1) $P(x_1|\Omega_i)\, P(x_2|\Omega_i) = P(x|\Omega_i)$

$$P(\Omega_j|x) = \frac{P(x_1|\Omega_i)\, P(x_2|\Omega_i)\, \cancel{P(\Omega_j)}^{\frac{1}{3}}}{P(x)}$$

$$P(\Omega_1|x) = \frac{P(x|\Omega_1) \times \frac{1}{3}}{P(x)}$$

$x_1 = 50 \quad x_2 = 0.5$

$$P(x|\Omega_1) = \frac{1}{\sqrt{\det(2\pi\Sigma)}}\, e^{\left[-\frac{1}{2}\begin{bmatrix} x_1-0 \\ x_2-0 \end{bmatrix}^T \begin{bmatrix} 1.42 & 0 \\ 0 & 1.42 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right]}$$

$$= \frac{1}{0.7\sqrt{2\pi}}\, e^{\left[-\frac{1}{2}\begin{bmatrix} 50 \\ 0.5 \end{bmatrix}^T \begin{bmatrix} 1.428 & 0 \\ 0 & 1.428 \end{bmatrix} \begin{pmatrix} 50 \\ 0.5 \end{pmatrix}\right]}$$

$$= \frac{1}{0.7\sqrt{2\pi}}\, e^{[-1785 - 0.1785]} \simeq 0$$

$$P(x|\Omega_2) = \frac{1}{\det(2\pi\Sigma_2)}\, e^{-\frac{1}{2}\left[\begin{pmatrix} 50-1 \\ 0.5-1 \end{pmatrix}^T \begin{bmatrix} 2.857 & -4.285 \\ -4.285 & 11.428 \end{bmatrix} \begin{bmatrix} 49 \\ -0.5 \end{bmatrix}\right]}$$

$$= \frac{1}{\sqrt{2\pi\, 0.07}}\, e^{-\frac{1}{2}(13824.149 + 212)} \simeq 0$$

$$P(x|\Omega_3) = \frac{1}{\sqrt{2\pi\, 0.52}}\, e^{-\frac{1}{2}\left((49 \quad -0.5)\begin{pmatrix} 1.538 & -0.384 \\ -0.384 & 1.346 \end{pmatrix}\begin{pmatrix} 49 \\ -0.5 \end{pmatrix}\right)}$$

$$= \frac{1}{\sqrt{2\pi\, 0.52}}\, e^{-\frac{1}{2}(3702 + 6.33)} \simeq 0$$

کلاس ۳ $\Leftarrow P(x|\Omega_3) >$ باقی

$x_1 = 0{,}5 \quad , \quad x_2 = 0{,}5$

$$P(x \mid \Omega_1) = \frac{1}{0{,}7\sqrt{2\pi}}\, e^{-\frac{1}{2}\left((0{,}5 \;\; 0{,}5)\begin{pmatrix} 1.42 & 0 \\ 0 & 1.42 \end{pmatrix}\begin{pmatrix} 0{,}5 \\ 0{,}5 \end{pmatrix}\right)}$$

$$= \frac{1}{0{,}7\sqrt{2\pi}}\, e^{-\frac{1}{2}(0.71)} = 0.39$$

$$P(x \mid \Omega_2) = \frac{1}{\sqrt{2\pi\, 0{,}07}}\, e^{-\frac{1}{2}\left((-0{,}5 \;\; -0{,}5)\begin{pmatrix} 2.857 & -4.285 \\ -4.285 & 11.428 \end{pmatrix}\begin{pmatrix} 0{,}5 \\ 0{,}5 \end{pmatrix}\right)}$$

$$= \frac{1}{\sqrt{2\pi\, 0.07}}\, e^{-\frac{1}{2}(1.428)} = 0.738$$

$$P(x \mid \Omega_3) = \frac{1}{\sqrt{2\pi \cdot 0{,}52}}\, e^{-\frac{1}{2}\left((-0{,}5 \;\; -0{,}5)\begin{pmatrix} 1.538 & -0.384 \\ -0.384 & 1.346 \end{pmatrix}\begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}\right)}$$

$$= \frac{1}{\sqrt{2\pi\, 0.52}}\, e^{-\frac{1}{2}(0.529)} = 0.424$$

2) $y(x_n, w) = w_0 + \sum_{j=1}^{D} w_i x_{ni}$ $\qquad \epsilon_i \sim N(0, \sigma^2)$

$x_i + \epsilon_i \rightarrow y(x_n, w) + \delta_n(x_i, \epsilon_i)$

$$\tilde{y}(x_n, w) = \underbrace{w_0 + \sum_{j=1}^{D} w_i x_{ni}}_{y(x_n, w)} + \underbrace{\sum_{i=1}^{D} w_i \epsilon_i}_{\delta(n)}$$

$$\tilde{E}_D(w) = \frac{1}{2} E\left\{ \sum \left(y(x_n, w) + \underline{\delta_n} - y_n \right) \right.$$

$$= \frac{1}{2} E\left\{ \sum_{n=1}^{N} \left(y(x_n, w) - y_n \right)^2 + \delta_n^2 - 2\delta_n y_n \right.$$

$$\left. + \underline{2 y(x_n, w) \delta_n} \right\}$$

$E(\epsilon_i) = 0$

$$= E_D(w) + \frac{1}{2} E\left(\sum_{i=1}^{D} w_i^2 \epsilon_i^2 \right) - y_n \cancel{E\left\{ \sum_{i=1}^{D} w_i \epsilon_i \right\}}^{\,0}$$

$$+ E\left\{ \left(w_0 + \sum_{i=1}^{D} w_i x_{ni} \right) \cancel{\left(\sum_{i=1}^{D} w_i \epsilon_i \right)}^{\,0} \right\}$$

$$= E_D(w) + \frac{1}{2} \sum_{i=1}^{D} \sigma^2 w_i^2$$

$$\sigma^2 = E(\epsilon_i^2) - \cancel{E(\epsilon_i)}_{0} \Rightarrow \sigma^2 = E(\epsilon_i^2)$$

3) a) $$P(\Omega_i \mid x) = \frac{e^{(w_i^T x + w_{0,i})}}{\sum_{j=1}^{k} e^{(w_j^T x + w_{0j})}} \quad \text{for } 1 \leqslant i < k$$

عدد ثابت

b) $$\sum_{i=1}^{n} \ln P(\Omega_i \mid x = x_i) = \sum_{i=1}^{n} \ln e^{(w_i^T x + w_{0i})} - \cancel{\ln \sum_{j=1}^{k} e^{(w_j^T x + w_{0j})}}$$

$$= \sum_{i=1}^{n} (w_i^T x + w_{0i}) - \text{Const}$$

c) $$= \frac{\partial}{\partial w_k} \sum_{i=1}^{n} (w_i^T x + w_{0i}) = \delta(k=i)\, x_i$$

d) $$\frac{\partial}{\partial w_k} \sum_{i=1}^{n} (w_i^T x + w_{0i}) - \frac{\lambda}{2} \sum_{j=1}^{k-1} |w_j|_2^2$$

$$= \delta(k=i)(x_i) - \frac{\lambda}{2} \delta(k=i)\, 2 w_k = \delta(k=i)(x_i - \lambda w_k)$$

4) a) in linear regression we aim to find a weight vector w that minimizes the sum of $\|X \underset{}{\overset{\text{weight}}{w}} - y\|^2$ → Labels

X → design matrix

$$\nabla L(w) = 2X^T(Xw - y) = 0$$

$$X^T X w = X^T y \rightarrow w_m = (X_m^T X_m)^{-1} X_m^T y$$

→ using m of the m features

for i feature

$$w_i = (x_j^T x_j)^{-1} x_j^T y$$

$$= \frac{x_i^T y}{x_i^T x_i}$$

b)

Columns of X matrix are orthogonal → independent

the optimization problem for linear regression can be like below

$$w_{all} = \arg\min \|Xw - y\|^2$$

$$X^T X = D \rightarrow \text{diagonal matrix}$$

now consider feature independently:

$$w_i = \arg\min_{w_i} \|x_i w_i - y\|^2$$

$$\frac{d}{dw_i}\|x_i w_i - y\|^2 = 2x_i^T(x_i w_i - y) = 0$$

$$\rightarrow w_i = \frac{x_i^T y}{x_i^T x_i} \rightarrow \text{from previous question}$$

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

$$W = \mathrm{diag}\left(\frac{x_1^T y}{x_1^T x_1} \; \cdots \; \frac{x_m^T y}{x_m^T x_m}\right)$$

you can see that each component of W corresponds to weight obtained when training on each feature

c)

$$y = w_0 + w_j x_j + \overset{\text{error}}{\varepsilon} \rightarrow$$ رگرسیون خطی

mean square error هدف رگرسیون مینیمم کردن

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_j x_{ij})\right)^2$$

$$\frac{\partial}{\partial w_j} \Rightarrow \frac{1}{n}\sum 2x_{ij}(w_0 + w_j x_{ij} - y_i) = 0$$

$$\rightarrow w_j = \frac{\sum_{i=1}^{n} x_{ij}(y_i - w_0)}{\sum_{i=1}^{n} x_{ij}^2} = \frac{\mathrm{Cov}(x_j, y)}{\mathrm{Var}(x_j)}$$

$$\frac{\partial}{\partial w_0} = \frac{1}{n}\sum_{i=1}^{n} 2(w_0 + w_j x_{ij} - y_i) = 0$$

$$w_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - w_j x_{ij}) = E(y) - w_j E(x_j)$$

5) a) $E(X) = \int_0^{\infty} x f(x) dx$

$X \geq \alpha > 0$

$E[X] = \int_0^{\infty} x f(x) dx \geq \int_\alpha^{\infty} x f(x) dx$

$$\frac{E[x]}{\alpha} \geq \frac{1}{\alpha} \int_\alpha^{\infty} x f(x) dx \overset{x \geq \alpha}{\geq} \frac{1}{\alpha} \int_\alpha^{\infty} \alpha f(x) dx)$$

$= P(x \geq \alpha)$ ✓

b) $|x - \mu| \geq \alpha$

$(x-\mu)^2 \geq \alpha^2 \qquad y = (x-\mu)^2$

markov

$$\boxed{P(Y \geq \alpha^2) \leq \frac{E(Y)}{\alpha^2}} \rightarrow$$

$E(y) = E((x-\mu)^2) = V(x) = \sigma^2$

$\rightarrow P(|x-\mu| \geq \alpha) = P((\underbrace{x-\mu)^2}_{y} \geq \alpha^2)$

$\leq \frac{\sigma^2}{\alpha^2}$ ✓

c) montecarlo method

$$\frac{\text{نسبت تعداد نقاط داخل دایره}}{\text{به کل نقاط}} \rightarrow \frac{\pi}{4}$$

$\frac{\pi}{4} = P \qquad$ چبیشف $= P(|Z-\mu| \geq 0.01) \leq \frac{\sigma^2}{0.01} \leq \frac{5}{100}$

95% قطعیت

$$E(X_i) = \pi/4 \times 1 + (1 - \pi/4) \times 0 = \pi/4$$

$$V(X_i) = \pi/4 \cdot (1 - \pi/4)$$

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

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

$$P(|\hat{\pi}(n) - \pi| \geq 0.01) \leq 0.05$$

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

$$\rightarrow \frac{\pi(4-\pi)}{0.05(0.01)^2} \leq n \rightarrow n \geq 539{,}353.24$$

6) a) $A = U \Sigma V^T \rightarrow \sigma_{max}(A) = \sigma_1 \rightarrow$ مقادیر ویژه روی قطر اصلی

$$A^{-1} = V \Sigma^{-1} U^T \qquad \sigma_{max}(A^{-1}) = \sigma_2$$

$$A A^{-1} = I \qquad \sigma_{min}(A^{-1}) = \frac{1}{\sigma_1}$$

$$\sigma_{max}(A) \times \sigma_{min}(A^{-1}) = \frac{1}{\sigma_1} \times \sigma_1 = 1$$

$$\xrightarrow{\sigma_{max}(A^{-1}) \geq \sigma_{min}(A^{-1})} \sigma_{max}(A) \times \sigma_{max}(A^{-1}) \geq 1$$

b) Consider the singular value for $X = \sigma_1, \dots, \sigma_r$ where $r = \text{rank}(X)$ such that $\sigma_1 \geq \dots \geq \sigma_r > 0$

$$\|A\|_F = \text{Frobenius norm} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_1$$

$$\rightarrow \sigma_1 \leq \sqrt{\sum_{i=1}^{r} \sigma_i^2} = \sqrt{\underbrace{\sigma_1^2 + \dots \sigma_r^2}_{\text{بزرگترین}}}$$

$$\max(\|A\|_F) \xrightarrow{\sigma_1 = \sigma_2 = \dots = \sigma_r} = \sqrt{r \times \sigma_1^2} = \sqrt{r}\, \sigma_1$$

$$= \sqrt{r}\, \|A\|_2$$

7) $y(x, w) = w_0 + \sum_{j=1}^{n} \left[w_j \, \sigma\left(2 \frac{x-\mu_j}{s}\right) \right]$

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

$$\tanh = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad \frac{x-\mu_j}{s} = k$$

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^{x}}{e^{x}+1}$$

$$y(x, w) = w_0 + \sum_{j=1}^{n} \left[w_j \frac{1}{1+e^{-2k}} \right]$$

$$y(x, u) = u_0 + \sum_{j=1}^{n} \left[u_j \left(\frac{1}{1+e^{-2k}} - \frac{e^{-2k}}{1+e^{-2k}} \right) \right]$$

$$\frac{e^{-2k}}{1+e^{-2k}} = 1 - \frac{1}{1+e^{-2k}}$$

$$\longrightarrow = u_0 + \sum_{j=1}^{n} \left[u_j \left(\frac{2}{1+e^{-2k}} - 1 \right) \right]$$

$$= w_0 + \sum_{j=1}^{n} \left[w_j \frac{1}{1+e^{-2k}} \right)$$

$$2u_j = w_j$$

$$u_0 - \sum u_j = w_0 \longrightarrow$$

therefore equal