~~Cost~~ Q3

For MSE loss,

$$J\text{theta} = \frac{1}{N}\sum_{i=1}^{N}(h_\theta(x) - y)^2$$

Incase of logistic regression

$$h_\theta(x) = \frac{1}{1+e^{-\theta x}}$$

$$\frac{\partial J_\theta}{\partial \theta} = \frac{2[h_\theta(x)-y]}{N}\sum_{i=1}^{N}\left(\frac{dh_\theta(x)}{d\theta}\right)$$

for 1 dataset,

$$= 2[h_\theta(x)-y]\frac{dh_\theta(x)}{d\theta}$$

$$= 2[h_\theta(x)-y] \times \frac{-1}{(1+e^{-\theta x})^2} \times -xe^{-\theta x}$$

$$= 2[h_\theta(x)-y] \times \frac{1+e^{-\theta x}-1}{(1+e^{-\theta x})^2}\, x$$

$$= 2[h_\theta(x)-y] \times \frac{1}{e^{-\theta x}+1} \times \left[1 - \frac{1}{e^{-\theta x}+1}\right] x$$

$$= 2[h_\theta(x) - y] \times h_\theta(x)[1 - h_\theta(x)]x$$

Hence, when

$$h_\theta(x) = 1, \quad \frac{\partial J_\theta}{\partial \theta} = 0$$

$$h_\theta(x) = 0, \quad \frac{\partial J_\theta}{\partial \theta} = 0$$

Here gradient will be same everytime as its derivative will always be 0, so model will not be able to learn.

Using cross entropy loss,

$$\frac{\partial J_\theta}{\partial \theta} = [h_\theta(x) - y]x$$

Using this model will be ~~abb~~ able to learn as it's value is ~~not~~ is not constant like before.

Qs. $Y = X\beta + \epsilon_0$

$Y - \epsilon_0 = X \cdot \beta$

$\Rightarrow \hat{Y} = X\beta$

$$J(\beta) = \frac{1}{N} \sum_{i=1}^{N} |\hat{Y} - Y|^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} |X\beta - Y|^2$$

$$\frac{\partial J\theta}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{N} (X_\beta - Y)^2$$

for datapoint

$$\frac{\partial J\theta}{\partial \theta} = \frac{\partial [X_\beta - Y]^2}{\partial \beta}$$

$$Y = \frac{\partial}{\partial \beta} [X\beta - Y]^T [X\beta - Y]$$

$$\cancel{= \frac{\partial}{\partial \beta} [X^T X\beta - X^T Y - Y]}$$

Using ~~N~~ rule

$$\cancel{= X^T X\beta - X^T Y + X^T X\beta - X Y^T}$$

$$= \frac{\partial}{\partial \beta} [(X\beta)^T X\beta - (X\beta)^T Y - Y^T (X\beta) + Y^T Y$$

$$= \frac{\partial}{\partial \beta}\left[x^T \beta^T x\beta - 2x^T\theta^T y - y^T y \right)$$

$$= \frac{\partial}{\partial \theta}\left[x^T \beta^T x \beta - 2x^T \beta^T y \right]$$

$$= x^T x \frac{\partial \beta^T \beta}{\partial \beta}$$

$$= 2x^T x \beta - 2x^T y$$

for $\frac{\partial J_\theta}{\partial \beta} = 0$,

$$2x^T x \beta = 2x^T y$$

$$\Rightarrow x^T x \beta = x^T y$$

$$\Rightarrow (x^T x)^{-1} \cancel{(x^T x)} \beta = (x^T x)^{-1} x^T y$$

$$\Rightarrow \beta = (x^T x)^{-1} \cdot x^T \cdot Y$$

Solution will exist only when $(x^T x)$ have inverse exist. So, there must not be a duplicate data.