# HW1-400100662

Thursday, October 12, 2023    3:33 PM

## Q1)

$$P(Y=1) = P(Y=2) = P(Y=3) = \frac{1}{3}$$

$1 \le i \le 3 :\ (X \mid Y=i) \sim \mathcal{N}(\mu_i, \Sigma_i)$ $\quad \mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mu_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\Sigma_1 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.8 \end{bmatrix}$$

$$\Rightarrow (X \mid Y=1) \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix} \right)$$

$\vdots$

---

Bayes Rule:
$$P(Y=c \mid X) = P(X \mid Y=c)\, P(Y=c)$$

multivariate normal distribution:

$$f_X(x_1, x_2, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}{\sqrt{(2\pi)^k |\Sigma|}} \qquad (1)$$

---

فرض استقلال در
naive bayes:
$$P(X \mid y) = \prod_{i=1}^{n} P(x_i \mid y) \qquad (*)$$

$$\rightarrow \max_{c \in Y} P(y=c \mid X) = \max_{c \in Y} P(X \mid y=c)\, P(y=c)$$

$$\overset{(*)}{=} \max_{c \in Y} \prod_{i=1}^{n} P(x_i \mid y=c)\, P(y=c)$$

$$\Rightarrow \max_{c \in Y} P(y=c \mid X) = \max_{c \in Y} \prod_{i=1}^{n} P(x_i \mid y=c)\, P(y=c)$$

اما در واقعیت معمولاً این فرض چندان مناسب نمی باشد و وابستگی بین فیچرهای مختلف وجود خواهد داشت!!

HW1-400100662

Thursday, October 12, 2023 3:33 PM

① $\rightarrow$

$$f_{X|Y=1}(x) = \frac{\exp\left(-\frac{1}{2} x^T \begin{bmatrix} \frac{10}{7} & 0 \\ 0 & \frac{10}{7} \end{bmatrix} x\right)}{2\pi \times 0.7} \qquad (\mathrm{I})$$

$$f_{X|Y=2}(x) = \frac{\exp\left(-\frac{1}{2}\left(x-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix} 2.85 & -4.28 \\ -4.28 & 11.4 \end{bmatrix}\left(x-\begin{bmatrix}1\\1\end{bmatrix}\right)\right)}{2\pi \times \sqrt{0.07}} \qquad (\mathrm{II})$$

$$f_{X|Y=3}(x) = \frac{\exp\left(-\frac{1}{2}\left(x-\begin{bmatrix}1\\1\end{bmatrix}\right)^T \begin{bmatrix} 1.04 & -0.38 \\ -0.38 & 1.35 \end{bmatrix}\left(x-\begin{bmatrix}1\\1\end{bmatrix}\right)\right)}{2\pi \times \sqrt{0.52}} \qquad (\mathrm{III})$$

$$\rightarrow \quad y_x = \underset{y\in\{1,2,3\}}{\arg\max}\ P(Y=y \mid x) = \underset{y\in\{1,2,3\}}{\arg\max}\ P(x \mid Y=y)\underbrace{P(Y=y)}_{\frac{1}{3}} \qquad (\mathrm{IIII})$$

$$= \underset{y\in\{1,2,3\}}{\arg\max}\ P(x \mid Y=y)$$

$\mathrm{I}, \mathrm{II}, \mathrm{III}, \mathrm{IIII} \rightarrow$

$$x_1 = \begin{bmatrix} 5\ \ 0 \\ 0.5 \end{bmatrix} \rightarrow y_{x_1} = 1$$

$$x_2 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \rightarrow y_{x_2} =$$

Q2) $y(x_n, w) = w_0 + \sum_{i=1}^{D} w_i x_{ni}$

$E_D(w) = \frac{1}{2}\sum_{n=1}^{N}\left[y(x_n,w) - y_n\right]^2$ , $\epsilon_i \sim N(0, \sigma^2 I)$

$$\mathbb{E}[\tilde{E}_D(w)] = \mathbb{E}\left[\frac{1}{2}\sum_{n=1}^{N}\left[y(x_n+\epsilon_n, w) - y_n\right]^2\right]$$

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

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

$$\Rightarrow \mathbb{E}\left[\frac{1}{2}\sum_{n=1}^{N}\left[\underbrace{w_0 + \sum_{i=1}^{D} w_i(x_{ni}+\epsilon_{ni})}_{y(x_n+\epsilon_n, w)} - y_n\right]^2\right] = \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\left[\left(\underbrace{y(x_n,w)}_{a} + \underbrace{\sum_{i=1}^{D} w_i\epsilon_{ni}}_{b} \underbrace{-y_n}_{c}\right)^2\right]$$

$$= \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\left[(a+b+c)^2\right] = \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\left[a^2+b^2+c^2+2ab+2ac+2bc\right]$$

$$= \frac{1}{2}\sum_{n=1}^{N}(a^2+c^2+2ac) + \frac{1}{2}\sum_{i=1}^{N}(2a+2c)\underbrace{\cancel{\mathbb{E}[b]}}_{0} + \frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[b^2] \quad (1)$$

$\left(b = \sum_{i=1}^{D} w_i \epsilon_{ni} \rightarrow \mathbb{E}[b] = \sum_{i=1}^{D} w_i \mathbb{E}[\epsilon_{ni}] = 0\right)$

$$\mathbb{E}[b^2] = \mathbb{E}\left[\left(\sum_{i=1}^{D} w_i\epsilon_{ni}\right)\left(\sum_{i=1}^{D} w_i\epsilon_{ni}\right)\right] = \mathbb{E}\left[\sum_{i=1}^{D} w_i^2\epsilon_{ni}^2 + \sum_{i=1}^{D}\sum_{\substack{j=1 \\ j\neq i}}^{D} w_i w_j \epsilon_{ni}\epsilon_{nj}\right]$$

$$= \sum_{i=1}^{D} w_i^2\,\mathbb{E}[\epsilon_{ni}^2] + \sum\sum w_i w_j \underbrace{\cancel{\mathbb{E}[\epsilon_{ni}\epsilon_{nj}]}}_{0}$$

because: $\epsilon_{ni}$ & $\epsilon_{nj}$ are independent $\rightarrow \mathbb{E}[\epsilon_i\epsilon_j] = \mathbb{E}[\epsilon_i]\mathbb{E}[\epsilon_j] = 0\times 0 = 0$

$var(\epsilon_{ni}) = \mathbb{E}[\epsilon_{ni}^2] - \cancel{\mathbb{E}[\epsilon_{ni}]^2}^{0} = \mathbb{E}[\epsilon_{ni}^2] = \sigma^2$

$$\Rightarrow \mathbb{E}[b^2] = \sigma^2\sum_{i=1}^{D} w_i^2 \quad (2)$$

HW1-400100662

Thursday, October 12, 2023 3:33 PM

(1),(2) $\Rightarrow \mathbb{E}[\tilde{E}_D(w)] = \frac{1}{2}\sum_{n=1}^{N}(a^2+c^2+2ac) + \frac{\sigma^2}{2}\sum_{n=1}^{N}\sum_{i=1}^{D} w_i^2$

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

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

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

---

Q3) logistic Regression for binary classification

$\mathbb{E}[y|x] = \sigma(w^T x) = \frac{1}{1+e^{-w^T x}} = \alpha$ ($w$ is learnable vector)

$P(y|x) = \text{Bernouli}(\alpha)$

$$\rightarrow \begin{cases} P(y=1|x) = \frac{1}{1+e^{-w^T x}} \\ P(y=0|x) = 1 - \frac{1}{1+e^{-w^T x}} \end{cases}$$

$$\rightarrow P(y|x) = \sigma(x^T w)^y \left(1 - \sigma(x^T w)\right)^{1-y}$$

الف)

alroach 1: "One_vs_Rest"

در این روش یک کلاس مشخص را به عنوان کلاس Positive و بقیه کلاس ها را به عنوان کلاس negative درنظر میگیریم

و این احتمال را با logistic Regression محاسبه می کنیم و همین عمل را برای تمام کلاس ها انجام می دهیم.

و در نهایت کلاسی که بالاترین احتمال را دارد، به عنوان کلاس مورد نظر predict می کنیم.

approach 2: "Multinomial Logistic Regression" or "softmax Regression"

for each class in $y \in \{1, 2, \dots, k\}$:

$$p(y=c \mid x, W) = \text{softmax}(x^T w_c) = \frac{\exp(x^T w_c)}{1 + \sum_{j=1}^{k-1} \exp(x^T w_j)}$$

where: $W = [w_1 \| w_2 \| \cdots \| w_{k-1} \| 0]$ → concatenation

$$\rightarrow \begin{cases} \text{if } c \neq k: & p(y=c \mid x, W) = \dfrac{\exp(x^T w_c)}{1 + \sum_{j=1}^{k-1} \exp(x^T w_j)} \\ \text{if } c = k: & p(y=c \mid x) = \dfrac{1}{1 + \sum_{j=1}^{k-1} \exp(x^T w_j)} \end{cases}$$

ب) $$L(w_1, w_2, \dots, w_{k-1}) = \sum_{i=1}^{n} \ln p(y=y_i \mid X = x_i) = \sum_{i=1}^{n} \ln\left(\frac{\exp(x_i^T w_{y_i})}{1 + \sum_{j=1}^{k-1} \exp(x_i^T w_j)}\right)$$

$$= \ln\left[\prod_{i=1}^{n}\left(\frac{\exp(x_i^T w_{y_i})}{1 + \sum_{j=1}^{k-1} \exp(x_i^T w_j)}\right)\right] = \ln\left(\prod_{i=1}^{n} \exp(x_i^T w_{y_i})\right) - \ln\left(\prod_{i=1}^{n}\left(\sum_{j=1}^{k-1} \exp(x_i^T w_j)\right)\right)$$

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

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

Thursday, October 12, 2023 3:33 PM

7)
$$L(w_1, \dots, w_{k-1}) = \sum_{i=1}^{n} x_i^T w_{y_i} - \sum_{i=1}^{n} \ln\left[\sum_{j=1}^{k-1} \exp(x_i^T w_j)\right]$$

(*)
$$\frac{\partial L}{\partial w_r} = \sum_{i=1}^{n} \mathbb{1}(y_i = r)\, x_i - \frac{\partial}{\partial w_r} \sum_{i=1}^{n} \ln\left[\sum_{j=1}^{k-1} \exp(x_i^T w_j)\right] \qquad (1 \le r < k)$$

① میدانیم:
$$\frac{\partial}{\partial w_r} \ln\Big[\underbrace{\sum_{j=1}^{k-1} \exp(x_i^T w_j)}_{A}\Big] = \frac{\partial \ln(A)}{\partial w_r} = \frac{1}{A} \frac{\partial A}{\partial w_r}$$

$$\frac{\partial A}{\partial w_r} = \frac{\partial \sum_{j=1}^{k-1} \exp(x_i^T w_j)}{\partial w_r} = \frac{\partial \exp(x_i^T w_r)}{\partial w_r} \overset{۲}{=} \exp(x_i^T w_r)\, x_i$$

$$\frac{\partial \exp(x_{i1} w_{r1} + x_{i2} w_{r2} + \dots)}{\partial w_{rx}} = x_{ix} \exp(x_i^T w_r) \qquad ۲$$

$$\Rightarrow \frac{\partial}{\partial w_r} \ln\left[\sum_{j=1}^{k-1} \exp(x_i^T w_j)\right] = \frac{\exp(x_i^T w_r)}{\sum_{j=1}^{k-1} \exp(x_i^T w_j)}\, x_i$$

(*)
$$\Rightarrow \frac{\partial L}{\partial w_r} = \sum_{i=1}^{n} \left[\mathbb{1}(y_i = r) - \frac{\exp(x_i^T w_r)}{\sum_{j=1}^{k-1} \exp(x_i^T w_j)}\right] x_i \qquad (1 \le r < k)$$

$$\frac{\partial L}{\partial w_k} = 0$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

ت)

$$f(w_1,\dots,w_{k-1}) = L(w_1,\dots,w_{k-1}) - \underbrace{\frac{\lambda}{2}\sum_{j=1}^{k-1} |w_j|_2^2}_{\frac{\lambda}{2}\sum_{j=1}^{k-1} w_j^T w_j}$$

$$\rightarrow \frac{\partial f}{\partial w_r} = \frac{\partial L}{\partial w_r} - \frac{\lambda}{2}\frac{\partial \sum_{j=1}^{k-1} w_j^T w_j}{\partial w_r} = \frac{\partial L}{\partial w_r} - \frac{\lambda}{2}\frac{\partial\, w_r^T w_r}{\partial w_r} = \frac{\partial L}{\partial w_r} - \lambda w_r$$

$$= \sum_{i=1}^{n}\left[ 1(y_i = r) - \frac{\exp(x_i^T w_r)}{\sum_{j=1}^{k-1}\exp(x_i^T w_j)}\right] x_i - \lambda w_r$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

## Q4)

الف)

$$\begin{cases} y \\ \hat{y} = w_i x_i \end{cases}$$

توجه: در رابطه‌ی بالا ترم $w_0$ را در نظر نمی‌گیریم و فرض می‌کنیم که سطر اول ماتریس $X$، بردار تمام $\underline{1}$ است که بیانگر ترم ثابت رگرسیون می‌باشد.

feature mat. $= X_{n\times m}$

$$\hat{y}_{n\times 1} = w_0 \mathbf{1}_{n\times 1} + Xw = \underbrace{\tilde{X}}_{n\times(m+1)} \tilde{w}_{(m+1)\times 1} \qquad (\mathrm{I})$$

$$\tilde{w} = \begin{bmatrix} w_0 \\ w \end{bmatrix} \qquad \tilde{X} = \begin{bmatrix} \mathbf{1}_{1\times n} & X \end{bmatrix}_{n\times(m+1)}$$

MSE loss: $L = \frac{1}{N}\|y-\hat{y}\|^2$

$$\rightarrow \frac{\partial L}{\partial w_i} = 0 \implies \frac{\partial \|y-\hat{y}\|^2}{\partial w_i} = \frac{\partial (y-\hat{y})^T(y-\hat{y})}{\partial w_i} = \frac{\partial (y^T-\hat{y}^T)(y-\hat{y})}{\partial w_i}$$

$$= \frac{\partial(\cancel{y^T y} - y^T\hat{y} - \hat{y}^T y + \hat{y}^T\hat{y})}{\partial w_i} = \frac{-\partial\left(y^T(w_i x_i) + (w_i x_i)^T y - (w_i x_i)^T w_i x_i\right)}{\partial w_i} = 0$$

تابعیت از $w_i$ ندارد.

$$\rightarrow \frac{\partial(w_i y^T x_i)}{\partial w_i} + \frac{\partial(w_i x_i^T y)}{\partial w_i} - \frac{\partial(w_i^2 x_i^T x_i)}{\partial w_i} = 0$$

$$y^T x_i + x_i^T y - \left[2 w_i x_i^T x_i\right] = 0$$

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

---

ب)

$$(\mathrm{I}) \rightarrow \hat{y}_{n\times 1} = X w$$

$$\rightarrow \frac{\partial \|\hat{y}-y\|^2}{\partial w} = \frac{\partial (Xw-y)^T(Xw-y)}{\partial w}$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

$w_{m\times 1} \quad X_{n\times m} \quad y_{n\times 1}$

$$= \frac{\partial (w^T X^T - y^T)(Xw - y)}{\partial w} = \frac{\partial \left[ \overbrace{w^T X^T X w - w^T X^T y - y^T X w}^{A} - \cancel{y^T y} \right]}{\partial w}$$

$x_i^T x_j = 0 \quad (i \neq j) \leftarrow$ ستون های ماتریس X متعامدند

$$X^T X = \begin{bmatrix} - x_1^T - \\ - x_2^T - \\ \vdots \\ - x_n^T - \end{bmatrix} \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & & & 0 \\ & x_2^T x_2 & & \\ & & \ddots & \\ 0 & & & x_n^T x_n \end{bmatrix}$$

ستون j ام ماتریس X

$$\Rightarrow A = \sum_{j=1}^{m} w_j x_j^T x_j w_j - \sum_{j=1}^{m} \sum_{i=1}^{n} w_j x_{ij} y_i - \sum_{j=1}^{n} \sum_{i=1}^{m} y_j x_{ji} w_i$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} w_j x_{ij} x_{ij} w_j - \sum_{j=1}^{m} \sum_{i=1}^{n} w_j x_{ij} y_i - \sum_{j=1}^{n} \sum_{i=1}^{m} y_j x_{ji} w_i$$

$$\Rightarrow \frac{\partial A}{\partial w} = \begin{bmatrix} \frac{\partial A}{\partial w_1} \\ \frac{\partial A}{\partial w_2} \\ \vdots \\ \frac{\partial A}{\partial w_m} \end{bmatrix} = \underline{0}$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

$$\frac{\partial A}{\partial w_1} = 2w_1\sum_{j=1}^{n} x_{j1}^2 - \sum_{j=1}^{n} x_{j1}y_j - \sum_{i=1}^{n} y_i x_{i1} = 2w_1\sum_{j=1}^{n} x_{j1}^2 \quad 2\sum_{j=1}^{n} y_j x_{j1} = 0$$

$$w_1 = \frac{\sum_{j=1}^{n} y_j x_{j1}}{\sum_{j=1}^{n} x_{j1}^2} = \frac{y^T x_1}{x_1^T x_1}$$

مشابها : $\frac{\partial A}{\partial w_i} = 0 \rightarrow w_i = \frac{y^T x_i}{x_i^T x_i}$ که برابر همان موردی است که در بخش قبل یافتیم.

ج) $\hat{y} = w_0 \underbrace{1_{n\times 1}}_{a} + w_i x_i$

$$\rightarrow \frac{\partial \|\hat{y}-y\|^2}{\partial w_i} = \frac{\partial (\hat{y}-y)^T(\hat{y}-y)}{\partial w_i} = \frac{\partial(\hat{y}^T\hat{y} - \hat{y}^T y - y^T\hat{y} + \cancel{y^T y})}{\partial w_i}$$

$$= \frac{\partial\left[(w_0 a^T + w_i x_i^T)(w_0 a + w_i x_i) - 2y^T(w_0 a + w_i x_i)\right]}{\partial w_i}$$

$$= \frac{\partial\left[w_0^2 a^T a + w_0 w_i a^T x_i + w_0 w_i x_i^T a + w_i^2 x_i^T x_i - 2w_0 a^T y - 2w_i x_i^T y\right]}{\partial w_i} = 0$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

$$\Rightarrow w_0 a^T x_i + w_0 x_i^T a + 2 w_i x_i^T x_i - 2 x_i^T y = 0$$

$$w_i = \frac{x_i^T y - w_0 x_i^T a}{x_i^T x_i} \quad (1)$$

$$\frac{\partial \|\hat{y} - y\|^2}{\partial w_0} = 0 \to 2 w_0 a^T a + 2 w_i a^T x_i - 2 a^T y = 0$$

$$\to w_i = \frac{a^T y - w_0 a^T a}{a^T x_i} \quad (2)$$

المان j ام بردار $y$

المان j ام بردار $x_i$

$$\Rightarrow \frac{a^T y - w_0 a^T a}{a^T x_i} = \frac{x_i^T y - w_0 x_i^T a}{x_i^T x_i} \Rightarrow \frac{\sum y_j - n w_0}{\sum x_j} = \frac{\sum x_j y_j - w_0 \sum x_j}{\sum x_j^2}$$

$$\Rightarrow \sum y_j \sum x_j^2 - n w_0 \sum x_j^2 = \sum x_j \sum x_j y_j - w_0 \sum x_j \sum x_j$$

$$w_0 = \frac{\sum x_j \sum x_j y_j - \sum y_j \sum x_j^2}{\sum x_j \sum x_j - n \sum x_j^2}$$

$$w_i = \frac{\sum y_j - \dfrac{n \sum x_j \sum x_j y_j - n \sum y_j \sum x_j^2}{\sum x_j \sum x_j - n \sum x_j^2}}{\sum x_j} =$$

$$w_i = \frac{(\sum y_j)(\sum x_j)^2 - n\cancel{(\sum y_j)}(\sum x_j^2) - n(\sum x_j)(\sum x_j y_j) + n\cancel{(\sum y_j)}(\sum x_j^2)}{(\sum x_j)^3 - n(\sum x_j)(\sum x_j^2)}$$

$$= \frac{(\sum y_j)(\sum x_j) - n(\sum x_j y_j)}{(\sum x_j)^2 - n(\sum x_j^2)} = \frac{\dfrac{(\sum x_j y_j)}{n} - \dfrac{(\sum x_j)}{n} \times \dfrac{(\sum y_j)}{n}}{\dfrac{(\sum x_j^2)}{n} - \left(\dfrac{\sum x_j}{n}\right)^2}$$

$$= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \simeq \frac{E[xy] - E[x]E[y]}{E[x^2] - E[x]^2} = \frac{cov(x,y)}{var(x)}$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

$$w_0 = \frac{\sum x_j \sum x_j y_j - \sum y_j \sum x_j^2}{(\sum x_j)^2 - n \sum x_j^2}$$

$$E[y] - w_1 E[x] = \frac{\sum y_j}{n} - w_1 \frac{\sum x_j}{n} = \frac{\sum y_j}{n} - \frac{(\sum x_j)^2 \sum y_j - n(\sum x_j y_j)\sum x_j}{n(\sum x_j)^2 - n^2 \sum x_j^2}$$

$$= \frac{\cancel{(\sum x_j)^2 \sum y_j} - n\sum y_j \sum x_j^2 - \cancel{(\sum x_j)^2 \sum y_j} + n(\sum x_j y_j)\sum x_j}{n(\sum x_j)^2 - n^2(\sum x_j^2)}$$

$$= \frac{\sum x_j \sum x_j y_j - \sum y_j \sum x_j^2}{(\sum x_j)^2 - n\sum x_j^2} = w_0 \quad \checkmark$$

Q5) الف)

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x)\, dx = \underbrace{\int_{-\infty}^{a} x f(x)\, dx}_{\geq 0} + \int_{a}^{+\infty} x f(x)\, dx$$

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

$$\rightarrow \mathbb{E}[X] \geq a P(X \geq a) \rightarrow \frac{\mathbb{E}[X]}{a} \geq P(X \geq a)$$

ب)

$$Z : \begin{cases} \mathbb{E}[Z] = \mu \\ Var(Z) = \sigma^2 \end{cases}$$

let $Y := (Z - \mu)^2$

markov's inequality $\rightarrow P(Y \geq a^2) \leq \frac{\mathbb{E}[Y]}{a^2}$

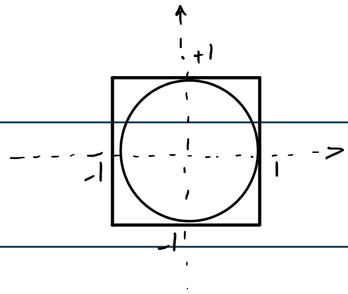
$$\rightarrow P(|Z - \mu|^2 \geq a^2) \leq \frac{\mathbb{E}[(Z-\mu)^2]}{a^2} = \frac{Var(Z)}{a^2}$$

$$Y \geq a^2 \equiv |Z - \mu| \geq a \rightarrow P(|Z - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

ج)



با استفاده از نامساوی چبیشف می‌خواهیم عدد $\pi$ را تخمین بزنیم.

تعریف می‌کنیم:

$$I(\bar{X}) = \begin{cases} 1 & x_1^2+x_2^2 < 1 \\ 0 & \text{O.W.} \end{cases} \quad \& \quad U(\bar{X}) := \text{Uniform}(-1 \leq x_1, x_2 \leq 1)$$

$$S_n = \frac{\sum_{i=1}^{n} I(\bar{X}_i)}{n} \rightarrow S'_n = 4S_n$$

$$\rightarrow \Pr\left[\frac{|S'_n - \pi|}{\pi} \leq 0.01\right] \geq 0.95 \Rightarrow \Pr\left[|S'_n - \pi| > 0.01\pi\right] < 0.05 \quad (1)$$

$$\mathbb{E}[S'_n] = \frac{4}{n} E\left[\sum_{i=1}^{n} I(\bar{X}_i)\right] = \frac{4}{n} \sum_{i=1}^{n} \mathbb{E}[I(\bar{X}_i)] = 4\,\mathbb{E}[I(\bar{X})]$$

$$\mathbb{E}[I(\bar{X})] = \int_{\bar{x}} I(\bar{x})\,U(\bar{x})\,d\bar{x} = \int_{x_1^2+x_2^2 \leq 1} \frac{1}{4}\,dx_1\,dx_2 = \frac{\pi}{4} \checkmark$$

$$\rightarrow \mathbb{E}[S'_n] = \pi$$

$$\mathbb{E}[I^2(\bar{X})] = \int_{\bar{x}} I^2(\bar{x})\,U(\bar{x})\,d\bar{x} = \int_{x_1^2+x_2^2 \leq 1} \frac{1}{4}\,dx_1\,dx_2 = \frac{\pi}{4}$$

$$\text{Var}(S'_n) = \text{Var}\left(\frac{4}{n}\sum_{i=1}^{n} I(\bar{X}_i)\right) = \frac{16}{n^2}\text{Var}\left(\sum I(\bar{X}_i)\right) \underset{\text{استقلال}}{=} \frac{16}{n}\text{Var}(I(\bar{X}))$$

$$= \frac{16}{n}\left(E[I^2(X)] - E[I(X)]^2\right) = \frac{16}{n}\left(\frac{\pi}{4} - \frac{\pi^2}{16}\right) = \frac{1}{n}(4\pi - \pi^2) \quad (2)$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

①, ② →

$$\frac{1}{n}(4\pi - \pi^2) \leq 0.05$$

$$\rightarrow n \geq 20 \times (4\pi - \pi^2)$$

$$\rightarrow n \geq 53.93$$

$$\rightarrow \boxed{n = 54}$$

---

Q7) sigmoid function : $\sigma(x) = \dfrac{1}{1+e^{-x}}$

$$\mathrm{Tanh}(x) = \frac{e^{x} - e^{-x}}{e^{x}+e^{-x}} = \frac{1-e^{-2x}}{1+e^{-2x}} = -1 + \frac{2}{1+e^{-2x}} = -1 + 2\sigma(2x)$$

$$\rightarrow \mathrm{Tanh}(x) = -1 + 2\sigma(2x) \rightarrow \sigma(x) = \frac{1+\mathrm{Tanh}(x/2)}{2} \quad ①$$

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

$$= \underbrace{w_0 + \frac{1}{2}\sum_{j=1}^{n} w_j}_{u_0} + \sum_{j=1}^{n} \underbrace{\frac{w_j}{2}}_{u_j} \mathrm{Tanh}\left(\frac{x-\mu_j}{s}\right) = u_0 + \sum_{j=1}^{n}\left[u_j \,\mathrm{Tanh}\left(\frac{x-\mu_j}{s}\right)\right]$$

constant

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

Q6) SVD decomposition: $A = U \Sigma V^T$ ۳

where: $U$ & $V$ are unitary matrix $\rightarrow \begin{cases} U^T U = V^T V = I \\ \rightarrow V^{-1} = V^T \ \& \ U^{-1} = U^T \end{cases}$

$$\Sigma_{n\times n} = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \ \sigma_n \end{bmatrix} \qquad \rightarrow AA^T u_i = \sigma_i^2 u_i \quad ۲$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$

اثبات می‌کنیم که: $\sigma_i \neq 0$

ماتریس $A$ معکوس‌پذیر است $\rightarrow$ ستون‌های $A$ مستقل خطی اند $\rightarrow N(A) = Z$

$$\Rightarrow Ax = 0 \rightarrow x = 0$$

واضحاً چون $Rank(A) = Rank(A^T) = n$ داریم:

$N(A^T) = Z \Rightarrow A^T x = 0 \rightarrow x = 0$ (1)

حال نشان می‌دهیم که: $N(AA^T) = Z$

$$AA^T x = 0 \rightarrow x^T A A^T x = 0 \rightarrow (A^T x)^T (A^T x) = 0 \rightarrow \|A^T x\|_2^2 = 0$$

$$\xrightarrow{(1)} x = 0$$

پس داریم: $N(AA^T) = Z$

و از رابطه‌ی ۲ می‌یابیم که به‌ازای $u_i \neq 0$ ، $\sigma_i^2$ مخالف صفر باشد.

$$\Rightarrow \begin{cases} \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \\ \sigma_i \neq 0 \end{cases} \Rightarrow \det(\Sigma) = \prod_{i=1}^{n} \sigma_i \neq 0$$

$$\rightarrow \Sigma \text{ is invertible} \rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{bmatrix}$$

$$\xrightarrow{۳} A^{-1} = (V^T)^{-1} \Sigma^{-1} U^{-1} = V \Sigma^{-1} U^T$$

مقادیر تکین $A$: $\sigma_1, \sigma_2, \dots, \sigma_n > 0$

مقادیر تکین $A^{-1}$: $\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n} \rightarrow \sigma_{max}(A^{-1}) = \frac{1}{\sigma_{min}(A)}$

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

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

ب) $A_{m\times n}$ $\quad \|A\|_2 \leq \|A\|_F \leq \sqrt{\text{rank}(A)}\ \|A\|_2$

$$\Rightarrow \|A\|_2^2 \leq \|A\|_F^2 \leq \text{rank}(A)\ \|A\|_2^2$$

میدانیم: $\|A\|_2^2 = \lambda_{max}(A^HA) = \sigma_{max}^2$ $\qquad$ ($\sigma$ : singular value of $A$)

$$\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2 = \text{trace}(A^HA)$$

برای ماتریس مربعی X میدانیم:

$$\text{trace}(X) = \sum_{i=1}^{n} \lambda_i$$

$$\rightarrow \text{trace}(A^HA) = \sum_{i=1}^{n} \lambda_i(A^HA) = \sum_{i=1}^{n} \sigma_i^2 \overset{①}{=} \sum_{i=1}^{R} \sigma_i^2$$

همچنین داریم:

$$\sigma_{max}^2 \leq \sum_{i=1}^{R} \sigma_i^2 \leq R\,\sigma_{max}^2$$

$$\Rightarrow \|A\|_2^2 \leq \|A\|_F^2 \leq \text{Rank}(A)\ \|A\|_2^2$$

$$\Rightarrow \|A\|_2 \leq \|A\|_F \leq \sqrt{\text{Rank}(A)}\ \|A\|_2$$

Proof of ①:

میدانیم: eigen values of $A^HA$ $= \sigma_i^2$

حال نشان می‌دهیم که تعداد مقادیر ویژه‌ی یک ماتریس دلخواه $A^{n\times n}$، برابر Rank آن ماتریس می‌باشد:

$$A x_i = \lambda_i x_i \xrightarrow{\text{if } \lambda_i = 0} A x_i = 0 \qquad (x_i \neq 0)$$

بنابراین تعداد λ های صفر ماتریس برابر است با:

$$\#(\lambda_i = 0) = \dim(N(A))$$

و از قضیه‌ی شمارش می‌دانیم:

$$\dim(N(A)) + \dim(C(A)) = n$$

$$\rightarrow \dim(N(A)) = n - \text{Rank}(A) = n - R$$

# HW1-400100662

Thursday, October 12, 2023 3:33 PM

پس برای ماتریس $A^{n\times n}$ با رنک $R$ ، $n-R$ مقدار ویژه ی صفر و $R$

مقدار ویژه ی ناصفر وجود دارد.

حال اگر نشان دهیم ، $Rank(A^HA) = Rank(A) = R$ ، می توانیم بگوییم

$A^HA$ ، $R$ مقدار ویژه ی ناصفر دارد که معادل این است که $A$ ،

$R$ مقدار تکین ناصفر دارد.

$$Rank(A) = Rank(A^H) = dim(C(A^H)) = R$$

$$\text{if } x \in N(A) : Ax = 0 \rightarrow A^HAx = 0 \rightarrow x \in N(A^HA)$$

وبرعکس : $$\text{if } x \in N(A^HA) : A^HAx = 0 \rightarrow x^HA^HAx = 0 \rightarrow (Ax)^H(Ax) = 0$$

$$\rightarrow \|Ax\|_2^2 = 0 \rightarrow Ax = 0 \rightarrow x \in N(A)$$

بنابراین : $N(A) = N(A^HA)$ و از قضیه ی شکاف رتبه ای داریم :

$$n = dim(N(A)) + Rank(A)$$
$$\rightarrow N(A) = N(A^HA) \Rightarrow Rank(A) = Rank(A^HA) \quad //$$

و به همین دلیل است که در تجزیه ی SVD ماتریس $A$ داریم :

$$A = \sum_{i=1}^{R} \sigma_i u_i v_i^T$$