Gradient Descent

05 November 2023 11:27

Gradient Descent

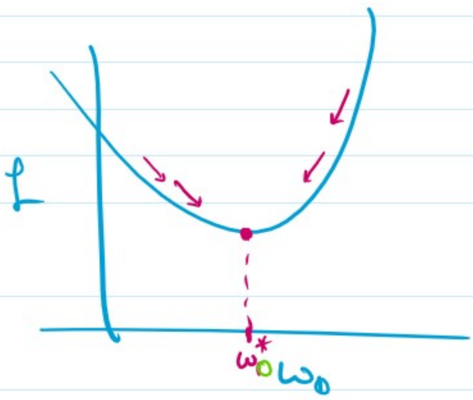
MSE $L = \frac{1}{N} \sum (\hat{y} - y)^2$; $\hat{y} = w_0 + w_1 x$

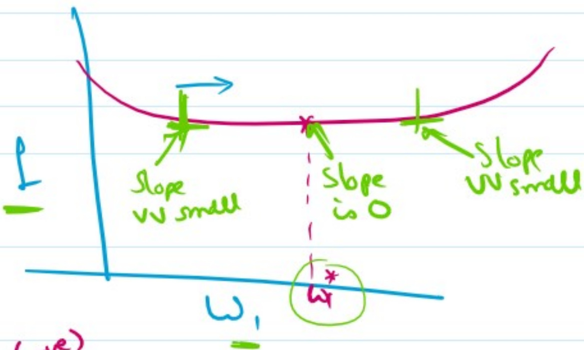
(Derivatives)
Gradients of the Loss f^n w.r.t. each var (w_j) ..

$$\frac{\partial L}{\partial w_0} = \frac{2}{N} \sum (w_0 + w_1 x) \cdot 1 \quad \text{—— (1)}$$

$$\frac{\partial L}{\partial w_1} = \frac{2}{N} \sum (w_0 + w_1 x) \cdot x \quad \text{—— (2)}$$

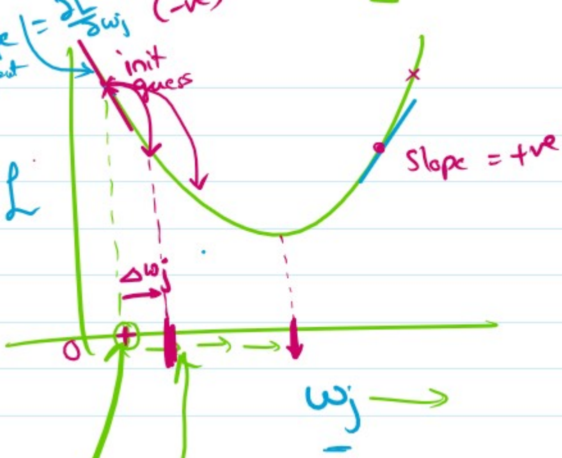
Objective is find the optimal values of w_j
for which the MSE Loss f^n gets minimized
Simultaneously w.r.t. each w_j





Gradient Descent Algo

Slope of tangent



iteration no.

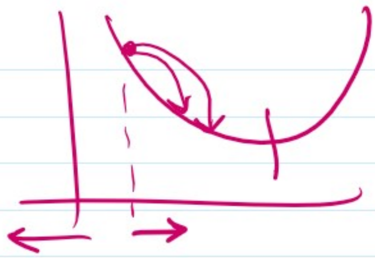
$$w_j^{k+1} = w_j^{k} - \Delta w_j$$

amt. of update

Value of w_j in the current iteration

Updated w_j
(Value in the next iteration)

Updated w_j
(Value in the next iteration)



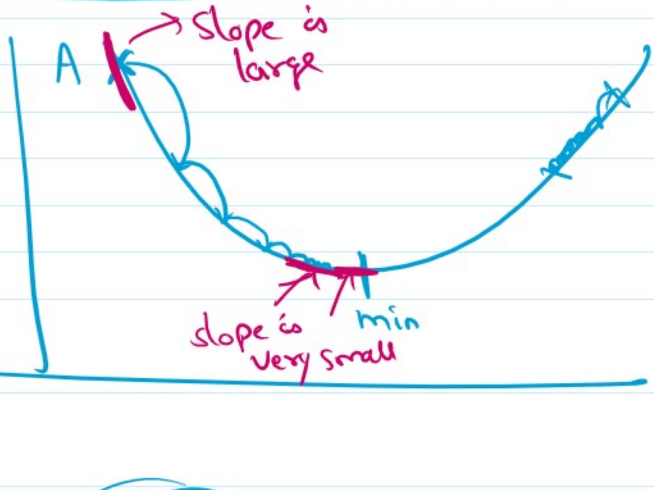
Amt. of Update

Direction of Update → Gradient $\left(\frac{\partial L}{\partial w_j}\right)$

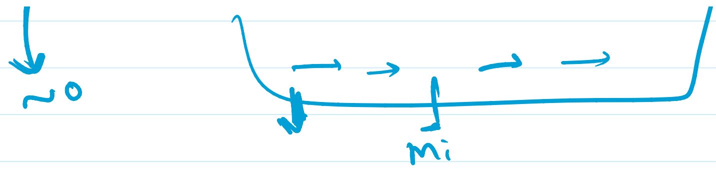
Quantum of "jump" → Learning Rate (α) → Hyperparameter (means we specify that value)

GD update Rule

$$w_j^{k+1} = w_j^{k} - \underbrace{\left[\alpha \cdot \left(\frac{\partial L}{\partial w_j}\right)\right]}_{\Delta w_j}$$



$$\Delta w_j = \underbrace{\alpha \cdot \frac{\partial L}{\partial w_j}}_{\sim 0} + \underbrace{\text{momentum team}}$$


~0
mi