#1

(a) $l_i(\theta) = \frac{1}{2}(X_i^T\theta - Y_i)^2$ $\quad \left(\theta = (\theta_1, \cdots, \theta_p) \in \mathbb{R}^p\right)$ $\quad (x_{ij} \in \mathbb{R}$ for $1 \le j \le p)$

$$\frac{\partial l_i}{\partial \theta_j} = (X_i^T\theta - Y_i) \cdot \frac{\partial}{\partial \theta_j}(X_i^T\theta - Y_i) = (X_i^T\theta - Y_i) \cdot x_{ij} \quad (X_i^T = (x_{i1}, x_{i2}, \cdots, x_{ip}))$$

$$\Rightarrow \nabla_\theta l_i(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} l_i(\theta) \\ \vdots \\ \frac{\partial}{\partial \theta_p} l_i(\theta) \end{bmatrix} = (X_i^T\theta - Y_i)\begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} = (X_i^T\theta - Y_i) X_i.$$

(b) $$X\theta - Y = \begin{bmatrix} X_1^T\theta - Y_1 \\ \vdots \\ X_N^T\theta - Y_N \end{bmatrix} \Rightarrow \mathcal{L}(\theta) = \sum_{i=1}^N \frac{1}{2}(X_i^T\theta - Y_i)^2 = \sum_{i=1}^N l_i(\theta).$$

$$\nabla_\theta \mathcal{L}(\theta) = \sum_{i=1}^N \nabla_\theta l_i(\theta) = \sum_{i=1}^N (X_i^T\theta - Y_i) X_i = [X_1 \cdots X_N] \cdot \begin{bmatrix} X_1^T\theta - Y_1 \\ \vdots \\ X_N^T\theta - Y_N \end{bmatrix} = X^T(X\theta - Y)$$

#2

$$\theta^{k+1} = \theta^k - \alpha\theta^k = (1-\alpha)\theta^k.$$

$$\Rightarrow \theta^i = (1-\alpha)^i \cdot \theta^0$$

If $\alpha > 2 \Rightarrow |1-\alpha| > 1 \Rightarrow |\theta^i| = |1-\alpha|^i \cdot |\theta^0|$

$\Rightarrow \lim_{i\to\infty} |\theta^i| \to \infty$ since $|1-\alpha|^i \to \infty$ when $|\theta^0| \neq 0$.

#3.

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

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

$$= (I_p - \alpha X^TX)\theta^k + \alpha X^TY.$$

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

$$\Rightarrow \theta^k - \theta^* = (I_p - \alpha X^TX)^k \cdot (\theta^0 - \theta^*).$$

X^TX is a real valued symmetric matrix, hence by the Spectral thm, X^TX is diagonalizable.

Denote the eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$.

Let) $X^TX = QDQ^{-1}$ for $\exists$ invertible matrix $Q \in \mathbb{R}^{p\times p}$, $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_n \end{bmatrix}$

$\Rightarrow \theta^k - \theta^* = [Q(I_p - \alpha D)Q^{-1}]^k \cdot (\theta^0 - \theta^*) = Q \cdot (I_p - \alpha D)^k \cdot Q^{-1} \cdot (\theta^0 - \theta^*)$

$$(I_p - \alpha D)^k = \begin{bmatrix} (1-\alpha\lambda_1)^k & & & 0 \\ & (1-\alpha\lambda_2)^k & & \\ & & \ddots & \\ 0 & & & (1-\alpha\lambda_n)^k \end{bmatrix}$$

Since $\alpha > \dfrac{2}{\rho(X^TX)} = \dfrac{2}{\lambda_1}$, $1-\alpha\lambda_1 < -1$, and this implies at least the first component in the diagonal diverges as $k \to \infty$.

Define $S = \{\theta^0 \mid \theta^k \text{ converges as } k \to \infty\}$

$Q^{-1}(\theta^0 - \theta^*) \in \{v \mid v \in \mathbb{R}^p,\ v = (0, c_2, \cdots, c_p), c_i \in \mathbb{R}\}$ is necessary for θ^k to converge.

$\Rightarrow S \subset T := \{Qv + \theta^* \mid v \in \mathbb{R}^p,\ v = (0, c_1, \cdots, c_p), c_i \in \mathbb{R}\}$

$\dim(T) = p-1$ since Q is invertible, and hence a 1-1 mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$.

$\therefore \dim(S) < p$, and this means that θ^k diverges for most starting points $\theta^0 \in \mathbb{R}^p$.