متغیر برنولی در تعداد n تا

$$\Rightarrow \frac{n\frac{\pi}{4}\left(1-\frac{\pi}{4}\right)}{10^{-4}\pi^2 n^2} = 0.95 \Rightarrow n = \left\lceil \frac{\frac{\pi}{4}\left(1-\frac{\pi}{4}\right)}{\pi^2 \times 10^{-4} \times 0.95} \right\rceil \Rightarrow \boxed{n = 180}$$

6- الف)

$$\|A\|_2 = \max_x \frac{\|Ax\|}{\|x\|} = \sqrt{\max \frac{x^TA^TAx}{\underbrace{x^Tx}_{\text{اسکالر}}}} = \sqrt{\max \frac{x^TA^TAx}{\|x^Tx\|}} = \sqrt{\lambda_{max}(A^TA)}$$

$$\Rightarrow \boxed{\|A\|_2 = \sigma_{max}(A)}$$

اگر λ_i مقدار ویژه‌ی B باشد

$$\lambda_i x = Bx \Rightarrow B^{-1}(\lambda_i x) = x \Rightarrow \lambda_i B^{-1}x = x \Rightarrow B^{-1}x = \frac{1}{\lambda_i}x$$

$\frac{1}{\lambda_i}$ مقدار ویژه‌ی B^{-1} است.

$$\|A^{-1}\|_2 = \sqrt{\lambda_{max}((A^{-1})^TA^{-1})} = \sqrt{\lambda_{max}((AA^T)^{-1})} = \sqrt{\frac{1}{\lambda_{min}(AA^T)}} = \sqrt{\frac{1}{\lambda_{min}(A^TA)}}$$

$$\|A^{-1}\|_2 = \sigma_{max}(A^{-1}) = \sqrt{\frac{1}{\lambda_{min}(A^TA)}}$$

$$\boxed{\sigma_i(A) = (\sigma_i(A^{-1}))^{-1}}$$

مقادیر تکین ماتریس A^{-1} معکوس مقادیر تکین ماتریس A هستند.

$$\Rightarrow \sigma_{max}(A)\,\sigma_{max}(A^{-1}) = \sqrt{\frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)}} \geq 1$$

$$\lambda_{max}(A^TA) \geq \lambda_{min}(A^TA) \Rightarrow \frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)} \geq 1$$

ب)

$$\|A\|_F^2 = tr(AA^T)$$

$$AA^T = U\Sigma \underbrace{V^T V}_{I} \Sigma U^T = U\Sigma^2 U^T$$

$$tr(AA^T) = \sum_{i=1}^{n} U_i \Sigma^2 U_i^T = \sum_{i=1}^{m}\sum_{j=1}^{n} \sigma_j^2 u_{ij}^2 = \sum_{j=1}^{n} \sigma_j^2 \sum_{i=1}^{m} u_{ij}^2 \overset{\text{یکه}}{=} \sum_{j=1}^{n} \sigma_j^2$$

$$\Rightarrow \boxed{\|A\|_F^2 = \sum_{j=1}^{n} \sigma_j^2(A)} \xrightarrow[\text{فول رنک است}]{\text{معکوس‌پذیر است پس}} \|A\|_F^2 = \sum_{j=1}^{rank(A)} \sigma_j^2(A)$$

$n = rank(A)$

$$\boxed{\|A\|_2^2 = \sigma_{max}^2(A)} \leftarrow \text{طبق الف}$$

$$\Rightarrow \|A\|_F^2 = \sigma_{max}^2(A) + \sum_{j=1}^{rank(A)-1} \sigma_j^2(A) \geqslant \sigma_{max}^2(A) = \|A\|_2^2$$

$$\Rightarrow \|A\|_F^2 \geqslant \|A\|_2^2 \Rightarrow \boxed{\|A\|_F \geqslant \|A\|_2} \leftarrow \text{نامساوی سمت چپ}$$

$$\|A\|_F^2 = \sum_{j=1}^{rank(A)} \sigma_j^2(A) \leqslant \sum_{j=1}^{rank(A)} \sigma_{max}^2(A) = rank(A)\,\sigma_{max}^2(A) = rank(A)\,\|A\|_2^2$$

رادیکال می‌گیریم

$$\Rightarrow \boxed{\|A\|_F \leqslant \sqrt{rank(A)}\,\|A\|_2} \leftarrow \text{نامساوی راست}$$

پایان

-7

$$2\sigma(2a) - 1 = \frac{2e^{2a}}{e^{2a}+1} - 1 = \frac{e^{2a}-1}{e^{a}+e^{-a}} = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} = \tanh(a)$$

$$y(x,u) = u_0 + \sum_{j=1}^{n} u_j \tanh\left(\frac{x-\mu_j}{s}\right) = u_0 + \sum_{j=1}^{n} u_j \left(2\sigma\left(2\frac{x-\mu_j}{s}\right) - 1\right)$$

$$= \underbrace{u_0 - \sum_{j=1}^{n} u_j}_{w_0} + \sum_{j=1}^{n} \underbrace{2u_j}_{w_j} \sigma\left(2\frac{x-\mu_j}{s}\right) \checkmark$$