سوال 1

naive bayes : $\hat{y} = \underset{y}{\arg\max}\; p(y|x) = \underset{y}{\arg\max} \frac{p(x|y)\,p(y)}{p(x)} \xrightarrow[\substack{p(y=1)=p(y=2)=p(y=3) \\ p(x) \text{ is Constant}}]{} \underset{y}{\arg\max}\; p(x|y)$ (I)

$(x|y) \sim N(\mu, \Sigma) \Rightarrow p(x|y) = \frac{1}{(2\pi)^{D/2}\,|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$

$\Rightarrow \ln p(x|y) = -\ln (2\pi)^{D/2}\,|\Sigma|^{1/2} - \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$

$= -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$ (II)

$\xrightarrow{\text{(I), (II)}} \hat{y} = \underset{y}{\arg\max}\; p(x|y) = \underset{y}{\arg\max}\; \ln p(x|y) = \underset{y}{\arg\max} \left\{ -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right\}$

$\xrightarrow{\text{leave out Constants}} = \underset{y}{\arg\max} \left\{ -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right\} = \underset{y}{\arg\min} \left\{ \ln|\Sigma| + (x-\mu)^T \Sigma^{-1}(x-\mu) \right\}$

$|\Sigma_1| = 0.49 \quad , \quad \Sigma_1^{-1} = \begin{bmatrix} 10/7 & 0 \\ 0 & 10/7 \end{bmatrix} \quad , \quad \ln|\Sigma_1| = -0.713$

$|\Sigma_2| = 0.16 - 0.09 = 0.07 \; , \; \Sigma_2^{-1} = \begin{bmatrix} \frac{20}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{80}{7} \end{bmatrix} \; , \; \ln|\Sigma_2| = -2.660$

$|\Sigma_3| = 0.56 - 0.04 = 0.52 \; , \; \Sigma_3^{-1} = \begin{bmatrix} \frac{20}{13} & -\frac{5}{13} \\ -\frac{5}{13} & \frac{35}{26} \end{bmatrix} \; , \; \ln|\Sigma_3| = -0.654$

(ت

$$x=[50, 0.5] \xrightarrow{y=1} (50, 0.5)\begin{pmatrix} 10/7 & 0 \\ 0 & 10/7 \end{pmatrix}\begin{pmatrix} 50 \\ 0.5 \end{pmatrix} = \frac{50005}{14} = 3571.786$$

$$\xrightarrow{y=2} (49, -0.5)\begin{pmatrix} \frac{20}{7} & \frac{-30}{7} \\ \frac{-30}{7} & \frac{80}{7} \end{pmatrix}\begin{pmatrix} 49 \\ -0.5 \end{pmatrix} = \frac{49510}{7} = 7072.857$$

$$\xrightarrow{y=3} (49, -0.5)\begin{pmatrix} \frac{20}{13} & \frac{-5}{13} \\ -5/13 & \frac{35}{26} \end{pmatrix}\begin{pmatrix} 49 \\ -0.5 \end{pmatrix} = \frac{386155}{104} = 3713.029$$

$$\text{argmin}\{3571.786 - 0.713, 7072.857 - 2.660, 3713.029 - 0.654\} = \boxed{1 = \hat{y}}$$

(ب

$$x=[0.5, 0.5] \xrightarrow{y=1} (0.5, 0.5)\begin{pmatrix} 10/7 & 0 \\ 0 & 10/7 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = 5/7 \to 5/7 - 0.713 = 0.001286$$

$$\xrightarrow{y=2} (-0.5, -0.5)\begin{pmatrix} 2/7 & -3/7 \\ -3/7 & 8/7 \end{pmatrix}\begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} = 10/7 \to 10/7 - 2.660 = -1.2314$$

$$\xrightarrow{y=3} (-0.5, -0.5)\begin{pmatrix} 2/13 & -5/13 \\ -5/13 & 35/26 \end{pmatrix}\begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix} = 55/104 \to 55/104 - 0.654 = -0.125$$

$$\xrightarrow{\text{argmin}} \boxed{\hat{y} = 2}$$

سوال 2

$$y(x_n,w)=w_0+\sum_{i=1}^{D} w_i x_{ni} \rightarrow E_D(w)=\frac{1}{2}\sum_{n=1}^{N}\left[y(x_n,w)-y_n\right]^2$$

$$\tilde{y}_n=w_0+\sum_{i=1}^{D} w_i(x_{ni}+\epsilon_{ni})=w_0+\sum_{i=1}^{D} w_i x_{ni}+\sum_{i=1}^{D} w_i\epsilon_{ni}=y(x_n,w)+\sum_{i=1}^{D} w_i\epsilon_{ni}$$

$$\tilde{E}_D(w)=\frac{1}{2}\sum_{n=1}^{N}\left\{\tilde{y}_n-y_n\right\}^2=\frac{1}{2}\sum_{n=1}^{N}\left\{\tilde{y}_n^2-2\tilde{y}_n y_n+y_n^2\right\}$$

$$=\frac{1}{2}\sum_{n=1}^{N}\Big\{y(x_n,w)^2+\underbrace{2y(x_n,w)\sum_{i=1}^{D} w_i\epsilon_{ni}}_{2}+\underbrace{\Big(\sum_{i=1}^{D} w_i\epsilon_{ni}\Big)^2}_{3}-2y(x_n,w)y_n\underbrace{-2\sum_{i=1}^{D} w_i\epsilon_{ni}y_n}_{5}+y_n^2\Big\} \quad \text{I}$$

حال باید از عبارت فوق ($\tilde{E}_D(w)$) امید ریاضی بگیریم.

با توجه به اینکه $\epsilon_i \sim N(0,\sigma^2 I)$ است و $E(\epsilon_i)=0$ پس میانگین جملات 2 و 3 با توجه به خاصیت خطی بودن میانگین صفر می‌شود. حال برای جمله سوم داریم:

$$3 \quad E\left[\Big(\sum_{i=1}^{D} w_i\epsilon_{ni}\Big)^2\right]=E\left[\sum_{i=1}^{D}\sum_{j=1,j\neq i}^{D} w_i\epsilon_{ni}w_j\epsilon_{nj}+\sum_{i=1}^{D} w_i^2\epsilon_{ni}^2\right]$$

$$=\sum_{i=1}^{D}\sum_{j=1,j\neq i}^{D} w_i w_j E(\epsilon_{ni}\epsilon_{nj})+\sum_{i=1}^{D} w_i^2 E(\epsilon_{ni}^2)$$

$$\xrightarrow{\epsilon_i \text{ ها iid}} \left.\begin{array}{l} E(\epsilon_{ni}\epsilon_{nj})=E(\epsilon_{ni})E(\epsilon_{nj})=0 \\ E\left((\epsilon_{ni}-0)^2\right)=var(\epsilon_{ni})=\sigma^2 \end{array}\right\} \rightarrow =0+\sum_{i=1}^{D} w_i^2\sigma^2 \quad \text{II}$$

$$\xrightarrow{\text{I},\text{II}} E(\tilde{E}_D(w))=\frac{1}{2}\sum_{n=1}^{N}\left\{y(x_n,w)^2+0+\sum_{i=1}^{D} w_i^2\sigma^2-2y(x_n,w)y_n-0+y_n^2\right\}$$

$$=\frac{1}{2}\sum_{n=1}^{N}\left\{y(x_n,w)^2-2y(x_n,w)y_n+y_n^2\right\}+\frac{N}{2}\sum_{i=1}^{D} w_i^2\sigma^2$$

$$=\frac{1}{2}\sum_{n=1}^{N}\left[y(x_n,w)-y_n\right]^2+\frac{N\sigma^2}{2}\sum_{i=1}^{D} w_i^2$$

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

سوال 3

الف) به جای اینکه $p(c_1|x)$ را حساب کنیم، احتمالات کلاس‌ها $p(c_i|x)$ را به صورت زیر درنظر می‌گیریم:

$$p(c_i|x) = y_i(x) = \frac{\exp(-w_i^T x)}{\sum_{j=1}^{k} \exp(-w_j^T x)}$$

حال آرا ماتریسی درنظر می‌گیریم که هر سطر آن بردار one-hot انکود شده $t^{(i)}$ است و هر دیتا پوینت را به صورت $x^{(i)}, t^{(i)}$ درنظر می‌گیریم. تابع loss برای این مدل به صورت زیر خواهد بود:

$$p(t^{(i)}|x^{(i)}, w) = y_{t^{(i)}}(x^{(i)})$$

$$L(w) = -\sum_{i=1}^{N} \left(\sum_{k=1}^{N} T_{nk} \log(y_k(x^{(n)})) \right)$$

ب) ابتدا $Y^{(i)}$ را به عنوان بردار one-hot برای $x^{(i)}$ در نظر می‌گیریم. حال $o^{(i)}$ به صورت زیر خواهد بود:

$$o^{(i)} = \begin{bmatrix} w_1^T x^{(i)} \\ w_2^T x^{(i)} \\ \vdots \\ w_{k-1}^T x^{(i)} \\ 0 \end{bmatrix}$$

حال اگر برای $x^{(i)} \in C_k$ برقرار باشد، آنگاه خواهیم داشت:

$$\log p(y^{(i)}|x^{(i)}) = \log \frac{e^{o_k^{(i)}}}{\sum_j e^{o_j^{(i)}}} = o_k^{(i)} - \log\left(\sum_j e^{o_j^{(i)}}\right) = (Y^{(i)})^T o^{(i)} - \log\left(\sum_j e^{o_j^{(i)}}\right)$$

حال برای L خواهیم داشت:

$$L(w_1, w_2, \dots, w_{k-1}) = \sum_{i=1}^{n} \left[(Y^{(i)})^T o^{(i)} - \log\left(\sum_j e^{o_j^{(i)}}\right) \right]$$

پ) گرادیان L را نسبت به w_m ها به دست می‌آوریم:

$$\frac{\partial L}{\partial w_{ml}} = \sum_{i=1}^{n} \left[Y_m^{(i)} x_l^{(i)} - \frac{e^{o_m^{(i)}}}{\sum_j e^{o_j^{(i)}}} x_l^{(i)} \right]$$

$$= \sum_{i=1}^{n} \left[Y_m^{(i)} - P_m^{(i)} \right] x_l^{(i)}$$

حال با $p_m^{(i)} = p(Y=m \mid X = x^{(i)})$ برقرار است که $x^{(i)}$ عضو کلاس m است:

$$\frac{\partial L}{\partial w_j} = \sum_{i=1}^{n} \left(Y_j^{(i)} - P_j^{(i)} \right) x^{(i)} \Rightarrow \frac{\partial L}{\partial w} = (Y - P)^T X$$

که در آن:

$$X = [x^{(1)}, x^{(2)}, \dots, x^{(n)}]^T$$

$$Y = [y^{(1)}, y^{(2)}, \dots, y^{(n)}]^T$$

$$P = [p^{(1)}, p^{(2)}, \dots, p^{(n)}]^T$$

$$W = [w_1, \dots, w_{k-1}, 0]^T$$

ت) گرادیان به صورت مقابل می‌باشد:

$$\boxed{\frac{\partial f}{\partial w} = (Y - P)^T X - \lambda w}$$

Scanned with CamScanner

سوال ۴

الف) در این حالت ماتریس X به صورت x_j خلاصه می‌شود

$$\hat{y} = w^T X \Rightarrow \hat{y} = w_j^T x_j \Rightarrow E(w_j) = (y - w_j^T x_j)^2 = \|y - w_j^T x_j\|^2 = (y - w_j^T x_j)^T (y - w_j^T x_j)$$

$$= (y^T - x_j^T w_j)(y - w_j^T x_j) = y^T y - 2y^T x_j w_j + \underbrace{w_j^T x_j^T x_j w_j}_{= w_j^T w_j x_j^T x_j}$$

$$w_j^* \Rightarrow \frac{\partial E(w_j)}{\partial (w_j)} = 0 \Rightarrow \frac{\partial E(w_j)}{\partial (w_j)} = -2y^T x_j + 2w_j x_j^T x_j = 0$$

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

ب) از آنجایی که x_j ها متعامدند پس برای هر $j \neq i$ داریم: $x_j \cdot x_i = 0$. بنابراین $X^T X$ ماتریس قطری است:

$$X^T X = \mathrm{diag}(x_1^T x_1, x_2^T x_2, \dots, x_m^T x_m) \Rightarrow (X^T X)^{-1} = \mathrm{diag}((x_1^T x_1)^{-1}, \dots, (x_m^T x_m)^{-1})$$

می‌دانیم در حالت کلی که از همه ویژگی‌ها استفاده کنیم مشابه بخش اول w^* به صورت زیر محاسبه می‌شود:

$$y = w^T X \Rightarrow E(w) = \sum_{i=1}^{n} (y^{(i)} - w^T x_i)^2 = \|y - Xw\|^2 = (y - Xw)^T (y - Xw)$$

$$= w^T X^T X w - 2y^T X w + y^T y$$

$$\frac{\partial E(w)}{\partial (w)} = 0 \Rightarrow w X^T X = y^T X \Rightarrow w = (X^T X)^{-1} X^T y$$

بنابراین:

$$w = (X^T X)^{-1} X^T y = \mathrm{diag}((x_1^T x_1)^{-1}, \dots, (x_m^T x_m)^{-1}) X^T y$$

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

$$= (x_j^T x_j)^{-1} (X^T y)_j = \frac{(X^T y)_j}{x_j^T x_j} = \frac{(X^T)_j\, y}{x_j^T x_j} = \frac{x_j^T y}{x_j^T x_j}$$

که همانطور که می‌بینیم با نتیجه بخش الف که حاصل از train کردن روی تنها یک ویژگی بود برابر است و در نتیجه حکم اثبات می‌شود.

2) توجه داریم که در فرم ماتریسی، در حالتی که تنها یکی از ویژگی‌های داده X مثل x_j استفاده کنیم خواهیم داشت:

$$X = \begin{bmatrix} x_j & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \end{bmatrix}_{n\times 2} \rightarrow X^T X = \begin{bmatrix} x_j^T x_j & \mathrm{sum}(x_j) \\ \mathrm{sum}(x_j) & n \end{bmatrix}_{2\times 2}$$

منظور از $\mathrm{sum}(x_j)$ مجموع n تا عضو ستون x_j است.

$$\Rightarrow (X^T X)^{-1} = \frac{1}{n\|x_j\|^2 - \mathrm{sum}(x_j)^2} \begin{bmatrix} n & -\mathrm{sum}(x_j) \\ -\mathrm{sum}(x_j) & x_j^T x_j \end{bmatrix}$$

از طرفی:

$$X^T y = \left(\begin{bmatrix} x_j & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \end{bmatrix} \right)^T y = \begin{bmatrix} x_j^T y \\ \mathrm{sum}(y) \end{bmatrix}$$

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

$$1 \rightarrow w_j = \frac{n x_j^T y - \mathrm{sum}(x_j)\,\mathrm{sum}(y)}{n\|x_j\|^2 - \mathrm{sum}(x_j)^2} = \frac{\frac{x_j^T y}{n} - \frac{\mathrm{sum}(x_j)}{n}\frac{\mathrm{sum}(y)}{n}}{\frac{\|x_j\|^2}{n} - \left(\frac{\mathrm{sum}(x_j)}{n}\right)^2} = \frac{E(x_j y) - E(x_j)E(y)}{E(x_j^2) - E(x_j)^2} = \boxed{\frac{Cov(x_j, y)}{var(x_j)}}$$

$$2 \rightarrow w_0 = \frac{\mathrm{sum}(y)\|x_j\|_2^2 - \mathrm{sum}(x_j)(x_j^T y)}{n\|x_j\|^2 - \mathrm{sum}(x_j)^2} = \frac{\mathrm{sum}(y)\|x_j\|^2 - \mathrm{sum}(x_j)(x_j^T y)}{n\|x_j\|^2 - \mathrm{sum}(x_j)^2}$$

$$= \frac{n^2 E(y) E(x_j^2) - n^2 E(x_j) E(x_j y)}{n^2 var(x_j)} = \frac{E(y) E(x_j^2) - E(x_j) E(x_j y)}{var(x_j)}$$

$$= \frac{E(y)\left(E(x_j^2) - E(x_j)^2\right) + E(y)E(x_j)^2 - E(x_j)E(x_j y)}{var(x_j)} = E(y) + \frac{E(y)E(x_j)^2 - E(x_j)E(x_j y)}{var(x_j)}$$

$$= E(y) + E(x_j)\frac{E(y)E(x_j) - E(x_j y)}{var(x_j)} \overset{1}{=} \boxed{E(y) - w_j E(x_j)}$$

سوال 5

(آ)

$$E(x)=\sum_{x} x\,p(x) \quad , x\geqslant 0$$

$$E(X)=\sum_{x\geqslant 0} x\,p(x) \geqslant \sum_{x\geqslant a} x\,p(x) \geqslant \sum_{x\geqslant a} a\,p(x) = a\sum_{x\geqslant a} p(x) = a\,P(X\geqslant a) \Rightarrow E(X)\geqslant a\,P(X\geqslant a)$$

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

(ب)

$$\begin{cases} Z=(X-\mu)^2 \xrightarrow[Z\geqslant 0]{\text{آ}} \dfrac{E(Z)}{\varepsilon^2} \geqslant P(Z\geqslant \varepsilon^2) \\ E(Z)=E((X-\mu)^2)=\mathrm{var}(x)=\sigma^2 \end{cases}$$

$$\Rightarrow P(|X-\mu|\geqslant\varepsilon)=P(|X-\mu|^2\geqslant\varepsilon^2)=P(Z\geqslant\varepsilon^2)\leqslant\frac{E(Z)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2} \Rightarrow \boxed{P(|X-\mu|\geqslant\varepsilon)\leqslant\frac{\sigma^2}{\varepsilon^2}}$$

(ج)

$$P(|X-\pi/4|\geqslant 0.01\times\pi/4)\leqslant\frac{\sigma^2}{n}\,\frac{1}{(\pi/4\times 0.01)^2}=\frac{\pi/4\,(1-\pi/4)}{n}\times\frac{1}{(\pi/4\times 0.01)^2}\leqslant 0.05$$

$$\Rightarrow \frac{\frac{3.1415}{4}\left(1-\frac{3.1415}{4}\right)}{n}\times\frac{1}{\left(\frac{3.1415}{4}\times 0.01\right)^2}\leqslant 0.05$$

منظور از 6 متغیر برنولی با $P=\pi/4$ است.

$$\Rightarrow \frac{2732.39}{n}\leqslant 0.05 \Rightarrow 54647.9\leqslant n \Rightarrow \boxed{n=54648}$$

سوال 6

(الف)

فرض: $\Sigma=\begin{bmatrix}\sigma_1 & & 0\\ & \sigma_2 & \\ 0 & & \ddots\ \sigma_n\end{bmatrix} \quad , \sigma_1\geqslant\sigma_2\geqslant\sigma_3\geqslant\dots\geqslant\sigma_n>0$

$$\xrightarrow[\Sigma\Sigma^{-1}=I]{} \Sigma^{-1}=\begin{bmatrix}1/\sigma_1 & & 0\\ & 1/\sigma_2 & \\ 0 & & \ddots\ 1/\sigma_n\end{bmatrix} \quad 0<\frac{1}{\sigma_1}\leqslant\frac{1}{\sigma_2}\leqslant\frac{1}{\sigma_3}\leqslant\dots\leqslant\frac{1}{\sigma_n} \quad \text{(1)}$$

$$\sigma(A^{-1}):\ (A^{-1})^T A^{-1}x=\sigma^2(A^{-1})x \longrightarrow (A^{-1})^T A^{-1}x=U(\Sigma^{-1})^T\underbrace{V^T V}_{I}\Sigma^{-1}U^T=U(\Sigma^{-1})^2U^T$$

$$\xrightarrow{\text{ضرب}}\ \text{از } U^T \text{ چپ}\quad \underbrace{U^TU}_{I}(\Sigma^{-1})^2U^T=(\Sigma^{-1})^2U^T=\sigma^2(A^{-1})$$

$$\longrightarrow \underbrace{\left((\Sigma^{-1})^2-\sigma^2(A^{-1})I\right)}_{=0}U^Tx=0 \rightarrow (\Sigma^{-1})^2=\sigma^2A^{-1}I$$

$$\xrightarrow{\text{(1)}} \boxed{\sigma_i^2(A^{-1})=\frac{1}{\sigma_i^2(A)}} \xrightarrow{\sigma_i>0} \sigma_i(A^{-1})=\frac{1}{\sigma_i(A)}$$

$$\sigma_{max}(A)\,\sigma_{max}(A^{-1})=\sigma_{max}(A)\frac{1}{\sigma_{min}(A)}=\frac{\sigma_{max}(A)}{\sigma_{min}(A)}\xrightarrow[(\sigma_{max}(A)/\sigma_{min}(A)\geqslant 1)]{\sigma_{max}(A)\geqslant\sigma_{min}(A)}\geqslant 1 \rightarrow \boxed{\sigma_{max}(A)\sigma_{max}(A^{-1})\geqslant 1}$$

ب)

$$A = U\Sigma V^T = U \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ & & \sigma_r & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} V^T \quad , \; r = \text{rank}(A) \; , \; \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

$$\|A\|_2 = \max_{\|v\|_2 = 1} \|Av\| = \max(\sigma) = \sigma_1$$

$$\|A\|_F^2 = \sum_i \sum_j A(i,j)^2 = tr(A^T A) = \sum_{i=1}^{r} \underbrace{\lambda_i}_{\text{eigenvalue of } A^TA} = \sum_{i=1}^{r} \sigma_i^2$$

$$\|A\|_2 = \sigma_1 = \sqrt{\sigma_1^2} \leq \|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \leq \sqrt{r\,\sigma_1^2} = \sqrt{r}\,.\sigma_1 = \sqrt{r}\,\|A\|_2$$

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$$

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

سوال 7 sigmoid function

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

حال قرار دهیم: $a_j = \dfrac{(x - \mu_j)}{s}$

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

$$= w_0 + \sum_{j=1}^{n} \left[\frac{w_j}{2} \left(2\sigma(2a_j) - 1 + 1 \right) \right] = w_0 + \sum_{j=1}^{n} \frac{w_j}{2} \left(2\sigma(2a_j) - 1 \right) + \sum_{j=1}^{n} \frac{w_j}{2}$$

$$\overset{*}{=} w_0 + \sum_{j=1}^{n} \frac{w_j}{2} \tanh(a_j) + \sum_{j=1}^{n} \frac{w_j}{2}$$

$$u_0 \triangleq w_0 + \sum_{j=1}^{n} \frac{w_j}{2}$$

$$u_j \triangleq \frac{w_j}{2}$$

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

$$= u_0 + \sum_{j=1}^{n} u_j \tanh\left(\frac{x - \mu_j}{s} \right) \quad \square.$$