Derivation of Gradient Descent

Cost function in Linear Regression → MSE (Mean squared error).

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

where $f_{w,b}(x) = wx + b$ $\qquad$ $m \to$ no. of tupples.

Gradient Descent algorithm →

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

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

(Repeat until convergence).

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

$$\Rightarrow \frac{\partial}{\partial w} \frac{1}{2m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)} \right)^2$$

$$\Rightarrow \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)} \right)^{\cancel{2}} \times \cancel{2}\, x^{(i)}$$

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

(By applying chain rule).

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

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

$$\Rightarrow \frac{\partial}{\partial b}\frac{1}{2m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)^2$$

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

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m}\left(wx^{(i)} + b - y^{(i)}\right)$$

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

$$\underbrace{\begin{aligned} w &= w - \alpha \sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)} \\ b &= b - \alpha \sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right) \end{aligned}}_{\text{Gradient descent for linear regression (MSE)}}$$