Gradient of MSE Derivation -

$$C = \frac{1}{n} \Sigma (\hat{y}_i - y_i)^2$$

$\hat{y}_i$ - predicted value

y_i - actual value

$$C = \frac{1}{n} \Sigma u^2 \qquad u = \hat{y}_i - y_i$$

$$\frac{\partial u}{\partial \hat{y}_i} = 1 - 0 = 1$$

$$\frac{\partial C}{\partial u} = \frac{1}{n} \Sigma 2u$$

$$= \frac{2}{n} \Sigma u = \frac{2}{n} \Sigma (\hat{y}_i - y_i)$$

$$\therefore \frac{\partial C}{\partial u} = \frac{2}{n} \Sigma (\hat{y}_i - y_i) \quad \& \quad \frac{\partial u}{\partial \hat{y}_i} = 1$$

$$\hat{y}_i = m x_i + b$$

$$\frac{\partial \hat{y}_i}{\partial m} = x_i \qquad \frac{\partial \hat{y}_i}{\partial b} = 1$$

Using Chain Rule:

$$\frac{\partial C}{\partial b} = \frac{\partial C}{\partial u} \cdot \frac{\partial u}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial b} = \frac{2}{n} \Sigma (\hat{y}_i - y_i) \cdot 1 \cdot 1$$

$$\frac{\partial C}{\partial m} = \frac{\partial C}{\partial u} \cdot \frac{\partial u}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial m} = \frac{2}{n} \Sigma (\hat{y}_i - y_i) \cdot 1 \cdot x_i$$

Gradient Descent Update Rules (MSE) -

$$b := b - \alpha \frac{\partial C}{\partial b}$$

$$m := m\cancel{b} - \alpha \frac{\partial C}{\partial m}$$

$$\Rightarrow \quad b := b - \alpha \cdot \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

$$m := m - \alpha \frac{2}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \cdot x_i$$