

# Derivation of Gradient Descent

Cost function Linear regression  $\rightarrow$  MSE

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})^2$$

where  $f(w, b)(x) = w_1 x + b$   $m \rightarrow$  no. of samples

Gradient Descent algorithm

$$w = w - \alpha \frac{\partial J(w, b)}{\partial w}$$

$$b = b - \alpha \frac{\partial J(w, b)}{\partial b}$$

(Repeat till convergence).

$$\frac{1}{2m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})^2$$

$$\Rightarrow \frac{1}{2m} \sum_{i=1}^m (w x^{(i)} + b - y^{(i)})^2 \cdot x^{(i)}$$

$$\Rightarrow \frac{1}{m} \sum_{i=1}^m (w x^{(i)} + b - y^{(i)}) \cdot (x^{(i)})$$

(By applying chain rule)

$$\underbrace{J(w, b)}_{\text{JW}} = \frac{1}{m} \sum_{i=1}^m (f_{w, b}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

$$\frac{\partial J(w, b)}{\partial b} = \frac{1}{m} \sum_{i=1}^m (f_{w, b}(x^{(i)}) - y^{(i)})^2$$

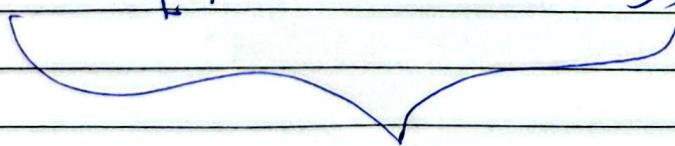
$$\rightarrow \frac{\partial}{\partial b} \frac{1}{m} \sum_{i=1}^m (w x^{(i)} + b - y^{(i)})^2$$

$$\rightarrow \frac{1}{m} \sum_{i=1}^m (w x^{(i)} + b - y^{(i)})$$

$$\frac{\partial J(w, b)}{\partial w} = \frac{1}{m} \sum_{i=1}^m (f_{w, b}(x^{(i)}) - y^{(i)})$$

$$w \leftarrow w - \frac{1}{m} \sum_{i=1}^m (f_{w, b}(x^{(i)}) - y^{(i)}) x^{(i)}$$

$$b \leftarrow b - \frac{1}{m} \sum_{i=1}^m (f_{w, b}(x^{(i)}) - y^{(i)})$$



Gradient Descent for Linear Regression (MLP)