Mathematical Foundations of Deep Neural Networks. HW #1 2021-14424 이서준

Problem 1.

a) $l_i(\theta) = \frac{1}{2}(X_i^T\theta - Y_i)^2 = \frac{1}{2}\left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right)^2$

$$\Rightarrow \frac{\partial}{\partial\theta_j} l_i(\theta) = \left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right) X_{ij}$$

$$\therefore \nabla l_i(\theta) = \left(\left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right) X_{i1}, \left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right) X_{i2}, \cdots, \left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right) X_{ip}\right)^T$$

$$= \left(\sum_{k=1}^{p} X_{ik}\theta_k - Y_i\right)(X_{i1}, X_{i2}, \cdots, X_{ip})^T = (X_i^T\theta - Y_i)X_i$$

b) $\mathcal{L}(\theta) = \frac{1}{2}\|X\theta - Y\|^2 = \frac{1}{2}\sum_{i=1}^{N}(X\theta - Y)_i^2$

$$= \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{j=1}^{p} X_{ij}\theta_j - Y_i\right)^2$$

$$\therefore \frac{\partial}{\partial\theta_k}\mathcal{L}(\theta) = \frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_k}\left(\sum_{j=1}^{p} X_{ij}\theta_j - Y_i\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} 2\left(\sum_{j=1}^{p} X_{ij}\theta_j - Y_i\right) X_{ik}$$

$$= \sum_{i=1}^{N}\left(\sum_{j=1}^{p} X_{ij}\theta_j - Y_i\right) X_{ik} = \sum_{i=1}^{N}(X\theta - Y)_i X_{ik}$$

$$\therefore \nabla_\theta\mathcal{L}(\theta) = \left(\sum_{i=1}^{N}(X\theta - Y)_i X_{i1}, \sum_{i=1}^{N}(X\theta - Y)_i X_{i2}, \cdots, \sum_{i=1}^{N}(X\theta - Y)_i X_{ip}\right)^T$$

$$= \left((X\theta - Y)^T X\right)^T = X^T(X\theta - Y)$$

□

Problem 2. $f'(\theta) = \theta$.

$$\Rightarrow \theta^{k+1} = \theta^k - \alpha\theta^k = (1-\alpha)\theta^k = (1-\alpha)^2\theta^{k-1} = \cdots = (1-\alpha)^{k+1}\theta^0$$

if $\alpha > 2$, $\lim_{k\to\infty}\theta^{k+1} = \lim_{k\to\infty}(1-\alpha)^{k+1}\theta^0$ diverges since $|1-\alpha| > 1$.

□

Problem 3. $f(\theta) = \frac{1}{2}\|X\theta - Y\|^2 \Rightarrow \nabla_\theta f(\theta) = X^T(X\theta - Y)$

$$\Rightarrow \theta^{k+1} = \theta^k - \alpha X^T(X\theta^k - Y).$$

Subtracting $\theta^* = (X^TX)^{-1}X^TY$ from both sides yields

$$\theta^{k+1} - \theta^* = \theta^k - \alpha X^T(X\theta^k - Y) - (X^TX)^{-1}X^TY$$

$$= (I - \alpha X^TX)\theta^k + \alpha X^TY - (X^TX)^{-1}X^TY$$

$$= (I - \alpha X^TX)\theta^k - (I - \alpha X^TX)(X^TX)^{-1}X^TY$$

$$= (I - \alpha X^TX)(\theta^k - \theta^*)$$

$$\Rightarrow \|\theta^{k+1} - \theta^*\| = \|I - \alpha X^TX\|\ \|\theta^k - \theta^*\|.$$

Since $\|\cdot\|$ is an inner product, $1 = \| I - \alpha X^T X + \alpha X^T X \| \geq \| I - \alpha X^T X \| + \alpha \|X^T X\|$ (Triangle Inequality)

$\therefore \| I - \alpha X^T X \| \leq 1 - \alpha \|X^T X\|$

lemma. $\|X\| \leq \lambda$ for all eigenvector λ of X.

pf) Let $Xv = \lambda v$. Then, $\|Xv\| = \|X\| \|v\| \leq |\lambda| \|v\|$ (submultiplicativity)

$\Rightarrow \|X\| \leq |\lambda|$ ($\because \|v\| > 0$)

Since $\alpha > \frac{2}{\rho(X^T X)}$, $\alpha \|X^T X\| > 2$. ($\because \|X^T X\| \leq \rho(X^T X)$, by the above lemma)

Therefore, $\| I - \alpha X^T X \| \leq 1 - \alpha \|X^T X\| < -1$, which causes

$\|\theta^{k+1} - \theta^* \| = \| I - \alpha X^T X \|^{k+1} \| \theta^0 - \theta^k \|$ to diverge as $k \to \infty$ whenever $\theta^0 \neq \theta^k$.

Thus, θ^k diverges for $\alpha > \frac{2}{\rho(X^T X)}$ $\square$