تمرین سری ۱ یادگیری عمیق — 400101034 - حسام حسینی

سوال ۱-

$P(Y=1) = P(Y=2) = P(Y=3) = \frac{1}{3} \rightarrow$ Class Prior Probability

$f_{X|Y=i} = \mathcal{N}(\mu_i, \Sigma_i) \quad i \in \{1, 2, 3\}$

$|\Sigma_1| = 0.49 \qquad |\Sigma_2| = 0.07 \qquad |\Sigma_3| = 0.52$

$$\Sigma_1^{-1} = \begin{bmatrix} \frac{10}{7} & 0 \\ 0 & \frac{10}{7} \end{bmatrix} \qquad \Sigma_2^{-1} = \begin{bmatrix} \frac{20}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{80}{7} \end{bmatrix} \qquad \Sigma_3^{-1} = \begin{bmatrix} \frac{20}{13} & \frac{-5}{13} \\ \frac{-5}{13} & \frac{35}{26} \end{bmatrix}$$

$\hat{y} = \arg\max \mathbb{P}(Y=i \mid x) = \arg\max \mathbb{P}(x \mid Y=i)\,\mathbb{P}(Y=i)$

$X = [50, 0.5] \longrightarrow$

$\mathbb{P}(X \mid Y=1)\,\mathbb{P}(Y=1) \propto e^{-875} \qquad \mathbb{P}(X \mid Y=2)\,\mathbb{P}(Y=2) \propto e^{-953}$

$\mathbb{P}(X \mid Y=3) \propto e^{-835} \quad \Rightarrow \quad \hat{y} = 3$

$X = [0.5, 0.5] \Rightarrow$

$\mathbb{P}(X \mid Y=1) \times \mathbb{P}(Y=1) \simeq 0.12 \quad , \quad \mathbb{P}(X \mid Y=2)\,\mathbb{P}(Y=2) \simeq 0.78$

$\mathbb{P}(X \mid Y=3) \times \mathbb{P}(Y=3) \simeq 0.1 \quad \Rightarrow \quad \hat{y} = 2$

$$y(x_n, w) = w_0 + \sum_{i=1}^{D} w_i x_{ni} \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)$$ سوال ۲-

$$\tilde{x}_i = x_i + \epsilon_i \Rightarrow y(\tilde{x}_n, w) = w_0 + \sum_{i=1}^{D} w_i (x_i + \epsilon_i) = y(x_n, w) + \sum_{i=1}^{D} w_i \epsilon_i$$

$$\mathbb{E}[\tilde{E}_D(w)] = \mathbb{E}\left[\frac{1}{2}\sum_{n=1}^{N}\left[\underbrace{y(x_n, w)}_{a} + \underbrace{\sum_{i=1}^{D} w_i \epsilon_i}_{b} - \underbrace{y_n}_{c}\right]^2\right]$$

$$= \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}\left[a^2 + b^2 + c^2 + 2ab - 2ac - 2bc\right] \qquad E(b) = \sum_{n=1}^{D} w_i \cancel{E(\epsilon_i)} = 0$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left(a^2 + c^2 - 2ac + E(b^2)\right)$$

$$E(b^2) = E\left(\sum_{i=1}^{D} w_i^2 \epsilon_i^2 + \sum_{i=1}^{D}\sum_{i=1}^{D} w_i w_j \epsilon_i \epsilon_j\right) = \sum_{i=1}^{D} w_i^2 \mathbb{E}(\epsilon_i^2) = \sigma^2 \sum_{i=1}^{D} w_i^2$$

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

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

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

سوال ۳- الف) در Logistic Regression اصلی ما تنها دو کلاس داریم و با عبور $w^T x$ از تابع سیگموید احتمال اینکه در کلاس مورد نظر باشیم بدست می‌آید

برای حالت چند کلاسه می‌توان دو راهکار را مد نظر گرفت اگر بخواهیم احتمال اینکه به کلاس k تعلق داشته باشیم را حساب کنیم کلیه کلاس های دیگر را یک کلاس در نظر می‌گیریم و همان کاری که در Logistic Regression عادی انجام می‌دادیم را انجام می‌دهیم به این کار one-vs-Rest گویند

راهکار دیگر استفاده از تابع softmax است:

$$P(Y=i \mid x, w) = \frac{e^{w_i^T x}}{\sum_{j=1}^{k} e^{w_j^T x}}$$

برای اینکه $\sum_{i=1}^{k} P(Y=i \mid x,w) = 1$ باشد احتمال کلاس آخر را برابر $1 - \sum_{i=1}^{k-1} P(Y=i \mid x,w)$ در نظر می‌گیریم و w_k را learn نمی‌کنیم و $w_k = 0$ تنظیم می‌کنیم $\Leftarrow$ $e^{w_k^T x} = 1$

ب)

$$L(w_1, \dots, w_{k-1}) = \sum_{i=1}^{n} \ln\left(P(Y=i \mid X=x_i)\right) = \sum_{i=1}^{n} \ln\left(\frac{e^{w_i^T x}}{\sum_{j=1}^{n} e^{w_j^T x_i}}\right)$$

$$= \sum_{i=1}^{n} \left(w_{y_i}^T x_i - \underbrace{\ln\left(\sum_{j=1}^{k} e^{w_{y_j}^T x_i}\right)}\right)$$

جمله آخر $= 0$

پ)

$$\frac{\partial L}{\partial w_t} = \underbrace{\frac{\partial}{\partial w_t} \sum_{i=1}^{n} \left(w_{y_i}^T x_i\right)} + \frac{\partial}{\partial w_t} \ln\left(\sum_{j=1}^{k-1} e^{w_j^T x_i}\right)$$

$$\sum_{i=1}^{n} [y_i = t]\, x_i$$

$$\frac{\partial}{\partial w_t} \ln\left(\sum_{i=1}^{k-1} e^{w_i x_i}\right) = \frac{x_i\, e^{w_t^T x_i}}{\sum_{j=1}^{k-1} e^{w_j^T x_i}}$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial w_t} = \sum_{i=1}^{n} \left(x_i [y_i = t] - \frac{x_i\, e^{w_t^T x_i}}{\sum_{j=1}^{k-1} e^{w_j^T x_i}} \right)}$$

ت) تنها کافی است گرادیان ترم دوم را پیدا کنیم ترم اول را بخش قبلی محاسبه کرده‌ایم

$$\frac{\partial}{\partial w_t} \frac{\lambda}{2} \sum_{j=1}^{k-1} |w_j|^2 = \frac{\partial}{\partial w_t} \frac{\lambda}{2} \sum_{j=1}^{k-1} w_j^T w_j = \lambda w_t$$

$$\frac{\partial f}{\partial w_t} = \sum_{i=1}^{n} \left(x_i [y_i = t] - \frac{x_i\, e^{w_t^T x_i}}{\sum_{j=1}^{k-1} e^{w_j x_i}} \right) - \lambda w_t$$

سوال 4 _ الف)

$$L(w) = \frac{1}{N}\sum_{n=1}^{N}\left(h(x_n) - y_n\right)^2 \quad ; MSE$$

$$X_0 = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix}$$

$$= \frac{1}{N}\|Xw - Y\|_2^2 = \frac{1}{N}(Xw - Y)^T(Xw - Y)$$

$$= \frac{1}{N}\left(Y^TXw - Y^TY - (Xw)^T(Xw) + (Xw)^TY\right)$$

$$Xw = \sum_{i=1}^{n} w_i x_i \quad (I) \qquad \frac{\partial L}{\partial w_j} = 0 \xrightarrow{(I)} Y^T x_j + x_j^T Y - 2 w_j x_j x_j^T = 0$$

$$\longrightarrow w_j = \frac{x_j^T Y}{x_j^T x_j}$$

حالت خاص: $w_0 = \frac{1^T Y}{1^T 1} \Longleftarrow$ mean (Y) وقتی هیچ ویژگی از دیتا نداریم بهترین

تخمین همان میانگین label هایمان است

ب)

$$\nabla L(w) = \frac{2}{N}X^T(Xw - Y) = 0 \longrightarrow X^TXw = X^TY$$

$$x_i^T x_j = 0 \quad i \neq j \Longrightarrow X^TX = \begin{bmatrix} x_1^T x_1 & & 0 \\ & \ddots & \\ 0 & & x_{m+1}^T x_{m+1} \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} x_0^T x_0 & & 0 \\ & \ddots & \\ 0 & & x_m^T x_m \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} x_0^T \\ \vdots \\ x_m^T \end{bmatrix} Y$$

$$\Longrightarrow x_j^T x_j w_j = x_j^T Y \Longrightarrow w_j = \frac{x_j^T Y}{x_j^T x_j}$$

همان چیزی است که در بخش قبل به آن رسیدیم

$$Xw_j = w_0 \overset{n\times 1}{\hat{1}} + w_j x_j \Longrightarrow \frac{\partial}{\partial w_j} \|Xw - y\|^2 = \frac{\partial}{\partial w_j} (Xw - y)^T (Xw - y)$$

$$= \frac{\partial}{\partial w_j}\left[(w_j 1^T + w_j x_j^T)(w_0 1 + w_j x_j) - 2y^T (w_0 1 + w_j x_j) \right]$$

$$= w_0 1^T x_j + w_0 x_j^T 1 + 2 w_j x_j^T x_j - 2 x_j y = 0$$

$$\rightarrow w_j = \frac{x_j^T y - w_0 x_j^T 1}{x_j^T x_j}$$

$$\frac{\partial \|Xw - Y\|_2}{\partial w_0} = 0 \Rightarrow 2w_0 1^T 1 + 2 w_j 1^T x_j - 2 1^T y = 0$$

$$\Rightarrow w_j = \frac{1^T y - w_0 1^T 1}{1^T x_j} = \frac{x_j^T y - w_0 x_j^T 1}{x_j^T x_j}$$

$$\Rightarrow \frac{\sum_i y_i - n w_0}{\sum_i x_{ji}} = \frac{\sum_i x_{ji} y_i - w_0 \sum_i x_{ji}}{\sum_i x_{ji}^2}$$

$$\Rightarrow \sum_i y_i \sum_i x_{ji}^2 - n w_0 \sum_i x_{ji}^2 = \sum_i x_{ji} \sum_i x_{ji} y_i - w_0 \sum_i x_{ji} \sum_i x_{ji}$$

$$\Rightarrow w_0 = \frac{\sum_i x_{ji} \sum_i x_{ji} y_i - \sum y_i \sum x_{ji}^2}{\sum x_{ji} \sum x_{ji} - n \sum x_{ji}^2}$$

$$w_j = \frac{\sum y_i - \dfrac{n \sum x_{ji} \sum x_{ji} y_i - n \sum y_i \sum x_{ji}^2}{\sum_i x_{ji} \sum_i x_{ji} - n \sum x_{ji}^2}}{\sum x_{ji}} = \frac{(\sum_i y_i)(\sum_i x_{ji}) - n(\sum_i x_{ji} y_i)}{(\sum_i x_{ji})^2 - n(\sum_i x_{ji}^2)}$$

$$= \frac{\overline{x_j y} - \overline{x_j}\,\overline{y}}{\overline{x_j^2} - \overline{x_j}^2} = \frac{E(x_j y) - E(x_j)E(y)}{E(x_j^2) - E(x_j)^2} = \frac{\mathrm{Cov}(x_j, y)}{\mathrm{Var}(x_j)}$$

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

$$E(y) - w_j E(x) = \frac{\sum y_i}{n} - \frac{w_j \sum_i x_{ji}}{n} =$$

$$\frac{\sum_i y_i}{n} - \frac{(\sum_i x_{ji})^2 \sum_i y_i - n(\sum x_{ji} y_i) \sum x_{ji}}{n(\sum_i x_{ji})^2 - n^2 \sum x_{ji}^2}$$

$$= \frac{\sum_i x_{ji} \sum x_{ji} y_i - \sum y_i \sum x_{ji}^2}{(\sum x_{ji})^2 - n \sum x_j^2} = w_0$$

سوال 5 - الف)

$$E(x) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\alpha} x f_X(x)\,dx + \int_{\alpha}^{\infty} x f_X(x)\,dx$$

با فرض $X > 0$

$$\geq \int_{\alpha}^{\infty} x f_X(x)\,dx \geq \alpha \int_{\alpha}^{\infty} f_X(x)\,dx = \alpha \mathbb{P}(X > \alpha)$$

$$\Rightarrow \frac{E(X)}{\alpha} \geq P(X > \alpha)$$

ب)

markov's $\quad (z-\mu)^2 > 0$

$$\mathbb{P}(|z-\mu| \geq \epsilon) = \mathbb{P}((z-\mu)^2 \geq \epsilon^2)$$

$$\leq \frac{E((z-\mu)^2)}{\epsilon^2} = \frac{Var(z)}{\epsilon^2} \Rightarrow \mathbb{P}(|z-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

ج) به احتمال $\frac{\pi}{4}$ نقطه تصادفی بر روی دایره فرود می‌آید $\Rightarrow$ $\frac{\pi}{4}$: مساحت دایره

$X_i \sim Bernouli(\frac{\pi}{4})$: 1 : داخل دایره   0 : خارج دایره

$M_n = \frac{4}{n} \sum_{i=1}^{n} x_i \rightarrow$ تخمین‌گر عدد π

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

$$Var(M_n) = \left(\frac{4}{n}\right)^2 \sum Var(x_i) = \frac{16}{n^2} \sum \frac{\pi}{4}\left(1-\frac{\pi}{4}\right) = \frac{\pi(4-\pi)}{n}$$

$$\mathbb{P}(|M_n - E(M_n)| \geq 0.01) \leq \frac{Var(M_n)}{0.01^2}$$

$$\Rightarrow \mathbb{P}(|M_n - \pi| \geq 0.01) \leq \frac{\pi(4-\pi)}{0.01^2 n} \leq 0.05 \Rightarrow n \geq 539354$$

نکته‌ای که کمی ایراد داشت این است که از عدد π استفاده کردیم درصورتیکه هدف ما تخمین π است.

برای این منظور از $x(4-x) \leq 4$ استفاده می‌کنیم که در حالت قبل بجای x، π قرار گرفته بود

$$\Rightarrow \mathbb{P}(|M_n - \pi| \geq 0.01) \leq \frac{4}{0.01^2 n} \leq 0.05 \Rightarrow n \geq 8 \times 10^5$$

سوال ۶- الف)

$$A = U \Sigma V^T \qquad \underset{\substack{n\times n \\ A \text{ مربعی}}}{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

$\sigma_1^2, \sigma_2^2, \dots \sigma_n^2$ are eigenvalus of AA^T and A^TA

$$U = \underset{\text{vector}}{\mathrm{eig}}(AA^T) \Rightarrow U^TU = I \Rightarrow U^{-1} = U^T \qquad u_i \neq 0$$

$$V = \mathrm{eig}(A^TA) \Rightarrow V^TV = I \Rightarrow V^{-1} = V^T \qquad v_i \neq 0$$

چون A مربعی و معکوس پذیر است

$\mathrm{Rank}(A) = \mathrm{Rank}(A^T) = n$

$N(A) = N(A^T) = Z$

$N(AA^T) = Z$

اثبات: $AA^Tx = 0 \Rightarrow x^TAA^Tx = 0$

$\Rightarrow \|A^Tx\|^2 = 0 \Rightarrow x = 0$

$$AA^Tu_i = \sigma_i^2 u_i \Rightarrow \text{for } u_i \neq 0 \Rightarrow \sigma_i \neq 0$$

$$|\Sigma| = \prod \sigma_i \neq 0 \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix}$$

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

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

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

ب)

$$\|A_2\|_2^2 = \sigma_{max}^2, \quad \|A\|_F^2 = tr(A^TA) = \sum_{i=1}^{r} \sigma_i^2 \qquad r = \mathrm{Rank}(A)$$

$$\Rightarrow \sigma_{max}^2 \leq \sum_{i=1}^{r} \sigma_i^2 \Rightarrow \sigma_{max} \leq \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

$$\Rightarrow \|A_2\|_2 \leq \|A\|_F$$

از طرفی

$$\sqrt{r}\,\underbrace{\sigma_{max}}_{\|A\|_2} \geq \sqrt{\sum_{i=1}^{r} \sigma_i^2} = \|A\|_F \Leftarrow r\,\sigma_{max}^2 \geq \sum_{i=1}^{r} \sigma_i^2$$

$$\Longrightarrow \|A_2\|_2 \leq \|A\|_F \leq \sqrt{r}\,\|A\|_2$$

سوال 7 -

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

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

$$\Rightarrow \sigma(x) = \frac{1+\tanh\left(\frac{x}{2}\right)}{2}$$

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

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