

HW1

권규보
산업공학과
2018-18249

Problem 1. (a) $\ell_i(\theta) = \frac{1}{2} (X_i^T \theta - y_i)^2$

$$X_i^T \theta - y_i = g(\theta)$$

$$\ell_i(\theta) = \frac{1}{2} (g(\theta))^2$$

$$\begin{aligned} \nabla \ell_i(\theta) &= g(\theta) \nabla g(\theta) = g(\theta) \cdot (X_i^T)^T \\ &= (X_i^T \theta - y_i) \cdot X_i \end{aligned}$$

$$(b) \quad L(\theta) = \frac{1}{2} (X\theta - y)^T (X\theta - y)$$

$$= \frac{1}{2} (\theta^T X^T X \theta - 2y^T X \theta + y^T y)$$

$$\nabla L(\theta) = X^T X \theta - X^T y = X^T (X\theta - y)$$

Problem 2.

$$f(\theta) = \frac{\theta^2}{2}$$

$$f'(\theta) = \theta$$

$$\begin{aligned}\theta^{k+1} &= \theta^k - \alpha f'(\theta^k) \\ &= \theta^k (1 - \alpha)\end{aligned}$$

$$\therefore \theta^k = \theta^0 (1 - \alpha)^k$$

$\theta^0 \neq 0$, $\alpha > 2 \cdot 10^{-4}$ $1 - \alpha < -1$ \therefore diverges

Problem 3

$$\nabla f(\theta^k) = X^T(X\theta - Y)$$

$$\begin{aligned}\theta^{k+1} &= \theta^k - \alpha \nabla f(\theta^k) = \theta^k - \alpha(X^T X \theta^k - X^T Y) \\ &= (I - \alpha X^T X) \theta^k + \alpha X^T Y\end{aligned}$$

$$\begin{aligned}\theta^{k+1} - \theta^* &= (I - \alpha X^T X) \theta^k + \alpha X^T Y - (X^T X)^{-1} X^T Y \\ &= (I - \alpha X^T X) (\theta^k - \theta^*) \\ &= (I - \alpha X^T X)^k (\theta^0 - \theta^*)\end{aligned}$$

PSD인 $X^T X$ 은 대각화하면 $S \Lambda S^T$ 가 된다.

$$(I - \alpha X^T X)^k = S (I - \alpha \Lambda)^k S^{-1}$$

$I - \alpha \Lambda$ 의 대각성분 중 Λ 에서 가장 큰 원소 $\rho(X^T X)$ 의 위치만 생각하면

$$1 - \frac{2}{\rho(X^T X)} \cdot \rho(X^T X) < -1 \text{ 이므로 이 위치의 } \theta^0 - \theta^* \text{가}$$

0이 아니다면 (즉, 이 위치의 initial 성분이 0이 아니기 optimal이 되지 않으면) diverges.