

5 Multi layered Perceptron Model [Artificial Neural N/W]



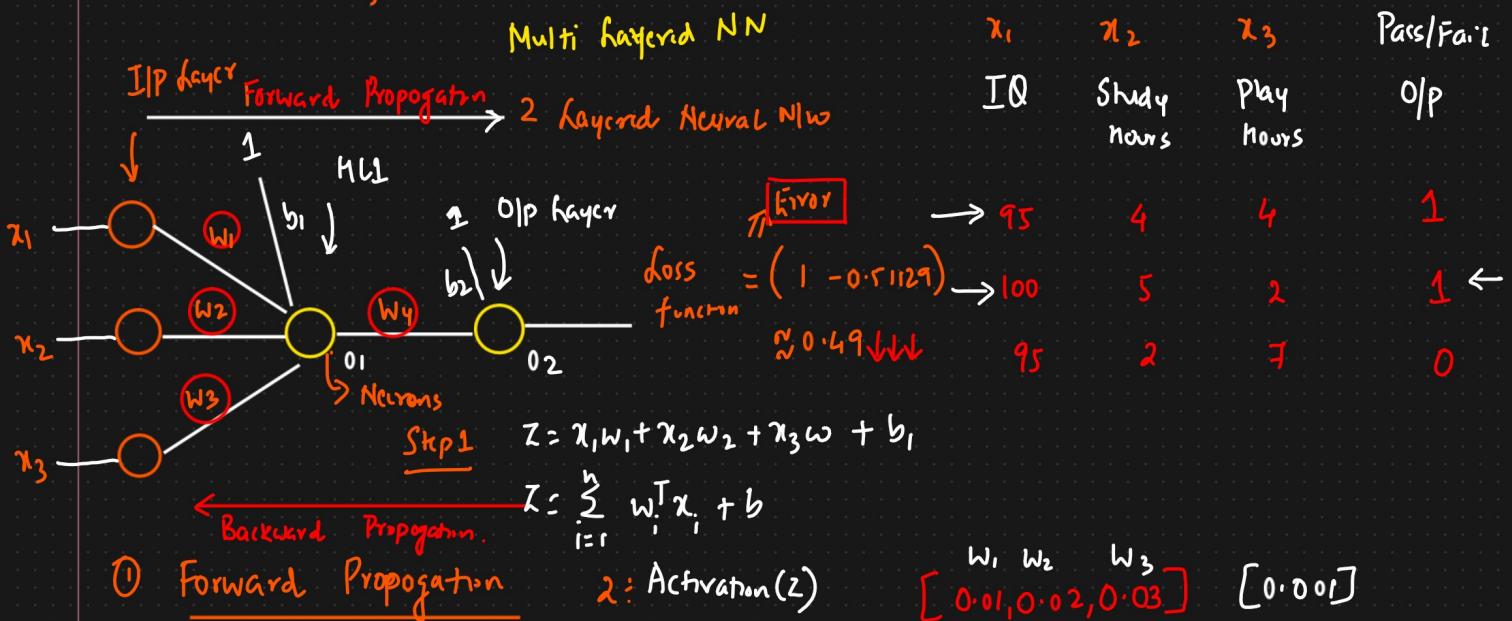
① Forward Propagation

② Backward Propagation → Geoffrey Hinton →

③ Loss function

④ Optimizers ✓

⑤ Activation function.



Hidden Layer 1

$$\begin{aligned} \text{Step 1: } Z &= 95 \times 0.01 + 4 \times 0.02 + 4 \times 0.03 + 1 \times 0.01 \\ &= 1.151 \end{aligned}$$

$$\text{Sigmoid.} = \frac{1}{1+e^{-z}}$$

$$\text{Step 2: Activation}(Z)$$

$$f(z) = \frac{1}{1+e^{-1.151}} = 0.759$$

$$o_1 = 0.759$$

Hidden Layer 2

$$w_4 = 0.02$$

$$b_2 = 0.03$$

$$\text{Step 1: } Z = o_1 * w_4 + b_2$$

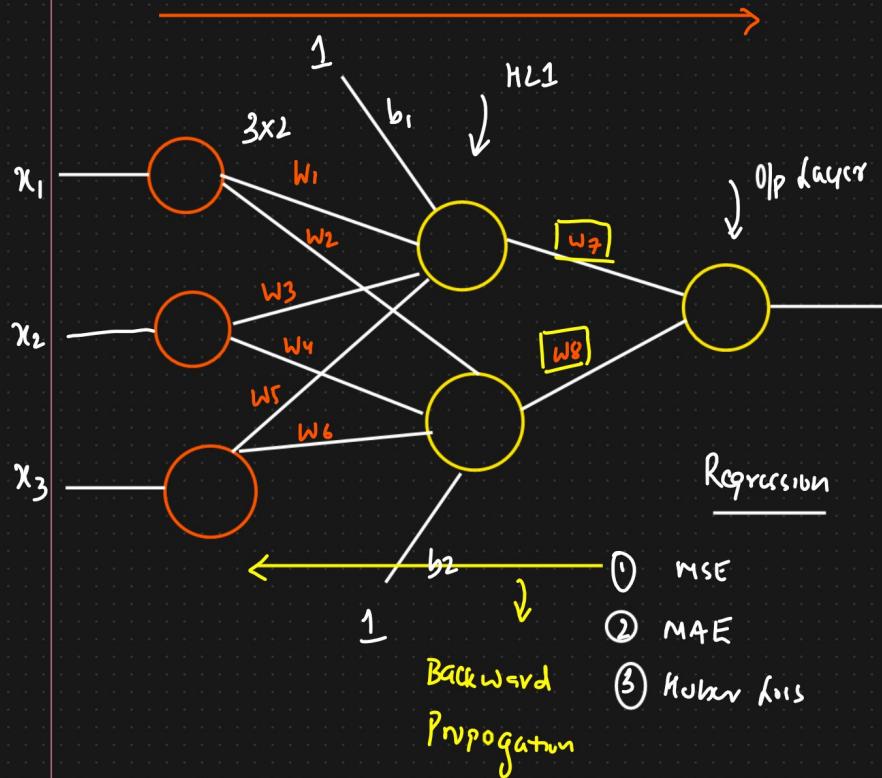
$$= 0.759 * 0.02 + 0.03$$

$$= 0.04518$$

Step 2 : Activation(z) $\frac{1}{1+e^{-(0.04518)}} = 0.51129$

$$O_2 = 0.51129 \Rightarrow \hat{y}$$

⑥ Back Propagation And Weight Updation Formula



Loss function
 $(y - \hat{y})^2$

Regression

- ① MSE
- ② MAE
- ③ Huber Loss

Classification

- ① Binary Cross Entropy
- ② Categorical Cross Entropy

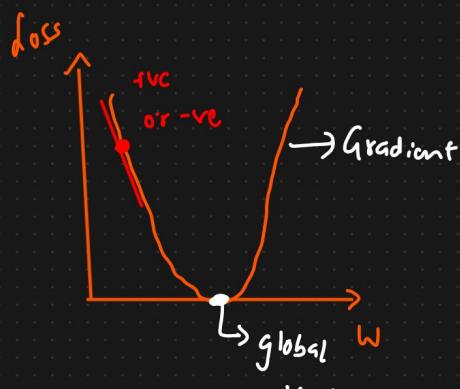
Weight update Formula

$$w_{7\text{new}} = w_{7\text{old}} - \eta \left[\frac{\partial h}{\partial w_{7\text{old}}} \right]$$

slope

$$w_{8\text{new}} = w_{8\text{old}} - \eta \left[\frac{\partial h}{\partial w_{8\text{old}}} \right]$$

Rate



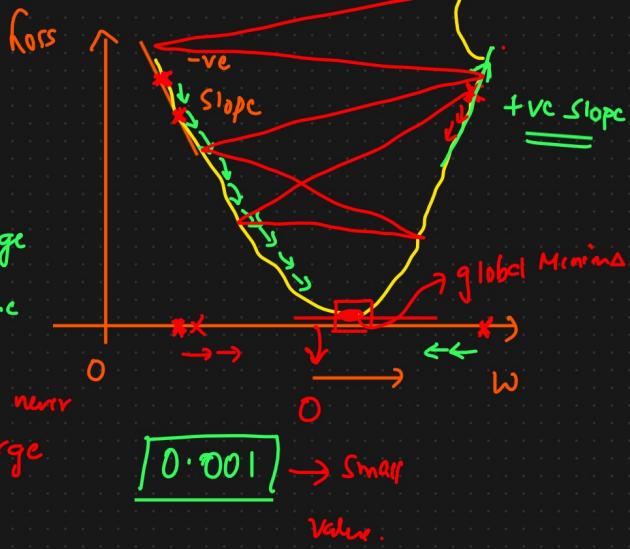
$$w_{\text{new}} = w_{\text{old}} - \eta \left[\frac{\partial h}{\partial w_{\text{old}}} \right] \Rightarrow \text{Weight Updation Formula.}$$

Gradient Descent

↗ Optimizers

I/P featur.s			Pass/Fail
x_1	x_2	x_3	
IQ	Study hours	Play hours	O/P
> 95	4	4	1
→ 100	5	2	1 ←
95	2	7	0

Optimizers : To reduce the loss value



$$W_{\text{new}} = W_{\text{old}} - \eta \quad (-\text{ve})$$

$$= W_{\text{old}} + \eta \quad (+\text{ve})$$

$$\boxed{W_{\text{new}} > W_{\text{old}}}$$

η = large value

It may never converge

Decreasing Rate

$$\boxed{0.001} \rightarrow \text{Small value.}$$

$$W_{\text{new}} = W_{\text{old}} - \eta \quad (+\text{ve})$$

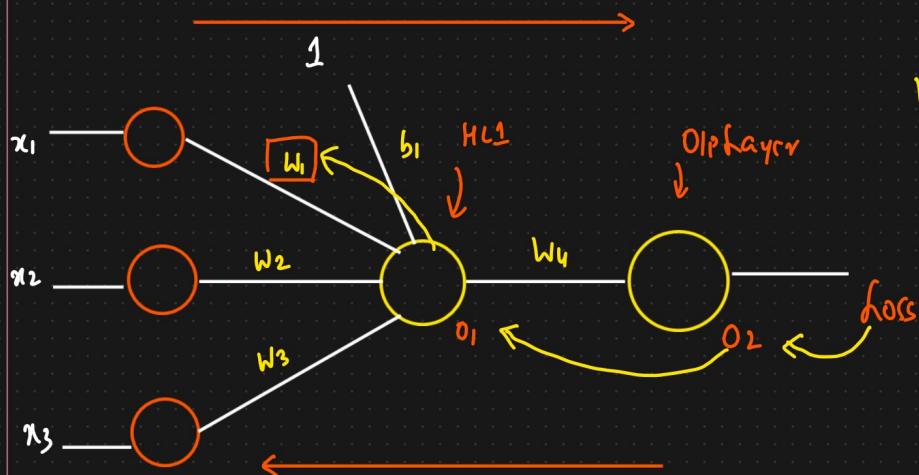
$$= W_{\text{old}} - \eta \quad (+\text{ve})$$

$$\boxed{W_{\text{new}} < W_{\text{old}}}$$

When w reaches Global Minima

$$\boxed{W_{\text{new}} = W_{\text{old}}}$$

⑦ Chain Rule of Derivative



$$W_{\text{new}} = W_{\text{old}} - \eta \frac{\partial h}{\partial W_{\text{old}}}$$

$$W_{4\text{new}} = W_{4\text{old}} - \eta \boxed{\frac{\partial h}{\partial W_{4\text{old}}}}$$

$$\frac{\partial h}{\partial w_{1,old}} = \frac{\partial h}{\partial o_2} * \frac{\partial o_2}{\partial w_{1,old}} \Rightarrow \text{Chain Rule of Derivation}$$

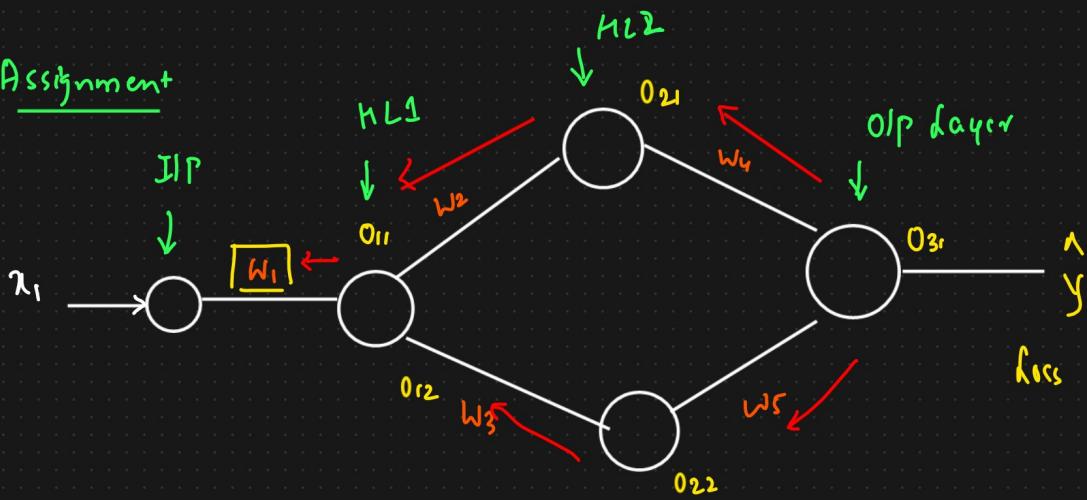
$$w_{1,new} = w_{1,old} - \eta \left[\frac{\partial h}{\partial w_{1,old}} \right]$$

$$\boxed{\frac{\partial h}{\partial w_{1,old}} = \frac{\partial h}{\partial o_2} * \frac{\partial o_2}{\partial o_1} * \frac{\partial o_1}{\partial w_{1,old}}}$$

$w_{2,new}$

$w_{3,new}$

Assignment



$$w_{1,new} = w_{1,old} - \eta \left[\frac{\partial h}{\partial w_{1,old}} \right]$$

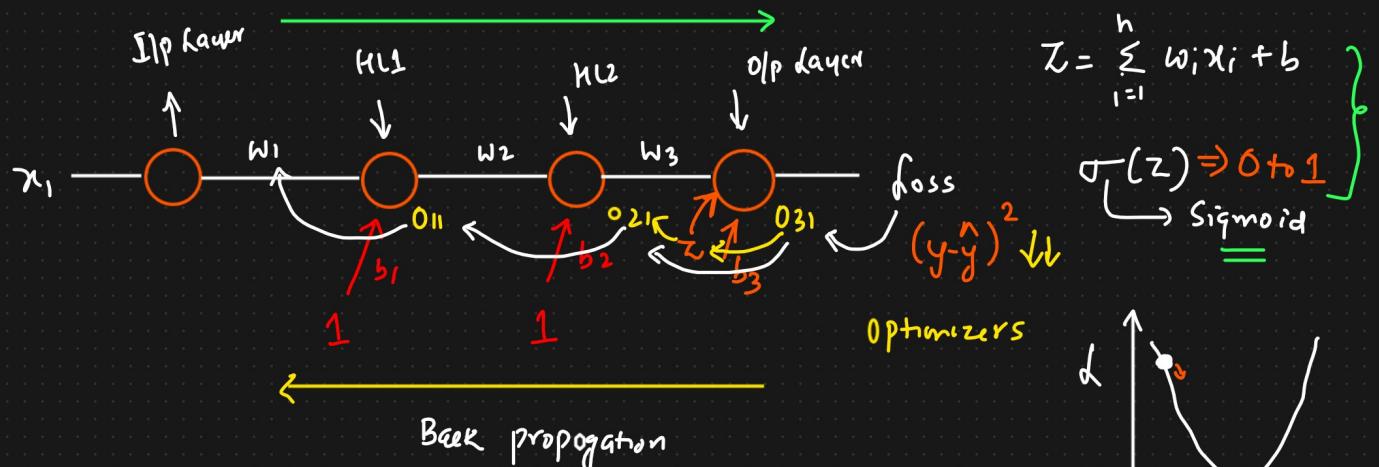


$$\frac{\partial h}{\partial w_{1,old}} = \left[\frac{\partial h}{\partial o_{31}} * \frac{\partial o_{31}}{\partial o_{21}} * \frac{\partial o_{21}}{\partial o_{11}} * \frac{\partial o_{11}}{\partial w_{1,old}} \right]$$

+

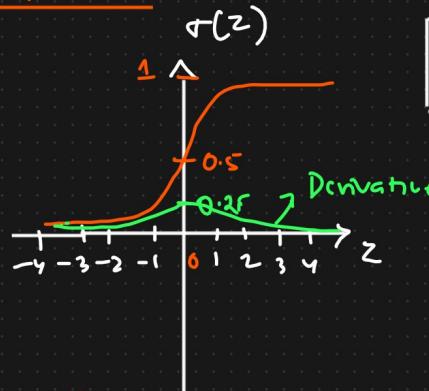
$$\left[\frac{\partial h}{\partial o_{31}} * \frac{\partial o_{31}}{\partial o_{22}} * \frac{\partial o_{22}}{\partial o_{12}} * \frac{\partial o_{12}}{\partial w_{1,old}} \right]$$

⑧ Vanishing Gradient Problem And Activation functions



⑨ Sigmoid Activation function

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



$$0 \leq \sigma(z) \leq 1$$

Derivative ($\sigma(z)$)

Vanishing
Gradient
Problem.

$$0 \leq \sigma(z) \leq 0.25$$

$$w_{1,\text{new}} = w_{1,\text{old}} - \eta \left[\frac{\partial h}{\partial w_{1,\text{old}}} \right] \Rightarrow \text{small value} \Rightarrow \boxed{w_{1,\text{new}} \approx w_{1,\text{old}}}$$

$$\frac{\partial h}{\partial w_{1,\text{old}}} = \frac{\partial h}{\partial w} * \boxed{\frac{\partial w}{\partial o_{31}}} * \boxed{\frac{\partial o_{31}}{\partial o_{21}}} * \boxed{\frac{\partial o_{21}}{\partial o_{11}}} * \boxed{\frac{\partial o_{11}}{\partial w_{11}}} \Rightarrow \text{smaller value}$$

$$o_{31} = \sigma \left(\underbrace{w_3 * o_{21}}_z + b_3 \right) \quad Z = w_3 * o_{21} + b_3$$

$$o_{31} = \sigma(z)$$

$$\frac{\partial o_{31}}{\partial o_{21}} = \frac{\partial (\sigma(z))}{\partial (z)} * \frac{\partial z}{\partial o_{21}} \quad \left\{ \text{Chain Rule} \right\}$$

$$0 \leq \sigma(z) \leq 0.25 \quad * \quad \frac{\partial((w_3 * o_{21}) + b_3)}{\partial(o_{21})}$$

↓

Derivative of
Sigmoid

$$\left| \frac{\partial o_{31}}{\partial o_{21}} = 0 \leq \sigma(z) \leq 0.25 \quad * \quad w_3 \right. \quad \left. \begin{matrix} \text{old} \\ \end{matrix} \right\}$$

④ To fix this problem Researchers started exploring other Activation function

① Tanh ② ReLU ③ PReLU ④ Swiss