تمرین سری ۴ درس ML - رادین خیام - 99101579

سؤال ۱-

$$J(\underline{w}) = E\left[(y-\underline{w}^T\underline{x})^2\right] = E[y^2] - 2\underline{w}^T \underbrace{E[y\underline{x}]}_{\hat{R}_{\underline{x}y}} + \underline{w}^T \underbrace{E[\underline{x}\underline{x}^T]}_{\hat{R}_x}\underline{w}$$

$$\longrightarrow J(\underline{w}) = \overbrace{E[y^2]}^{\sigma_y^2+m_y^2} - 2\underline{w}^T\hat{R}_{\underline{x}y} + \underline{w}^T\hat{R}_{\underline{x}}\underline{w}$$

$$J_2(\underline{w}) = E\left[(y-\underline{w}^T(\underline{x}+\underline{\varepsilon}))^2\right] = E[y^2] - 2\underline{w}^T\hat{R}_{\underline{x}y} - 2\underline{w}^T\underbrace{E[y\underline{\varepsilon}]}_{\cancel{E[y]E[\underline{\varepsilon}]}\,0} + \underline{w}^T\hat{R}_{\underline{x}}\underline{w} + \underline{w}^T\hat{R}_{\underline{\varepsilon}}\underline{w}$$

$$\longrightarrow J_2(\underline{w}) = E[y^2] - 2\underline{w}^T\hat{R}_{\underline{x}y} + \underline{w}^T\hat{R}_{\underline{x}}\underline{w} + \sigma^2\underline{w}^T w$$

$$\longrightarrow J_2(\underline{w}) = J(\underline{w}) + \sigma^2\underline{w}^T\underline{w}$$

سوال 2 -

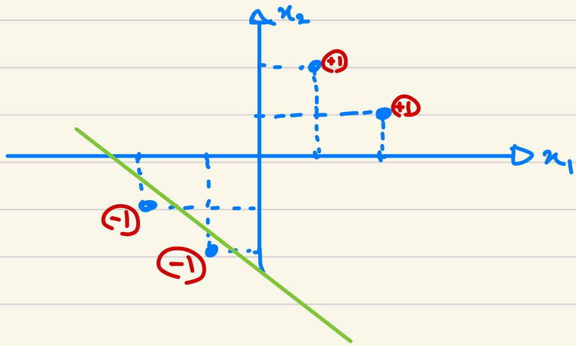
الف)

$\underline{w}_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \longrightarrow \left(\begin{bmatrix} -1 \\ -2 \end{bmatrix}, -1\right), \left(\begin{bmatrix} -2 \\ -1 \end{bmatrix}, -1\right)$ * غلط طبقه بندی می‌شوند.

$$\underline{w}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - (0.1)(1)\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} - (0.1)(1)\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.3 \\ 0.8 \end{bmatrix} \rightarrow -0.3-0.6+0.8=-0.1<0 \checkmark$$

همه داده ها درست طبقه بندی می‌شوند پس الگوریتم متوقف می‌شود.

$$0.3x_1 + 0.3x_2 + 0.8 = 0 \longrightarrow x_2 = -x_1 - \frac{8}{3}$$



ب)

$$\underline{w}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - (0.1)\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.9 \end{bmatrix}$$

$$\underline{w}_2 = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.9 \end{bmatrix} - (0.1)\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.3 \\ 0.8 \end{bmatrix}$$

* در انتهای ایپاک اول مشاهده می‌شود که با $\underline{w}_2$ تمام داده ها به درستی طبقه بندی می‌شوند پس الگوریتم متوقف می‌شود. شکل مرز مشابه الف می‌باشد.

پ)

$$X = \begin{bmatrix} 1 & 2 & -1 & -2 \\ 2 & 1 & -2 & -1 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}, \quad \underline{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\begin{array}{cccc} \underset{\sim}{\underline{x}_1} & \underset{\sim}{\underline{x}_2} & \underset{\sim}{\underline{x}_3} & \underset{\sim}{\underline{x}_4} \end{array} \qquad \begin{array}{cccc} \underset{\sim}{y_1} & \underset{\sim}{y_2} & \underset{\sim}{y_3} & \underset{\sim}{y_4} \end{array}$$

$$L(\underline{w}, w_0, \lambda) = \frac{1}{2}\underline{w}^T\underline{w} - \sum_{i=1}^{N} \lambda_i \left(y_i (\underline{w}^T \underline{x}_i + w_0) - 1 \right)$$

$$= \frac{w_1^2}{2} + \frac{w_2^2}{2} - \lambda_1 (w_1 + 2w_2 + w_0 - 1) - \lambda_2 (2w_1 + w_2 + w_0 - 1)$$

$$- \lambda_3 (w_1 + 2w_2 - w_0 - 1) - \lambda_4 (2w_1 + w_2 - w_0 - 1)$$

شرایط KKT :

$$KKT \begin{cases} \nabla_{\underline{w}, w_0} L(\underline{w}, w_0, \lambda) = 0 \\ \lambda_i \geq 0 \ , \ i = 1, 2, \dots, N \\ \lambda_i \left(y_i (\underline{w}^T \underline{x}_i + w_0) - 1 \right) = 0 , \ i = 1, 2, \dots, N \end{cases}$$

$$\frac{\partial}{\partial w_1} L = 0 \longrightarrow \boxed{w_1 = \lambda_1 + 2\lambda_2 + \lambda_3 + 2\lambda_4}$$

$$\frac{\partial}{\partial w_2} L = 0 \longrightarrow \boxed{w_2 = 2\lambda_1 + \lambda_2 + 2\lambda_3 + \lambda_4}$$

$$\frac{\partial}{\partial w_0} L = 0 \longrightarrow -\lambda_1 - \lambda_2 + \lambda_3 + \lambda_4 = 0 \longrightarrow \boxed{\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4}$$

$$\boxed{\lambda_1 (w_1 + 2w_2 + w_0 - 1) = 0} \qquad \boxed{\lambda_3 (w_1 + 2w_2 - w_0 - 1) = 0}$$

$$\boxed{\lambda_2 (2w_1 + w_2 + w_0 - 1) = 0} \qquad \boxed{\lambda_4 (2w_1 + w_2 - w_0 - 1) = 0}$$

$$\boxed{\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0}$$

الف- سؤال 2 قسمت پ)

- چون داده ها خطی جدای پذیر هستند پس جواب یکتا خواهیم داشت .

از طرفی به شکل واضح مشخص است که بهترین خط به ازای $\underline{w}=\begin{bmatrix}-1\\1\end{bmatrix}$ و $w_0=0$ بدست میاید کافی است این مقادیر را در معادلات بگذاریم تا به نتایج زیر برسیم :

$$\lambda_1+2\lambda_2+\lambda_3+2\lambda_4=-1$$
$$2\lambda_1+\lambda_2+2\lambda_3+\lambda_4=1$$
$$\lambda_1+\lambda_2=\lambda_3+\lambda_4$$

این معادلات چندین جواب خواهند داشت ما برای مثال یک دسته از ضرایب را بدست میاوریم :

$$\boxed{\lambda_1=1\ ,\ \lambda_3=0,\ \lambda_4=0\ ,\ \lambda_2=-1}$$

سئوال 3 -

الف)

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad , \quad \underline{y} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$$\underline{w} = (XX^T)^{-1} X \underline{y}$$

$$XX^T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 4 \end{bmatrix} \rightarrow (XX^T)^{-1} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

$$\rightarrow (XX^T)^{-1} X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \rightarrow (XX^T)^{-1} X \underline{y} = \underline{w} = \begin{bmatrix} 1 \\ 1 \\ -\frac{3}{2} \end{bmatrix}$$

$$g(\underline{x}) = \underline{w}^T \underline{x} \rightarrow g\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = x_1 + x_2 - 1.5$$

$$\rightarrow \boxed{x_1 + x_2 - 1.5 = 0}$$ معادله خط مرز

ب)

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad , \quad \underline{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \rightarrow (XX^T)^{-1} X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

$$\rightarrow \underline{w} = (XX^T)^{-1} X \underline{y} = \begin{bmatrix} 1 \\ 1 \\ -\frac{1}{2} \end{bmatrix} \rightarrow g\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = x_1 + x_2 - 0.5$$

$$\rightarrow \boxed{x_1 + x_2 - 0.5 = 0}$$

سؤال ۹ -

$$P(\omega_1) = P(\omega_2) , \quad P(\underline{x} | \omega_i) \sim \mathcal{N}(\underline{\mu}_i , \Sigma) , \quad \Sigma_1 = \Sigma_2 = \Sigma$$

برای طبقه بندی کننده بیز داشتیم:

$$g_i(\underline{x}) = P(\omega_i | \underline{x})$$

$$g_i(\underline{x}) = P(\omega_i) P(\underline{x} | \omega_i)$$

$$g_i(\underline{x}) = \ln P(\omega_i) + \ln P(\underline{x} | \omega_i)$$

$$= \ln(P(\omega_i)) + \ln\left(\frac{1}{\sqrt{(2\pi)^2 |\Sigma_i|}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i)}\right)$$

$$P(\omega_1) = P(\omega_2) , \ \Sigma_1 = \Sigma_2 = \Sigma \longrightarrow \boxed{g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i)}$$

برای مرز تصمیم: $g_1(\underline{x}) = g_2(\underline{x}) \longrightarrow g_1(\underline{x}) - g_2(\underline{x}) = 0$

$$\longrightarrow -\frac{1}{2}\left((\underline{x} - \underline{\mu}_1)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_2)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_2)\right) = 0$$

$$\longrightarrow -\frac{1}{2}\Big(\underbrace{-\underline{x}^T \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1^T \Sigma^{-1} \underline{x}}_{-2\underline{\mu}_1^T \Sigma^{-1} \underline{x}} + \underline{\mu}_1^T \Sigma^{-1} \underline{\mu}_1 + 2\underline{\mu}_2^T \Sigma^{-1} \underline{x} - \underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2 = 0$$

$$\longrightarrow \underline{\mu}_1^T \Sigma^{-1} \underline{x} - \underline{\mu}_2^T \Sigma^{-1} \underline{x} - \frac{1}{2}\left(\underline{\mu}_1^T \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2\right) = 0$$

$$\longrightarrow \left(\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2)\right)^T \underline{x} - \frac{1}{2}\left(\underline{\mu}_1^T \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2\right) = 0$$

$$\longrightarrow \boxed{\underline{w}^T \underline{x} + w_0 = 0 , \quad \underline{w} = \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) , \quad w_0 = -\frac{1}{2}\left(\underline{\mu}_1^T \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2^T \Sigma^{-1} \underline{\mu}_2\right)}$$

ادامه سؤال ۴

حال برای LD خواهیم داشت:

$$\ln \frac{P(\omega_1|\underline{x})}{P(\omega_2|\underline{x})} = \underline{W}^T\underline{x} + W_0$$

$$P(\omega_1|\underline{x}) + P(\omega_2|\underline{x}) = 1$$

$$\longrightarrow \begin{cases} P(\omega_1|\underline{x}) = \dfrac{\exp(\underline{W}^T\underline{x} + W_0)}{1+\exp(\underline{W}^T\underline{x} + W_0)} \\ P(\omega_2|\underline{x}) = \dfrac{1}{1+\exp(\underline{W}^T\underline{x} + W_0)} \end{cases}$$

$$\ln\left(\frac{P(\omega_1|\underline{x})}{P(\omega_2|\underline{x})}\right) = \ln\left(\frac{P(\underline{x}|\omega_1)}{P(\underline{x}|\omega_2)}\right) + \cancel{\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right)}^{0}$$

$$\ln\left(\frac{P(\underline{x}|\omega_1)}{P(\underline{x}|\omega_2)}\right) = \ln\left(\frac{\exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_1)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_1)\right)}{\exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_2)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right)}\right)$$

$$= -\frac{1}{2}(\underline{x}-\underline{\mu}_1)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_2)$$

$$= (\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}\underline{x} - \frac{1}{2}\left(\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1 - \underline{\mu}_2^T\Sigma^{-1}\underline{\mu}_2\right)$$

$$\longrightarrow \ln\left(\frac{P(\omega_1|\underline{x})}{P(\omega_2|\underline{x})}\right) = \underline{W}^T\underline{x} + W_0 = (\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}\underline{x} - \frac{1}{2}\left(\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1 - \underline{\mu}_2^T\Sigma^{-1}\underline{\mu}_2\right)$$

$$\longrightarrow \boxed{\underline{W} = (\underline{\mu}_1-\underline{\mu}_2)\Sigma^{-1} \quad , \quad W_0 = -\frac{1}{2}\left(\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1 - \underline{\mu}_2^T\Sigma^{-1}\underline{\mu}_2\right)}$$

← پاسخ LD و طبقه بند بیز در این حالت یکسان شدند.

ادامه سوال ۹- (با استفاده از مشتق گیری)

$$\ln \frac{P(\omega_1 \mid \underline{x})}{P(\omega_2 \mid \underline{x})} = \underline{W}^T \underline{x} + w_0$$

$$P(\omega_1 \mid \underline{x}) + P(\omega_2 \mid \underline{x}) = 1$$

$$\longrightarrow \begin{cases} P(\omega_1 \mid \underline{x}) = \dfrac{\exp(\underline{W}^T \underline{x} + w_0)}{1 + \exp(\underline{W}^T \underline{x} + w_0)} \\[2ex] P(\omega_2 \mid \underline{x}) = \dfrac{1}{1 + \exp(\underline{W}^T \underline{x} + w_0)} \end{cases}$$

$$L(\theta) = \ln\left(\prod_{m=1}^{2}\left(\prod_{k=1}^{N_m} P(\underline{x}_k^{(m)} \mid \omega_m ; \theta_m)\right)\right)$$

$$P(\underline{x}_k^{(m)} \mid \omega_m ; \theta_m) = \frac{P(\underline{x}_k^{(m)})\, P(\omega_m \mid \underline{x}_k^{(m)} ; \theta_m)}{P(\omega_m)}$$

$$\longrightarrow L(\theta) = \sum_{m=1}^{2}\left(\sum_{k=1}^{N_m} \ln P(\omega_m \mid \underline{x}_k^{(m)} ; \theta_m)\right) + \ln \frac{\prod_{m=1}^{2}\left(\prod_{k=1}^{N_m} P(\underline{x}_k^{(m)})\right)}{\prod_{m=1}^{2}\left(P(\omega_m)\right)^{N_m}}$$

$$= \sum_{k=1}^{N_1} \ln\left(\frac{\exp(\underline{W}^T \underline{x}_k + w_0)}{1 + \exp(\underline{W}^T \underline{x}_k^{(1)} + w_0)}\right) + \sum_{k=1}^{N_2} \ln\left(\frac{1}{1 + \exp(\underline{W}^T \underline{x}_k^{(2)} + w_0)}\right)$$

$$+ \ln \frac{\prod_{m=1}^{2}\left(\prod_{k=1}^{N_m} P(\underline{x}_k^{(m)})\right)}{\prod_{m=1}^{2}\left(P(\omega_m)\right)^{N_m}}$$

این ترم اصن مستقل از $\underline{W}$ و w_0 هست

حال باید مشتق گیری را نسبت به متغیرهای $\underline{W}$ و w_0 انجام دهیم:

$$\nabla_{\underline{W}} L(\theta) = 0$$

$$\frac{\partial}{\partial w_0} L(\theta) = 0$$

$$A_1 = \exp\left(\underline{w}^T \underline{x}_k^{(1)} + w_0\right)$$

$$A_2 = \exp\left(\underline{w}^T \underline{x}_k^{(2)} + w_0\right)$$

برای نوشتن ساده‌تر

$$\frac{\partial}{\partial w_0} L(\theta) = \sum_{k=1}^{N_1} \frac{1+A_1}{A_1} \times \frac{\left(A_1(1+A_1) - A_1^2\right)}{(1+A_1)^2} + \sum_{k=1}^{N_2} (1+A_2) \times \frac{-A_2}{(1+A_2)^2} = 0$$

$$\rightarrow \sum_{k=1}^{N_1} \frac{1}{1+A_1} = \sum_{k=1}^{N_2} \frac{A_2}{1+A_2} \rightarrow \sum_{k=1}^{N_1} \frac{1}{1+\exp\left(\underline{w}^T x_k^{(1)} + w_0\right)} = \sum_{k=1}^{N_2} \frac{\exp\left(\underline{w}^T x_k^{(2)} + w_0\right)}{1+\exp\left(\underline{w}^T x_k^{(2)} + w_0\right)}$$

سوال 5 -

$$P(\omega_1) = P(\omega_2) \quad , \Sigma_1 = \Sigma_2 = \Sigma \quad , P(\underline{x}|\omega_i) \sim N(\underline{\mu}_i, \Sigma)$$

$$\begin{cases} \underline{x}_k \in \omega_1 , & y_k = -\dfrac{N_1+N_2}{N_1} \\ \underline{x}_k \in \omega_2 , & y_k = -\dfrac{N_1+N_2}{N_2} \end{cases} \quad , \quad L=1 \rightarrow \underline{x}_k \in \mathbb{R}^2$$

روش SES : $\boxed{N_1+N_2=N}$

$$X = \overbrace{\begin{bmatrix} x_1 & x_2 & \cdots\cdots & x_N \\ 1 & 1 & \cdots\cdots & 1 \end{bmatrix}}^{N_1+N_2} \qquad \underline{y} = \begin{bmatrix} -\frac{N}{N_1} \\ \vdots \\ -\frac{N}{N_1} \\ -\frac{N}{N_2} \\ \vdots \\ -\frac{N}{N_2} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} N_1 \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} N_2 \end{matrix}$$

$$\nabla_{\underline{w}} J(\underline{w}) = 0 \longrightarrow \underline{w} = (XX^T)^{-1} X \underline{y}$$

$$X\underline{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} -\frac{N}{N_1} \\ \vdots \\ -\frac{N}{N_1} \\ -\frac{N}{N_2} \\ \vdots \\ -\frac{N}{N_2} \end{bmatrix} = \begin{bmatrix} -\frac{N}{N_1} \sum_{k \in \omega_1} x_k - \frac{N}{N_2} \sum_{k \in \omega_2} x_k \\ -\frac{N}{N_1} \times N_1 - \frac{N}{N_2} \times N_2 \end{bmatrix} = \begin{bmatrix} -N \overbrace{(\hat{\mu}_1 + \hat{\mu}_2)}^{\hat{\mu}} \\ -2N \end{bmatrix}$$

$$XX^T = \begin{bmatrix} \overbrace{\sum x_k^2}^{N(\hat{\sigma}^2+\hat{\mu}^2)} & \overbrace{\sum x_k}^{N\hat{\mu} = N(\hat{\mu}_1+\hat{\mu}_2)} \\ \sum x_k & N \end{bmatrix} \rightarrow (XX^T)^{-1} = \frac{1}{N^2\hat{\sigma}^2 + \cancel{N^2\hat{\mu}^2} - \cancel{N^2\hat{\mu}^2}} \begin{bmatrix} N & -N\hat{\mu} \\ -N\hat{\mu} & N(\hat{\sigma}^2+\hat{\mu}^2) \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum x_k^2 - \frac{1}{N^2}(\underbrace{\sum x_k}_{N\hat{\mu}})^2 = \frac{1}{N}\sum x_k^2 - \hat{\mu}^2 \rightarrow \sum x_k^2 = N\hat{\sigma}^2 + N\hat{\mu}^2$$

$$\rightarrow \underline{w} = \begin{bmatrix} w \\ w_0 \end{bmatrix} = (XX^T)^{-1} X \underline{y} = \begin{bmatrix} \frac{1}{N\hat{\sigma}^2} & -\frac{\hat{\mu}}{N\hat{\sigma}^2} \\ -\frac{\hat{\mu}}{N\hat{\sigma}^2} & \frac{1}{N} + \frac{\hat{\mu}}{N\hat{\sigma}^2} \end{bmatrix} \begin{bmatrix} -N\hat{\mu} \\ -2N \end{bmatrix}$$

ادامه سؤال 5 -

$$\rightarrow \quad w = \frac{-\hat{\mu}}{\hat{\sigma}^2} + \frac{2\hat{\mu}}{\hat{\sigma}^2} = \frac{\hat{\mu}}{\hat{\sigma}^2} = \frac{\hat{\mu}_1 + \hat{\mu}_2}{\hat{\sigma}^2}$$

$$w_0 = \frac{\hat{\mu}^2}{\hat{\sigma}^2} - 2 - \frac{2\hat{\mu}}{\hat{\sigma}^2} = \frac{\frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)^2 - (\hat{\mu}_1 + \hat{\mu}_2) - 1}{2\hat{\sigma}^2}$$

در حالت LDA داشتیم:

$$W = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}^2} , \quad W_0 = -\frac{\hat{\mu}_1^2 - \mu_2^2}{2\hat{\sigma}^2}$$

سوال 6 -

الف) فقط بردارهای پشتیبان در ساختن مرز مؤثر هستند اما همه بردارهای روی حاشیه لزوماً بردار پشتیبان نیستند.

دلیلش هم این هست که $\lambda_i(\underbrace{y_i(\underline{w}^T\underline{x}_i+w_0)}_{1}-1)$ بنابراین λ_i می‌تواند صفر یا غیر صفر باشد برای همین شاید بردار پشتیبان باشد و شاید نه.

ب)

سؤال 7 -

$Y = f(x) = e^{a+bx} \rightarrow \ln Y = a + bx$

از داده‌های Y، Ln می‌گیریم و مدل تبدیل به یک رگرسیون خطی می‌شود.

$\underbrace{\ln Y}_{Y'} = [0 \quad 0.42 \quad 1 \quad 1.63 \quad 1.89 \quad 1.95]^T$

$X = \begin{bmatrix} 0 & 1 & 2.5 & 3.51 & 4.2 & 7 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

$XY' = \begin{bmatrix} 30.23 \\ 6.89 \end{bmatrix}, \quad XX^T = \begin{bmatrix} 86.21 & 18.21 \\ 18.21 & 6 \end{bmatrix} \rightarrow (XX^T)^{-1} = \begin{bmatrix} 0.032 & -0.098 \\ -0.098 & 0.464 \end{bmatrix}$

$\rightarrow \underline{w} = (XX^T)^{-1} XY = \begin{bmatrix} 0.292 \\ 0.234 \end{bmatrix}$

$\rightarrow \ln Y = 0.292x + 0.234$

$\rightarrow Y = e^{0.292x + 0.234} \quad \rightarrow \boxed{\begin{matrix} a = 0.292 \\ b = 0.234 \end{matrix}}$

سؤال 8 -

الف)

$$I(\eta_i) = \frac{1}{2}(\eta_i)^2$$

$$J(\underline{w}, w_0, \eta) = \frac{1}{2}\|\underline{w}\|^2 + C\sum_{i=1}^{N} I(\eta_i) = \frac{1}{2}\|\underline{w}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\eta_i^2$$

$$\rightarrow L(\underline{w}, w_0, \underline{\eta}, \underline{\lambda}, \underline{\mu}) = \frac{1}{2}\underline{w}^T\underline{w} + \frac{C}{2}\sum_{i=1}^{N}\eta_i^2$$

$$- \sum_{i=1}^{N}\lambda_i\left(y_i(\underline{w}^T x_i + w_0) - 1 + \eta_i\right) - \sum_{i=1}^{N}\mu_i\eta_i$$

ب)

$$\text{KKT}\begin{cases} \nabla_{\underline{w}, w_0, \eta} L(\underline{w}, w_0, \underline{\eta}, \underline{\lambda}, \underline{\mu}) = 0 \\ \begin{cases} \lambda_i \geq 0, & i = 1, 2, \dots, N \\ \mu_i \geq 0, & i = 1, 2, \dots, N \end{cases} \\ \begin{cases} \lambda_i\left(y_i(\underline{w}^T\underline{x}_i + w_0) - 1 + \eta_i\right) = 0 \\ \mu_i\eta_i = 0, & i = 1, 2, \dots, N \end{cases} \end{cases}$$

$$\nabla_{\underline{w}} L = 0 \rightarrow w = \sum_{i=1}^{N}\lambda_i y_i x_i$$

$$\nabla_{w_0} L = 0 \rightarrow \sum_{i=1}^{N}\lambda_i y_i = 0$$

$$\nabla_{\underline{\eta}} L = 0 \rightarrow C\eta_i - \lambda_i - \mu_i = 0$$

پ)

$$w=\sum_{i=1}^{\tilde{N}}\lambda_i x_i y_i \quad , \quad C\eta_i=\lambda_i+\mu_i$$

$$\rightarrow L=\frac{1}{2}\Big(\sum_{i=1}^{\tilde{N}}\lambda_i x_i y_i\Big)^T\Big(\sum_{i=1}^{\tilde{N}}\lambda_i x_i y_i\Big)+\frac{C}{2}\sum_{i=1}^{\tilde{N}}\eta_i^2$$
$$-\sum_{i=1}^{\tilde{N}}\lambda_i\Big(y_i\Big(\big(\sum_{j=1}^{\tilde{N}}\lambda_j x_j y_j\big)^T x_i+w_0\Big)-1+\eta_i\Big)-\sum_{i=1}^{\tilde{N}}\mu_i\eta_i$$

$$\rightarrow L=\sum_{i=1}^{\tilde{N}}\lambda_i-\frac{1}{2}\sum_{i=1}^{\tilde{N}}\sum_{j=1}^{\tilde{N}}\lambda_i\lambda_j y_i y_j x_i^T x_j+\sum_{i=1}^{\tilde{N}}\Big(\frac{C}{2}\eta_i^2-(\mu_i+\lambda_i)\eta_i\Big)$$

$$\rightarrow L=\sum_{i=1}^{\tilde{N}}\lambda_i-\frac{1}{2}\sum_{i=1}^{\tilde{N}}\sum_{j=1}^{\tilde{N}}\lambda_i\lambda_j y_i y_j x_i^T x_j-\frac{C}{2}\sum_{i=1}^{\tilde{N}}\eta_i^2$$

$$\rightarrow \max_{\lambda}\Big(\sum_{i=1}^{\tilde{N}}\lambda_i-\frac{1}{2}\sum_{i=1}^{\tilde{N}}\sum_{j=1}^{\tilde{N}}\lambda_i\lambda_j y_i y_j x_i^T x_j-\frac{C}{2}\sum_{i=1}^{\tilde{N}}\eta_i^2\Big)$$

subject to : $\sum_{i=1}^{\tilde{N}}\lambda_i y_i=0$, $\quad 0\leq\lambda_i\leq C\eta_i$

ت)

چون قید η_i^2 وجود دارد دیگر لازم نیست نگران علامت آن باشیم چون همیشه مثبت است.

سؤال و-

$$\begin{array}{cc} & \begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \end{matrix} \\ \begin{matrix} \text{Class 1} \\ \text{Class 2} \\ \text{Class 3} \\ \text{Class 4} \end{matrix} & \begin{bmatrix} -1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & +1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & +1 \\ -1 & -1 & +1 & -1 & +1 & +1 \end{bmatrix} \end{array} \qquad \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

فاصله همینگ این داده را از کد ورد هر 4 کلاس بدست میاوریم:

$\boxed{\text{Class } 1 = 2}$, Class 2 = 3 , Class 3 = 3 , Class 4 = 3

چون کمترین فاصله همینگ از این کلاس است داده به کلاس 1 تعلق می‌گیرد.

سوال 10-

الف) $\underline{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ ، معادله مرز: $x_2 = -x_1 + 1 \rightarrow x_1 + x_2 - 1 = 0$

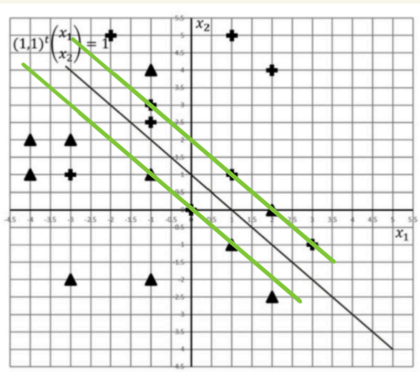
$\rightarrow \underline{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ w_0 = -1$

برای بدست آوردن معادلات حاشیه ها:

$\underline{w}^T \underline{x} + w_0 = 1 \rightarrow x_1 + x_2 - 1 = 1 \rightarrow \boxed{x_1 + x_2 - 2 = 0}$

$\underline{w}^T \underline{x} + w_0 = -1 \rightarrow x_1 + x_2 - 1 = -1 \rightarrow \boxed{x_1 + x_2 = 0}$

ب)



- $\eta_{(2,4)} \boxed{<} \eta_{(2,0)}$
- $\eta_{(-1,1)} \boxed{>} \eta_{(-1,-2)}$
- $\eta_{(-3,1)} \boxed{>} \eta_{(-1,2.5)}$
- $\eta_{(2,4)} \boxed{=} \eta_{(-1,-2)}$

$y_i (x_1 + x_2 - 1)$:

$(2,4) \rightarrow 5 \rightarrow \eta_i = 0$, $(2,0) \rightarrow -1 \rightarrow \eta_i > 1$

$(-1,1) \rightarrow 1 \rightarrow 1 \geq \eta_i > 0$, $(-1,-2) \rightarrow 4 \rightarrow \eta_i = 0$

$(-3,1) \rightarrow -3 \rightarrow \eta_i > 1$, $(-1,2.5) \rightarrow 0.5 \rightarrow 0 < \eta_i \leq 1$

$(2,4) \rightarrow 5 \rightarrow \eta_i = 0$, $(-1,-2) \rightarrow 4 \rightarrow \eta_i = 0$

پ)

$x = (2,4)$ ← چون دور از حاشیه هست و درست طبقه بندی شده ← قطعاً بردار پشتیبان نیست

$x = (1,1)$ ← چون روی حاشیه بالایی هست و درست طبقه بندی شده ← نمی توان گفت

$x = (2,0)$ ← چون روی حاشیه پایینی هست و غلط طبقه بندی شده ← قطعاً بردار پشتیبان است

سؤال ۱۱-

الف)

$$J(\underline{w}) = \underline{w}^T A \underline{w}$$

$$A = U \Lambda U^T \longrightarrow J(\underline{w}) = \underline{w}^T U \Lambda U^T \underline{w}$$

$$\underline{w}(t+1) = \underline{w}(t) - \rho_t \frac{\partial J(\underline{w})}{\partial \underline{w}}$$

$$\Delta \underline{w} = -\mu \frac{\partial J(\underline{w})}{\partial \underline{w}} \Big|_{\underline{w} = \underline{w}(old)}$$

$$\frac{\partial J(\underline{w})}{\partial \underline{w}} = \frac{\partial}{\partial \underline{w}} \underline{w}^T \underbrace{U \Lambda U^T}_{B} \underline{w} = (B + B^T)\underline{w} = 2 \underbrace{U \Lambda U^T}_{A} \underline{w} = \boxed{2 A \underline{w}}$$

$$\longrightarrow \underline{w}(t+1) = \underline{w}(t) - 2\rho_t A \underline{w}(t) \longrightarrow \underline{w}(t+1) = (I - 2\rho_t U \Lambda U^T)\underline{w}(t)$$

ب)

در عمل معمولاً الگوریتم را زمانی که اختلاف $\underline{w}(t+1)$ و $\underline{w}(t)$ از یک حدی کمتر باشد متوقف می‌کنیم، اما برای بررسی همگرایی به صورت ریاضی می‌توانیم بنویسیم:

$$\lim_{t \to \infty} \underline{w}(t+1) = \lim_{t \to \infty} \underline{w}(t) \longrightarrow \lim_{t \to \infty} 2\rho_t A \underline{w} = 0$$

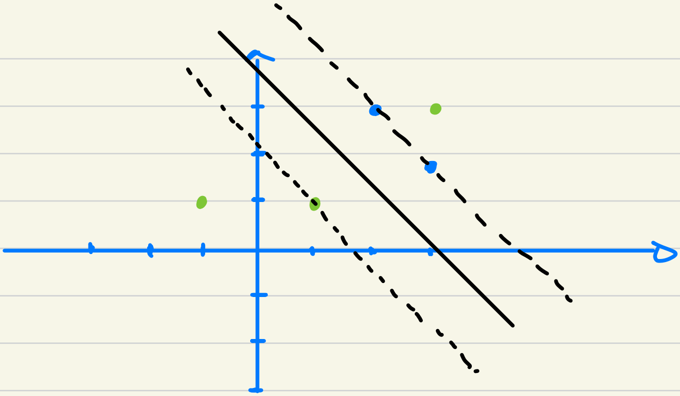
$$\longrightarrow \boxed{\lim_{t \to \infty} \rho_t = 0}$$

* شرط همگرایی به صورت مطلق

$$\boxed{\rho(t) = \frac{1}{t} \quad \text{مثلاً} *}$$

سوال 12 -

الف)



ب)

λ_1 درست می باشد.

λ_2 ← چون داده سوم غلط طبقه بندی می شود و $\mu_3=0$ است پس $C=\lambda_3=10$ باید باشد.

λ_3 ← چون داده پنجم درست طبقه بندی شده است و روی حاشیه نیست پس باید $\lambda_5^{(3)}=0$

λ_4 ← چون λ_1 را صفر گذاشته است ولی این داده یا روی حاشیه و یا داخل حاشیه هست بنابراین $0<\lambda_i<C$ باید باشد.

λ_5 ← چون $\sum_{i=1}^{n}\lambda_i y_i=0$ نقض شده است.

پ)

$\lambda_1=6.89$
$\lambda_2=6.89$
$\lambda_3=10$
$\lambda_4=3.78$
$\lambda_5=1$

$$\rightarrow \quad w_1=2\lambda_1+3\lambda_2-3\lambda_3-\lambda_4+\lambda_5=\boxed{0.67}$$

$$w_2=3\lambda_1+2\lambda_2-3\lambda_3-\lambda_4-\lambda_5=\boxed{0.67}$$

$$y_i(w_1x_1+w_2x_2+w_0-1)=0 \rightarrow \boxed{w_0=-2.35}$$

$$\rightarrow w_1x_1+w_2x_2+w_0=0 \rightarrow \boxed{0.67x_1+0.67x_2-2.35=0}$$

سؤال ۱۳-

الف)

$$X_j = \begin{bmatrix} x_{1,j} & x_{2,j} & \cdots & x_{N,j} \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad w_j = ?$$

$$J(w) = \sum_{k=1}^{N} (y_k - w_j x_{k,j})^2$$

$$\rightarrow \frac{\partial J(w)}{\partial w_j} = \sum_{k=1}^{N} 2(y_k - w_j x_{k,j}) \times -x_{k,j} = 0$$

$$\rightarrow \sum_{k=1}^{N} y_k x_{k,j} = w_j \sum_{k=1}^{N} x_{k,j}^2 \quad \rightarrow w_j = \frac{\sum_{k=1}^{N} y_k x_{k,j}}{\sum_{k=1}^{N} x_{k,j}^2} = \frac{X_j \underline{y}}{X_j X_j^T}$$

ب)

وقتی که ویژگی ها مستقل هستند XX^T یک ماتریس قطری می‌شود.

$$\rightarrow (XX^T)^{-1} = \begin{bmatrix} (X_0X_0^T)^{-1} & 0 & \cdots & 0 \\ 0 & (X_1X_1^T)^{-1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \underbrace{(X_LX_L^T)^{-1}}_{\text{اسکالر است}} \end{bmatrix}, \quad X\underline{y} = \begin{bmatrix} X_0 \underline{y} \\ X_1 \underline{y} \\ \vdots \\ X_L \underline{y} \end{bmatrix}$$

$$\rightarrow \underline{w} = (XX^T)^{-1} X \underline{y} = \begin{bmatrix} (X_0X_0^T)^{-1} X_0 \underline{y} \\ (X_1X_1^T)^{-1} X_1 \underline{y} \\ \vdots \\ (X_LX_L^T)^{-1} X_L \underline{y} \end{bmatrix}$$

→ مشابه با این که به ازای هر ویژگی رگرسیون را انجام دهیم.

ادامه سوال 13 -

پ)

$$X = \begin{bmatrix} x_{1,j} & x_{2,j} & \cdots & x_{N,j} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad , \quad \underline{w} = \begin{bmatrix} w_j \\ w_0 \end{bmatrix}$$

$$\rightarrow XX^T = \begin{pmatrix} \sum_{k=1}^{N} x_{k,j}^2 & \sum_{k=1}^{N} x_{k,j} \\ \sum_{k=1}^{N} x_{k,j} & N \end{pmatrix}$$

$$\rightarrow (XX^T)^{-1} = \frac{1}{N\sum_{k=1}^{N} x_{k,j}^2 - \left(\sum_{k=1}^{N} x_{k,j}\right)^2} \begin{bmatrix} N & -\sum_{k=1}^{N} x_{k,j} \\ -\sum_{k=1}^{N} x_{k,j} & \sum_{k=1}^{N} x_{k,j}^2 \end{bmatrix}$$

$$X\underline{y} = \begin{pmatrix} \sum_{k=1}^{N} x_{k,j}\, y_k \\ \sum_{k=1}^{N} y_k \end{pmatrix}$$

Made with Goodnotes