P1

(a)

$$l_i(\theta) = \frac{1}{2}(X_i^T\theta - Y_i)^2 = \frac{1}{2}\left(\sum_j X_{ij}\theta_j - Y_i\right)^2$$

$$\frac{\partial}{\partial\theta_k} l_i(\theta) = \frac{\partial}{\partial\theta_k}\left(\frac{1}{2}\left(\sum_j X_{ij}\theta_j - Y_i\right)^2\right)$$

$$= \left(\sum_j X_{ij}\theta_j - Y_i\right) X_{ik} = (X_i^T\theta - Y_i) X_{ik}$$

$$\therefore \nabla_\theta l_i(\theta) = (X_i^T\theta - Y_i)(X_{i1} \ X_{i2} \cdots X_{ip})^T = (X_i^T\theta - Y_i) X_i \quad \divideontimes$$

(b)

$$\mathcal{L}(\theta) = \frac{1}{2}\|X\theta - Y\|^2 = \frac{1}{2}\sum_{i=1}^{N}(X\theta - Y)_i^2 = \frac{1}{2}\sum_{i=1}^{N}\left(\sum_{j=1}^{p} X_{ij}\theta_j - Y_i\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{p}\frac{1}{2}(X_{ij}\theta_j - Y_i)^2$$

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

$$\therefore \nabla_\theta \mathcal{L}(\theta) = ((X\theta - Y)^T X)^T = X^T(X\theta - Y) \quad \divideontimes$$

P2

$$\theta_{k+1} = \theta_k - \alpha f'(\theta_k) = \theta_k - \alpha\theta_k = (1-\alpha)\theta_k$$

$$= (1-\alpha)^{k+1}\theta_0$$

if $|1-\alpha| > 1$, it diverges $\therefore \alpha > 2$ makes GD diverge.

P3

$$\min_{\theta\in\mathbb{R}^p} f(\theta), \quad f(\theta) = \frac{1}{2}\|X\theta - Y\|^2$$

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

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

$$= \theta_k - \theta_* - \alpha X^T X(\theta_k - \theta_*) \quad (\because X\theta_* = Y)$$

$$= \theta_k - \theta_* - \alpha\rho(\theta_k - \theta_*)$$

$$= (1-\alpha\rho)(\theta_k - \theta_*)$$

$$\therefore \|\theta_{k+1} - \theta_*\| = |1-\alpha\rho|\,\|\theta_k - \theta_*\|$$

Thus, using the result of problem 2, $\alpha\rho > 2 \iff \alpha > \frac{2}{\rho}$ makes $\{\theta_k\}$ diverge.