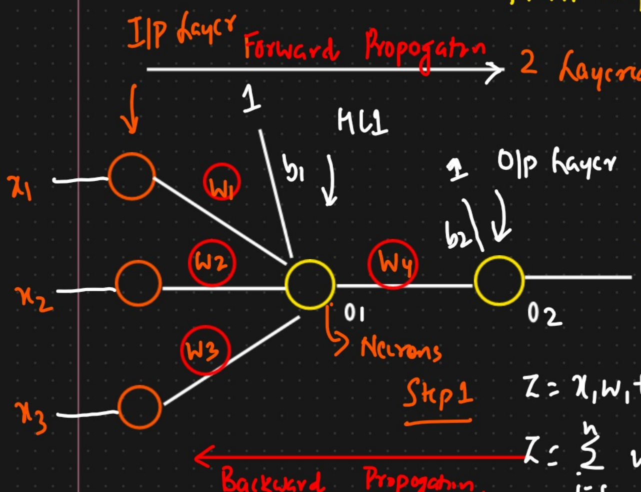
5 Multi Layered Perceptron Model [Artificial Neural N/W]

1. Forward Propogation
2. Backward Propogation → Geoffrey Hinton →

3. Loss function
4. Optimizers ✓
5. Activation function.

Multi Layered NN

I/P Layer — Forward Propogation → 2 Layered Neural N/w



Loss function $= (1 - 0.51129)$ ≈ 0.49 ↓↓↓ — Error

x_1 IQ	x_2 Study hours	x_3 Play hours	Pass/Fail O/P
95	4	4	1
100	5	2	1 ←
95	2	7	0

Step 1 $z = x_1 w_1 + x_2 w_2 + x_3 w + b_1$

$z = \sum_{i=1}^{n} w_i^T x_i + b$

2: Activation(z)

$W_1 \; W_2 \; W_3$ [0.01, 0.02, 0.03] [0.001]

1. Forward Propogation

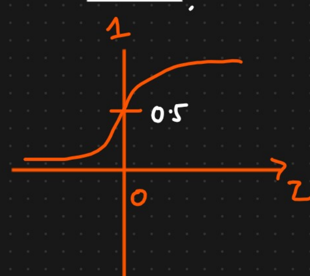
Hidden Layer 1

Step 1: $z = 95 \times 0.01 + 4 \times 0.02 + 4 \times 0.03 + 1 \times 0.01$

$= 1.151$

Step 2: Activation (z)

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



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

$O_1 = 0.759$

Hidden Layer 2

$W_4 = 0.02$

$b_2 = 0.03$

Step 1: $z = O_1 * W_4 + b_2$

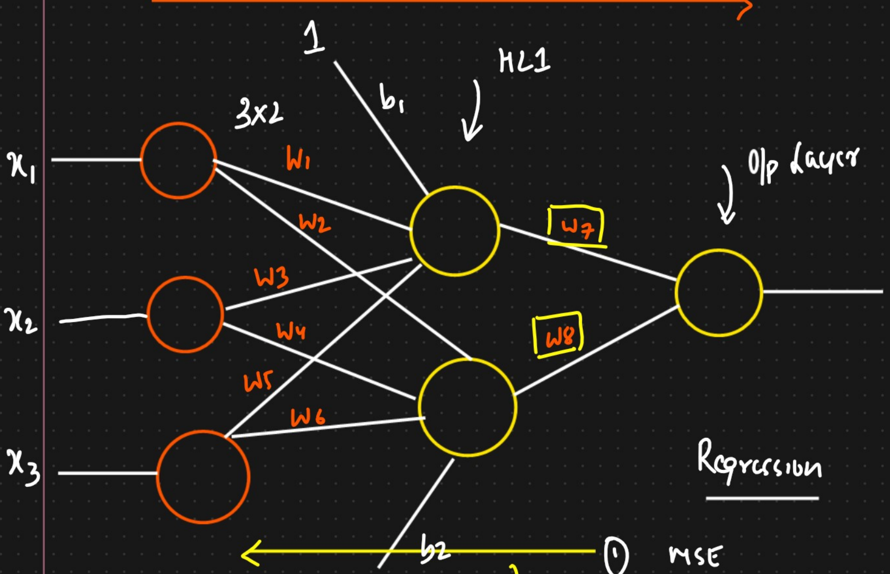
$= 0.759 * 0.02 + 0.03$

$= 0.04518$

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

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

⑥ Back Propogation And Weight Updation Formula



Backward Propogation

I/P features x_1 IQ	x_2 Study hours	x_3 Play hours	Pass/Fail O/p
95	4	4	1
100	5	2	1
95	2	7	0

Loss function

$(y-\hat{y})^2$ ↓↓↓↓↓↓

Regression

1. MSE
2. MAE
3. Huber Loss

Classification

1. Binary Cross Entropy
2. Categorical Cross Entropy

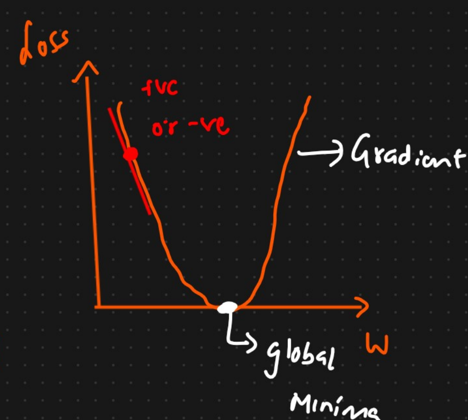
Weight updation Formula

$$W_{7_{new}} = W_{7_{old}} - \eta \frac{\partial L}{\partial W_{7_{old}}}$$

→ slope

$$W_{8_{new}} = W_{8_{old}} - \eta \frac{\partial L}{\partial W_{8_{old}}}$$

→ Learning Rate



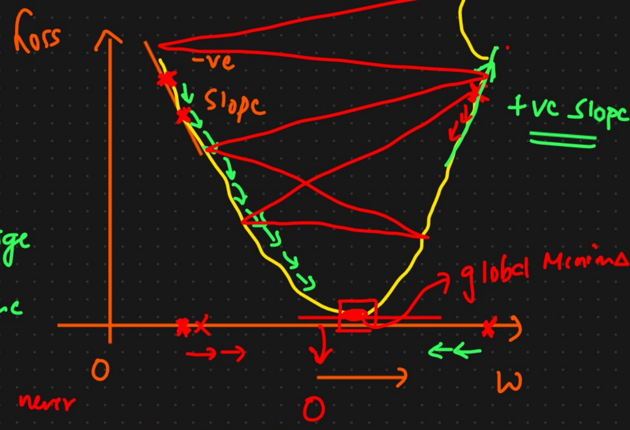
$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W_{old}}$$

⇒ Weight Updation Formula.

Gradient Descent

↗ Optimizers

Optimizers : To reduce the Loss value



η = large value

⇓

It may never converge

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

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

↳ Learning Rate

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

$\boxed{0.001}$ → small value.

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

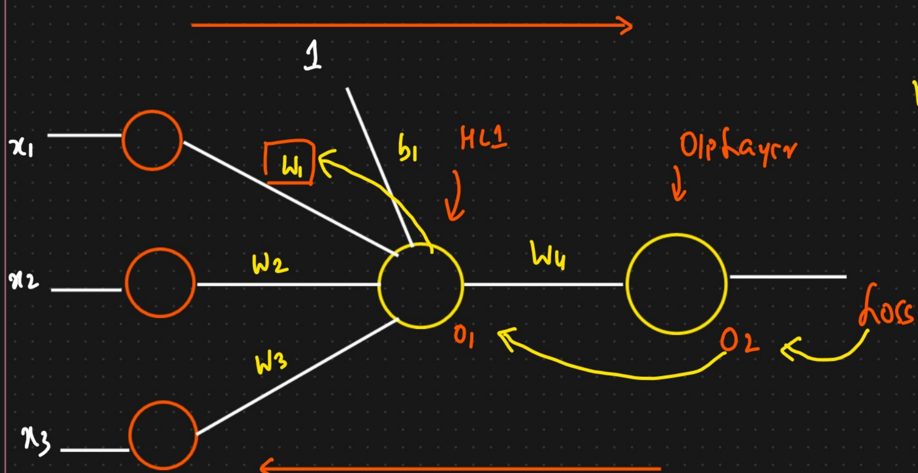
$$= W_{old} - \eta(+ve)$$

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

When W reaches global Minima

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

(7) Chain Rule of Derivative



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

$$W_{4_{new}} = W_{4_{old}} - \eta \boxed{\frac{\partial L}{\partial W_{4_{old}}}}$$

$$\frac{\partial L}{\partial W_{4old}} = \frac{\partial L}{\partial O_2} * \frac{\partial O_2}{\partial W_{4old}}$$ ⇒ Chain Rule of Derivation

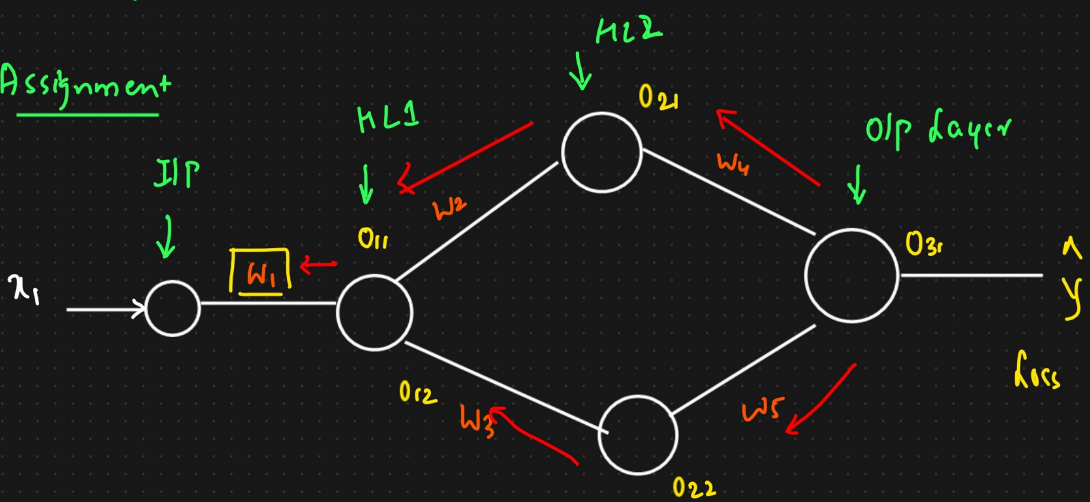
$$W_{1new} = W_{1old} - \eta \boxed{\frac{\partial L}{\partial W_{1old}}}$$

$$\boxed{\frac{\partial L}{\partial W_{1old}} = \frac{\partial L}{\partial O_2} * \frac{\partial O_2}{\partial O_1} * \frac{\partial O_1}{\partial W_{1old}}}$$

W_{2new}

W_{3new}.

Assignment



$$W_{1new} = W_{1old} - \eta \boxed{\frac{\partial L}{\partial W_{1old}}}$$

⇓

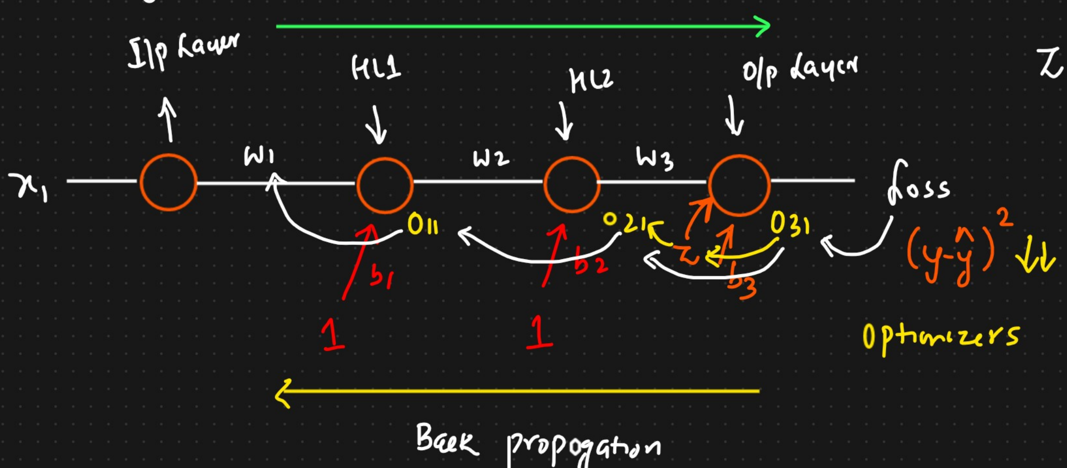
$$\frac{\partial L}{\partial W_{1old}} = \left[\frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial W_{1old}}\right]$$

$$+$$

$$\left[\frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{22}} * \frac{\partial O_{22}}{\partial O_{12}} * \frac{\partial O_{12}}{\partial W_{1old}}\right]$$

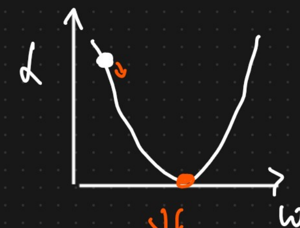
⑧ Vanishing Gradient Problem And Activation functions

$\phi(z)$



$$z = \sum_{i=1}^{n} w_i x_i + b$$

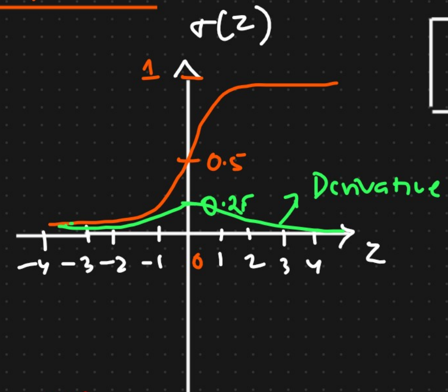
$\sigma(z) \Rightarrow 0$ to 1 → Sigmoid



⇓ Vanishing Gradient Problem.

⊛ Sigmoid Activation function

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



⇓

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

Derivative $(\sigma(z))$

⇓

$$0 \le \sigma'(z) \le 0.25$$

$$W_{1_{new}} = W_{1_{old}} - \eta \frac{\partial L}{\partial W_{old}}$$

η ↑ 0.01

$\frac{\partial L}{\partial W_{old}}$ ⇒ Small value

⇒ Small Value

⇒ $W_{1_{new}} \approx W_{1_{old}}$ ⇈ Equal

[ ] [0-0.25] * [0-0.25] * [0-0.25] ⇒ Smaller value ⇓ [0.000001]

$$\frac{\partial L}{\partial W_{1old}} = \frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial W_{11}}$$

$$O_{31} = \sigma(\underbrace{W_3 * O_{21} + b_3}_{\to z})$$

$z = W_3 * O_{21} + b_3$

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

$$\frac{\partial O_{31}}{\partial O_{21}} = \frac{\partial(\sigma(z))}{\partial(z)} * \frac{\partial z}{\partial O_{21}}$$

→ Sigmoid

{Chain Rule}.

$0 \le \sigma'(z) \le 0.25$ * $\dfrac{\partial((w_3 * O_{21}) + b_3)}{\partial(O_{21})}$

⇓

Derivative of Sigmoid

$$\frac{\partial O_{31}}{\partial O_{21}} = 0 \le \sigma'(z) \le 0.25 \quad * \quad w_{3_{old}}$$

- To fix this problem Researchers started exploring other Activation function

1. Tanh
2. Relu
3. Prelu
4. Swiss