یادگیری عمیق

دکتر فاطمی زاده

دانشگاه صنعتی شریف

مهندسی برق

برنا خدابنده ۴۰۰۱۰۹۸۹۸

تمرین شماره 1

تاریخ: 1402/7/26

$$P(Y=i|X=x) = f_X(x|Y=i)\,\frac{P(Y=i)}{f_X(x)}$$ 1)

$$P(Y=i|X=x) = \frac{P(Y=i)f_X(x|Y=i)}{\sum_i P(Y=i)f_X(x|Y=i)}, \quad i^* = \underset{i}{\operatorname{argmax}}\, P(Y=i|X=x)$$

$$\forall i: P(Y=i) = 1/3 \Rightarrow P(Y=i|X=x) = \frac{f_X(x|Y=i)}{\sum_i f_X(x|Y=i)}$$

$$(X|Y=i) \sim \mathcal{N}(\mu_i, \Sigma_i) \Rightarrow f_X(x|Y=i) = \frac{1}{2\pi\sqrt{\det(\Sigma_i)}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right)$$

$$\det(\Sigma_1) = 0.49, \quad \Sigma_1^{-1} = \frac{1}{0.7}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \det(\Sigma_2) = 0.07, \quad \Sigma_2^{-1} = \frac{1}{0.7}\begin{pmatrix} 2 & -3 \\ -3 & 8 \end{pmatrix}$$

$$\det(\Sigma_3) = 0.52, \quad \Sigma_3^{-1} = \frac{5}{13}\begin{pmatrix} 4 & -1 \\ -1 & 3.5 \end{pmatrix}$$

sharing denominator in $P(Y=i|X=x)$ $\Rightarrow$ $i^* = \underset{i}{\operatorname{argmax}}\, f_X(x|Y=i)$

الف)

$$f_X(x|Y=1) \propto e^{-875}, \quad f_X(x|Y=2) \propto e^{-953}, \quad f_X(x|Y=3) \propto e^{-835}$$

$$\Rightarrow \boxed{i^* = 3}$$

$$f_X(x|Y=1) = 0.2727, \quad f_X(x|Y=2) = 1.8615, \quad f_X(x|Y=3) = 0.2419$$ ب)

$$\Rightarrow P(Y=1|X=x) = 0.1146, \quad P(Y=2|X=x) = 0.7836, \quad P(Y=3|X=x) = 0.1016$$

$$\Rightarrow \boxed{i^* = 2}$$

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

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2 I), \quad x_i \to x_i + \epsilon_i$$

$$\tilde{E}_D(w) = \frac{1}{2}\sum_{i=1}^{N}[\tilde{y}_n - y_n]^2, \quad \tilde{y}_n = y(x_n + \epsilon_n, w) = w_0 + \sum_{i=1}^{D} w_i(x_{ni} + \epsilon_{ni}) = \hat{y}_n + \delta_n$$

$$E(\tilde{E}_D(w)) = \frac{1}{2}\sum_{j=1}^{N} E([\tilde{y}_n - y_n]^2) = \frac{1}{2}\sum_{i=1} E(\underbrace{[(\hat{y}_n - y_n) + \delta_n]^2}_{(\hat{y}_n - y_n)^2 + 2\delta_n[\hat{y}_n - y_n] + \delta_n^2})$$

$$= E_D(w) + E(\delta_n)\sum_{i=1}(\hat{y}_n - y_n) + \frac{N}{2}E(\delta_n^2)$$

$$\delta_n = \sum_{i=1}^{D} w_i \epsilon_{ni} \Rightarrow E(\delta_n) = \sum_{i=1}^{D} w_i E(\epsilon_{ni}) = 0$$

$$\delta_n^2 = \left(\sum_{i=1}^{D} w_i \epsilon_{ni}\right)^2 = \sum_{i=1}^{D} w_i^2 \epsilon_{ni}^2 + 2\sum_{i \neq j} w_i w_j \epsilon_{ni}\epsilon_{nj} \Rightarrow E(\delta_n^2) = \sum_{i=1}^{D} w_i^2 \underbrace{E(\epsilon_{ni}^2)}_{\sigma^2} + 2\sum_{i \neq j} w_i w_j \underbrace{E(\epsilon_{ni}\epsilon_{nj})}_{0}$$

$$\Rightarrow E(\delta_n^2) = \sigma^2 \sum_{i=1}^{D} w_i^2$$

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

influenced by cs188

logistic regression: $f(x) = x$ for simple regression.

$$P(y=1 \mid f(x), \theta) = \frac{1}{1+\exp(-\theta^T f(x))}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n+1} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$$

mult_class:

$$P(y=i \mid X=x, \theta) = \frac{e^{w_i^T f(x)}}{\sum_{k=1}^{K-1} e^{w_k^T f(x)}}, \quad W = [w_1, w_2, \dots, w_{K-1}]$$

$$l(W) = \prod_{i=1}^{n} P(y_i \mid W, f(x)) = \prod_{i=1}^{n} \prod_{k=1}^{K-1} \left(\frac{e^{w_k^T f(x)}}{\sum_{k=1}^{K-1} e^{w_k^T f(x)}} \right)^{\delta_{k,y_i}} ; \quad \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & \text{o.w} \end{cases}$$

$$L = \log(l(W)) = \boxed{\sum_{i=1}^{n} \sum_{k=1}^{K-1} \delta_{k,y_i} \log\left(\frac{e^{w_k^T f(x)}}{\sum_{k=1}^{K-1} e^{w_k^T f(x)}} \right) = \sum_{i=1}^{n} \sum_{k=1}^{K-1} \delta_{k,y_i} \left[w_k^T f(x) - \log\left(\sum_{k=1}^{K-1} e^{w_k^T f(x)} \right) \right]}$$

$$\frac{\partial L}{\partial w_k} = \sum_{i=1}^{n} \frac{\partial}{\partial w_k} \sum_{j=1}^{K-1} \delta_{j,y_i} \left[w_k^T f(x) - \log\left(\sum_{j=1}^{K-1} e^{w_j^T f(x)} \right) \right]$$

$$= \sum_{i=1}^{n} \left[\delta_{k,y_i} f(x) - \sum_{j=1}^{k-1} \delta_{k,y_i} \frac{e^{w_k^T f(x)}}{\sum_{j=1}^{k-1} e^{w_j^T f(x)}} f(x) \right] = \boxed{\sum_{i=1}^{n} \left(\delta_{k,y_i} - \frac{e^{w_k^T f(x)}}{\sum_{j=1}^{K-1} e^{w_j^T f(x)}} \right) f(x) = \nabla_{w_k} L}$$

$$\nabla_{w_k} f = \nabla_{w_k} L - \nabla_{w_k} \left[\frac{\lambda}{2} \sum_{j=1}^{K-1} w_j^T w_j \right] = \boxed{\nabla_{w_k} L - \lambda w_k = \nabla_{w_k} f}$$

(4

(الف

regression problem:

minimize $\|Xw - y\|_2^2$, (assuming Z is extended thes W contians biases)

on a single feature: $\|w_j x_j - y\|^2 \rightsquigarrow$ best w_j is the projection

$$\Rightarrow w_j x_j = y + e_\perp \Rightarrow w_j x_j^T x_j = x_j^T y + \cancel{x_j^T e} \Rightarrow \boxed{w_j = \frac{x_j^T y}{x_j^T x_j}}$$

(ب

let: $\forall i \neq j : x_i^T x_j = 0$, minimize $\|Xw - y\|_2^2 \Rightarrow Xw = y + e , e \in N(X)$

$$\Rightarrow \sum_i w_i x_i = y + e_\perp \Rightarrow x_j^T \sum w_i x_i = x_j^T y + \cancel{x_j^T e} = \sum_i w_i x_j^T x_i$$

$$= (0 + 0 + 0 \dots + w_j x_j^T x_j + 0 + 0 \dots + 0) = x_j^T y \Rightarrow \boxed{w_j = \frac{x_j^T y}{x_j^T x_j}}$$

جواب همان جواب قسمت الف است در نتیجه این مسئله حالت خاص با ویژگی مستقل است

(ج

$$Z = [X ; \mathbb{1}_{n\times 1}] , \quad w = \begin{bmatrix} \vec{w} \\ \hline w_0 \end{bmatrix} \Rightarrow \text{minimize } \|Zw - y\|_2^2 \Rightarrow w_{ls} = Z^\dagger y$$

$w \to \begin{bmatrix} w_j \\ \hline w_0 \end{bmatrix} = \omega^T$

$w_{ls} = (Z^T Z)^{-1} Z^T y$, on a single feature $\Rightarrow Z \to [x_j ; \mathbb{1}_{n\times 1}] = Z_j$

$$\Rightarrow \omega_{j,ls}^T = \left(\begin{bmatrix} x_j^T \\ \hline \mathbb{1}^T \end{bmatrix} [x_j ; \mathbb{1}] \right)^{-1} \begin{bmatrix} x_j^T \\ \hline \mathbb{1}^T \end{bmatrix} y = \begin{pmatrix} \overset{nE(x^2)}{x_j^T x_j} & \overset{nE(x)}{x_j^T \mathbb{1}} \\ \mathbb{1}^T x_j & n \end{pmatrix}^{-1} \begin{bmatrix} x_j^T \\ \hline \mathbb{1}^T \end{bmatrix} y$$

$$= \frac{1}{\cancel{n}} \cdot \frac{1}{\underbrace{E(x^2) - E(x)^2}_{Var(x)}} \begin{pmatrix} 1 & -E(x) \\ -E(x) & E(x^2) \end{pmatrix} \begin{bmatrix} \overset{\cancel{n}E(xy)}{x_j^T y} \\ \underset{\cancel{n}E(y)}{\mathbb{1}^T y} \end{bmatrix} = \frac{1}{Var(x)} \begin{bmatrix} \overbrace{E(xy) - E(x)E(y)}^{Cov(x,y)} \\ \underbrace{E(x^2)E(y) - E(x)E(xy)} \end{bmatrix} = \begin{bmatrix} w_j \\ w_0 \end{bmatrix}$$

$$\Rightarrow \boxed{w_j = \frac{Cov(x_j, y)}{Var(x_j)}} , \quad w_0 = \frac{1}{Var(x)} \Big[E(y) \underbrace{\{E(x^2) - E^2(x)\}}_{Var(x)} + \cancel{E^2(x)E(y)} - E(x) \underbrace{\{E(xy) - E(x)E(y)\}}_{Cov(x,y)} - \cancel{E^2(x)E(y)} \Big]$$

$$\hookrightarrow w_0 = E(y) - \frac{Cov(x_j, y)}{Var(x)} E(x) = \boxed{E(y) - w_j E(x)}$$

Influenced by ProbStat - Arash Amini

5) الف)

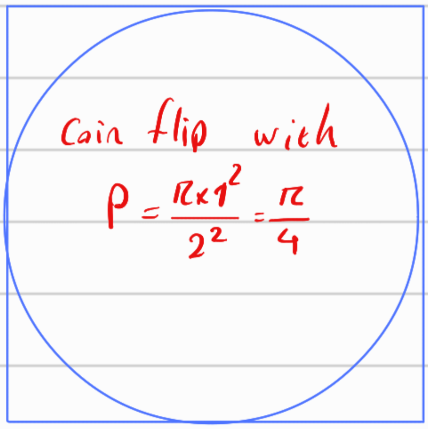
$$Y = \begin{cases} \alpha & X \geq \alpha \\ 0 & X < \alpha \end{cases} \Rightarrow \forall X : Y \leq X \Rightarrow E(Y) \leq E(X)$$

$$\Rightarrow E(Y) = \alpha P(Y = \alpha) = \boxed{\alpha P(X \geq \alpha) \leq E(X)}$$

ب)

$$P(|Z-\mu| \geq \varepsilon) = P(|Z-\mu|^2 \geq \varepsilon^2) \leq \frac{E((Z-\mu)^2)}{\varepsilon^2} = \frac{Var(Z)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

ج)



$$Z = 4\frac{N_{in}}{N}, \quad N_{in} = X_1 + X_2 + \cdots + X_n, \quad \pi = E(Z)$$

$$X_i : \begin{cases} 1 : \text{داخل دایره} \ (p = \pi/4) \\ 0 : \text{خارج دایره} \ (p = 1 - \pi/4) \end{cases}$$

$$Var(Z) = \frac{16}{N^2} Var(N_{in}) = \frac{16}{N^2} \times \sum_i Var(X_i) = \frac{16}{N} Var(X)$$

$$Var(X) = E(X^2) - E^2(X) = \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 = Var(X) \Rightarrow \boxed{Var(Z) = \frac{\pi}{N}(4-\pi)}$$

we need $P(|Z-\pi| \geq \varepsilon) \leq 5\%$, $\varepsilon = 2\% \times \pi = \eta\pi$

$$\Rightarrow 5\% \leq \frac{Var(Z)}{\varepsilon^2} = \frac{1}{N\eta^2}\left(\frac{4}{\pi} - 1\right) \geq 5\% \longrightarrow \boxed{\begin{array}{c} N \geq \frac{1}{\eta^2 \times 5\%}\left(\frac{4}{\pi} - 1\right) \\ N \geq 55000 \end{array}}$$

---

7)

$$y(x,w) = w_0 + \sum_j w_j \sigma\left(2 \overbrace{\frac{x-\mu_j}{s}}^{z_j}\right) = w_0 + \sum_j w_j \frac{1}{1+\exp(-2z_j)}$$

$$= w_0 + \sum_j w_j \frac{e^{z_j}}{e^{z_j} - e^{-z_j}} = w_0 + \sum_j w_j \times \frac{1}{2}\left\{1 + \frac{e^{z_j} + e^{-z_j}}{e^{z_j} - e^{-z_j}}\right\}$$

$$= w_0 + \sum_j \frac{w_j}{2}\left(1 + \tanh(z_j)\right)$$

$$= \underbrace{\left[w_0 + \sum_j \frac{w_j}{2}\right]}_{u_0} + \sum_j \underbrace{\frac{w_j}{2}}_{u_j} \tanh\left(\frac{x-\mu_j}{s}\right) = u_0 + \sum_j u_j \tanh\left(\frac{x-\mu_j}{s}\right)$$

$$\boxed{u_j = \frac{1}{2}w_j, \quad u_0 = w_0 + \sum_j \frac{1}{2} w_j}$$

6)

$$\Sigma^{-1} = \begin{pmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & \\ & & \ddots & \\ & & & 1/\sigma_n \end{pmatrix} \Rightarrow$$ مقادیر تکین A^{-1}، معکوس مقادیر A است.

a)

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

b)

$$\|A\| = \sigma_{max}(A) \Rightarrow \|A\|^2 = \sigma_{max}^2(A) \leq \sum_{i=1}^{rank(A)} \sigma_i^2(A) \overset{\mathrm{I}}{=} \|A\|_F^2$$

$$\|A_F\|^2 = \sum_{i=1}^{rank(A)} \sigma_i^2(A) \leq \sum_{i=1}^{rank(A)} \sigma_{max}^2(A) = rank(A)\,\sigma_{max}^2(A) = rank(A)\|A\|^2$$

$$\boxed{\begin{aligned} &\|A\|^2 \leq \|A\|_F^2 \leq rank(A)\|A\|^2 \\ &\|A\| \leq \|A\|_F \leq \sqrt{rank(A)}\,\|A\| \end{aligned}}$$

اثبات I :

$$\|A\|_F^2 = \sum_{i,j} a_{ij}^2 = Tr(A^T A)$$

$$A = U\Sigma V^T \Rightarrow A^T = V\Sigma^T U^T \Rightarrow A^T A = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T$$

$$= V\Sigma^2 V^T \Rightarrow \|A\|_F^2 = Tr(A^T A) = Tr(V\Sigma^2 V^T) = Tr(\Sigma^2 V^T V)$$

$$= Tr(\Sigma^2) = \sum_{i=1}^{rank(A)} \sigma_i^2 = \|A\|_F^2$$