Derivation of Gradient Descent

Cost function linear regression → MSE

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

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

Gradient Descent algorithm →

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

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

(Repeat till convergence).

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

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

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

(By applying chain rule)

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

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

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

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

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

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

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

Gradient Descent for linear regression (IMP)