2020-14281 이시윤 HW1..

#1.

(a)

- $\dfrac{\partial l_i(\theta)}{\partial \theta_j} = \dfrac{\partial \frac{1}{2}(X_{i1}\theta_1 + X_{i2}\theta_2 + \cdots + X_{ip}\theta_p - Y_i)^2}{\partial \theta_j} = (X_{i1}\theta_1 + X_{i2}\theta_2 + \cdots + X_{ip}\theta_p - Y_i)X_{ij}$

$= (X_i^T\theta - Y_i)X_{ij}$. Thus, $\nabla_\theta l_i(\theta) = (X_i^T\theta - Y_i)X_i$

$*\ X_i^T = [X_{i1}, X_{i2}, \cdots, X_{ip}]$

- $X^T(X\theta - Y) = (X_1^T\theta - Y_1)X_1 + (X_2^T\theta - Y_2)X_2 + \cdots + (X_N^T\theta - Y_N)X_N$

$L(\theta) = \frac{1}{2}\left((X_1^T\theta - Y_1)^2 + (X_2^T\theta - Y_2)^2 + \cdots + (X_N^T\theta - Y_N)^2\right)$

$\dfrac{\partial L(\theta)}{\partial \theta_j} = \dfrac{\partial l_1(\theta)}{\partial \theta_j} + \dfrac{\partial l_2(\theta)}{\partial \theta_j} + \cdots + \dfrac{\partial l_N(\theta)}{\partial \theta_j}$

$= (X_1^T\theta - Y_1)X_{1j} + (X_2^T\theta - Y_2)X_{2j} + \cdots + (X_N^T\theta - Y_N)X_{Nj}$

$\therefore \nabla_\theta L(\theta) = (X_1^T\theta - Y_1)X_1 + (X_2^T\theta - Y_2)X_2 + \cdots + (X_N^T\theta - Y_N)X_N = X^T(X\theta - Y)$

#2 $f(\theta) = \theta^2/2$

$f'(\theta^k) = \theta^k.$ * θ^k: kth iterate of Gradient descent

$\theta^1 = \theta^0 - \alpha f'(\theta^0) = \theta^0 - \alpha\theta^0$

$|\theta^1| = |1-\alpha||\theta^0|$

$|\theta^2| = |1-\alpha|^2|\theta^0|$

$\vdots$

$|\theta^n| = |1-\alpha|^n|\theta^0|.$

If $\alpha > 2$ and $|\theta^0| > 0$, $|1-\alpha| > 1$ Thus $|\theta^n| \to \infty$ as $n \to \infty$.

It diverges if $\alpha > 2$!

#3 $f(\theta) = \frac{1}{2}\|X\theta - Y\|^2$

$\nabla f(\theta^k) = X^T(X\theta^k - Y)$ holds (proved in #1).

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

Let $\theta^* = (X^TX)^{-1}X^TY$. then,

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

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

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

$\theta^n - \theta^* = (I - \alpha X^TX)(\theta^{n-1} - \theta^*) = (I - \alpha X^TX)^2(\theta^{n-2} - \theta^*) = \cdots = (I - \alpha X^TX)^n(\theta^0 - \theta^*)$

X^TX 의 largest eigenvalue 를 ρ라 하자. (절대값이 최대인 것)

$(\alpha X^TX)v = (\alpha\rho)v$; $(I - \alpha X^TX)v = v - \alpha\rho v = (1 - \alpha\rho)v$

만약 $\alpha > \frac{2}{\rho}$ 라면, $|1 - \alpha\rho| > 1$ 이고 $I - \alpha X^TX$ 의 eigenvalue 중 하나의 절대값은 1을 초과함을 알 수 있다.

$I - \alpha X^TX$ 는 Real Symmetric 이므로 diagonalizable 하다. 즉 $I - \alpha X^TX = PDP^{-1}$.

$\therefore (\theta^n - \theta^*) = (PD^nP^{-1})(\theta^0 - \theta^*) = P\begin{pmatrix} \lambda_1^n & & \\ & \lambda_2^n & \\ & & \ddots & \\ & & & \lambda_p^n \end{pmatrix}P^{-1}(\theta^0 - \theta^*)$

$\lambda^* = \max\{|\lambda_1|, \ldots, |\lambda_p|\} > 1$ 이므로 $(\lambda^*)^n \to \infty$ as $n \to \infty$ 이고

따라서 $\alpha > \frac{2}{\rho}$ 라면 대부분의 θ^0 에 대해 $\theta^n - \theta^*$ 가 발산, 즉 θ^n 이 발산한다.