2025-07-22 - Compare and Contrast among activation functions, loss functions, optimizers and regularization for choosing the appropriate method for the given application

22 July 2025 11:35

⇒ Optimization Procedures :

⇒ Gradient Descent :

θ ~ Parameters

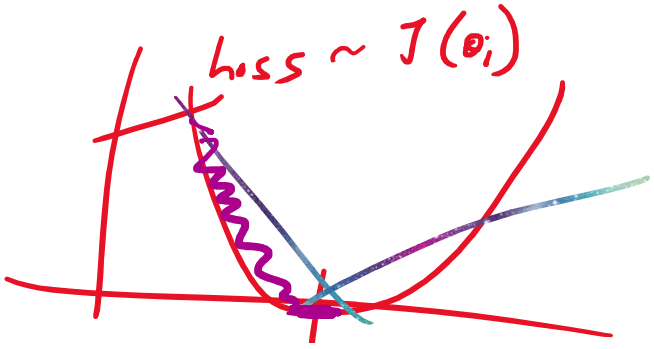
J ~ Loss Functions

$\frac{d}{d\theta_i} J(\theta_i)$ — Parameters, Loss Fn

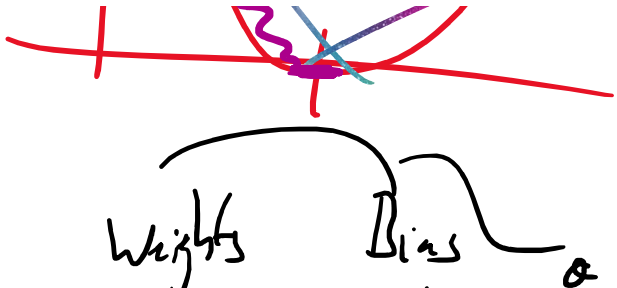
Gradient Calculation

⇒ Process (Textbooks) :

$$\theta_i := \theta_i - \alpha \times \frac{d}{d\theta_i} J(\theta_i)$$



(Step Increments)
Learning Rate (α)



$[0.00005 \text{ to } 0.005]$

$$\Rightarrow \hat{y} = \underline{w} \times x_i + \underline{b}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\Rightarrow \quad d/dw = \frac{1}{N} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$d/dw = \frac{1}{N} \sum_{i=1}^{n} (y_i - (wx_i + b))^2$$

$$\rightarrow d/dw = \frac{1}{N} \sum 2x (y_i - (wx_i + b)) \times d/dw \left(y_i - (w \cdot x_i + b)\right)$$

$$= \frac{2}{N} \sum (y_i - (wx_i + b)) \times (-x_i)$$

$$d/dw = -\frac{2}{N} \sum (x_i) \times (y_i - (wx_i + b))$$

$$d/db = \frac{1}{N} \sum (y_i - (wx_i + b))^2$$

$$\partial/\partial L = \frac{1}{N} \Sigma (y_i - (W x_i + b))$$

$$= \frac{2}{N} \Sigma (y_i - (W x_i + b)) \times (-1)$$

$$\partial/\partial b = -\frac{2}{N} \Sigma (y_i - (W x_i + b))$$

$\Rightarrow$ Parameter Recalculation:

$$W = W - (\alpha \times \{ -\frac{2}{N} \Sigma (x_i) \times (y_i - (W x_i + b)) \})$$

$$b = b - (\alpha \times \{ -\frac{2}{N} \Sigma (y_i - (W x_i + b)) \})$$