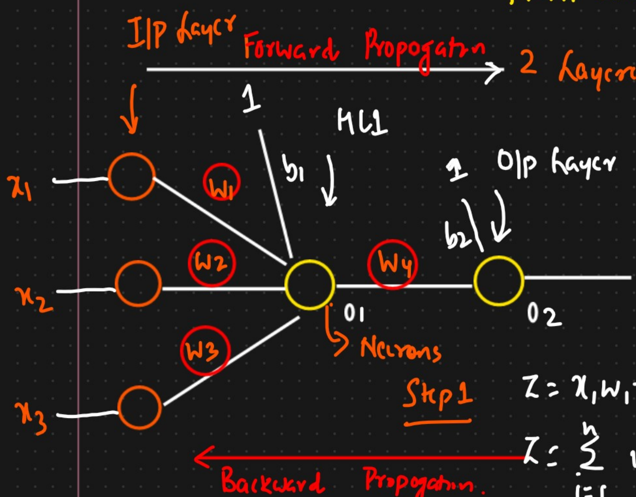
5 Multi Layered Perceptron Model [Artificial Neural N/W]

① Forward Propogation
② Backward Propogation → Geoffrey Hinton →

③ Loss function
④ Optimizers ✓
⑤ Activation function.

Multi Layered NN



Error

Loss function $= (1 - 0.5129)$

$\approx 0.49 \downarrow\downarrow\downarrow$

x_1	x_2	x_3	Pass/Fail
IQ	Study hours	Play Hours	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

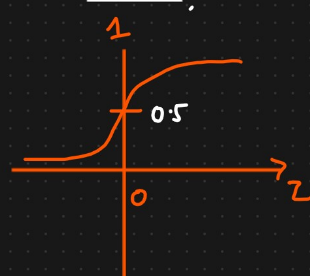
Hidden Layer 1

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

$= 1.151$

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

Step 2: Activation (z)



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

$O_1 = 0.759$

Hidden Layer 2

$w_4 = 0.02$

$b_2 = 0.03$

Step 1: $z = O_1 * w_4 + b_2$

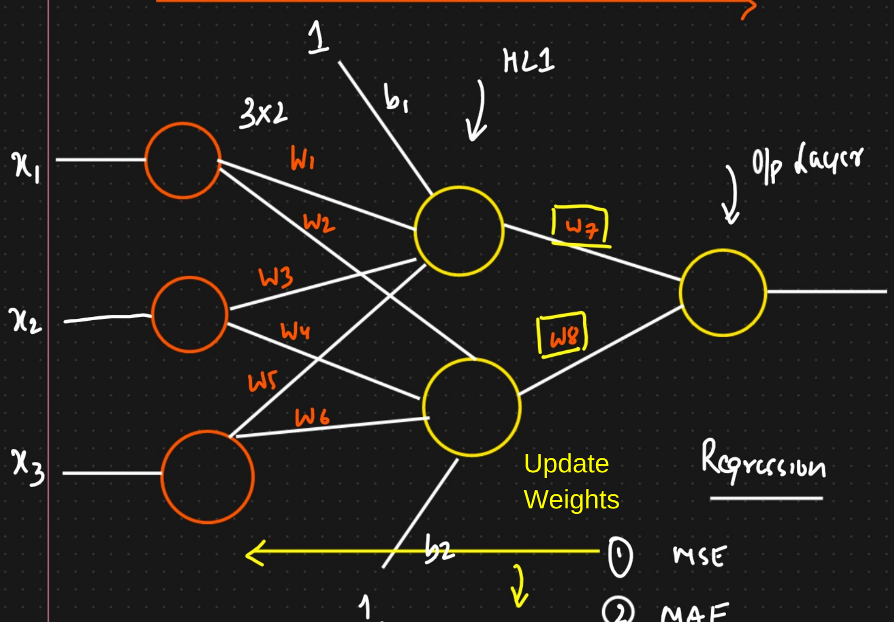
$= 0.759 * 0.02 + 0.03$

$= 0.04518$

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



I/P features

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

Loss function

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

Backward Propogation

Regression

① MSE
② MAE
③ Huber Loss

Classification

① Binary Cross Entropy
② Categorical Cross Entropy

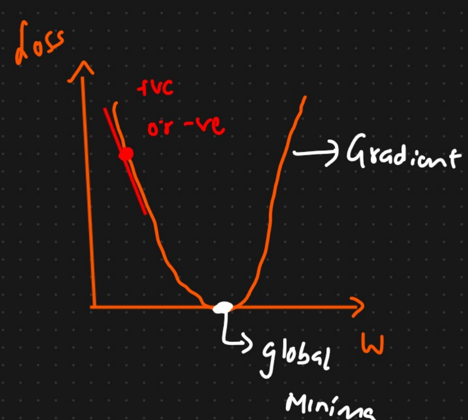
Weight updation Formula

$$W_{7new} = W_{7old} - \eta \left[\boxed{\frac{\partial L}{\partial W_{7old}}} \right]$$

slope

$$W_{8new} = W_{8old} - \eta \left[\frac{\partial L}{\partial W_{8old}} \right]$$

Learning Rate



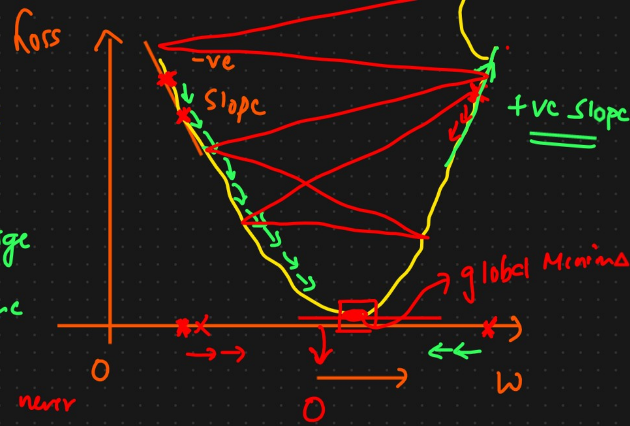
$$\boxed{W_{new} = W_{old} - \eta \left[\frac{\partial L}{\partial W_{old}} \right]}$$

$\Rightarrow$ Weight Updation Formula.

Gradient Descent

Optimizers

Optimizers : To reduce the loss value



η = large value

$\Downarrow$

It may never converge

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

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

η ↳ Learning Rate

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

$\boxed{0.001}$ → small value.

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

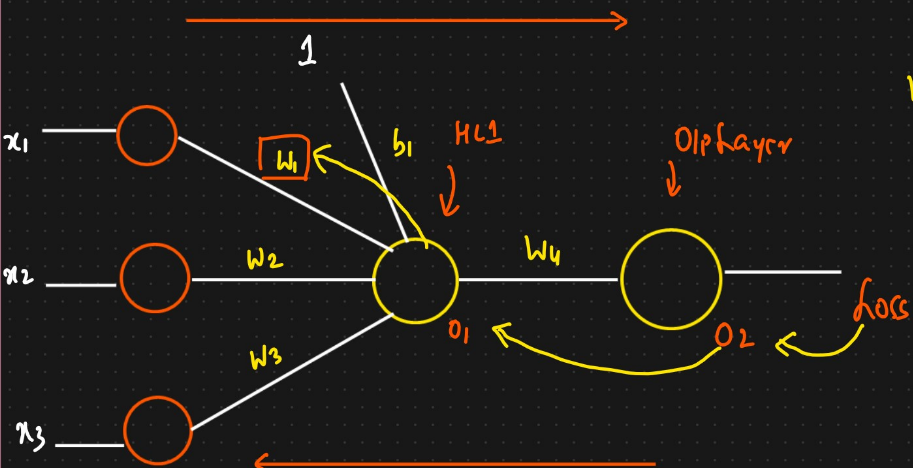
$$= W_{old} - \eta\,(+ve)$$

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

When W reaches global Minima

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

⑦ **Chain Rule of Derivative**



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

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

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

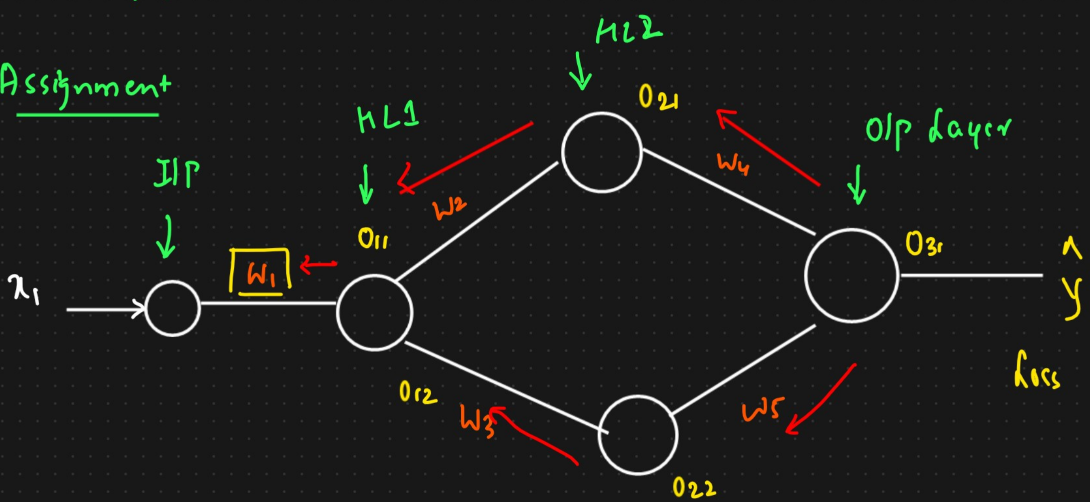
$$W_{1new} = W_{1old} - \eta \boxed{\frac{\partial L}{\partial W_{1old}}}$$

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

W_{2new}

W_{3new}.

Assignment



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

$$\Downarrow$$

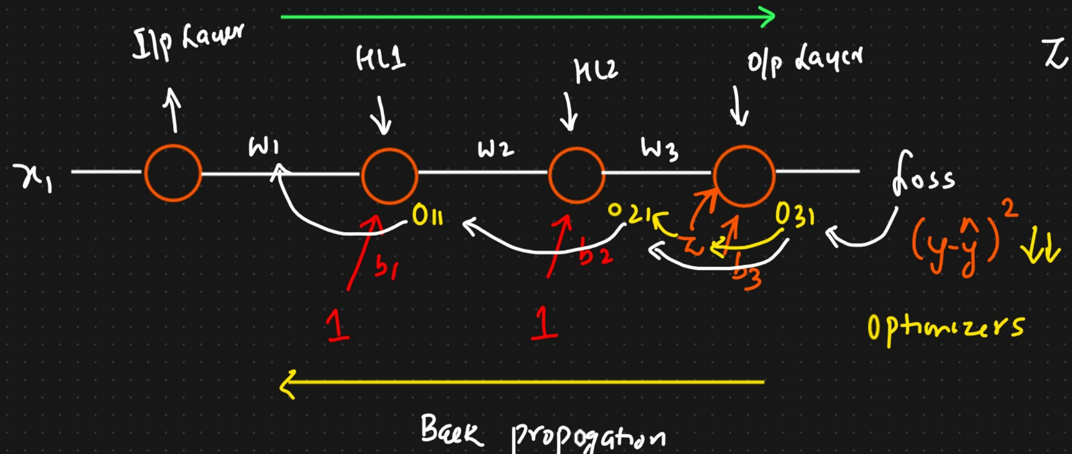
$$\frac{\partial L}{\partial W_{1old}} = \left[\frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial W_{1old}}\right]$$

$$+$$

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

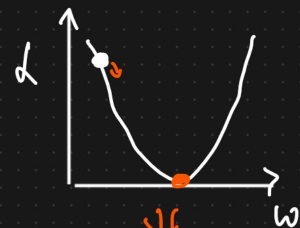
⑧ Vanishing Gradient Problem And Activation functions

$\phi(z)$



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

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

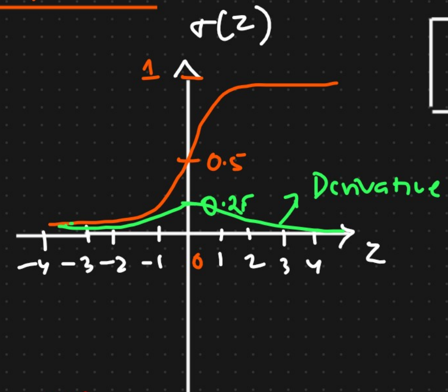


⇓

Vanishing Gradient Problem.

(*) Sigmoid Activation function

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



⇓

$$\boxed{0 \le \sigma(z) \le 1}$$

Derivative $(\sigma(z))$

⇓

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

0.01

$$W_{1\,new} = W_{1\,old} - \eta \boxed{\frac{\partial L}{\partial W_{1\,old}}} \Rightarrow \text{Small value} \Rightarrow \boxed{W_{1\,new} \approx W_{1\,old}}$$

Equal

⇒ Small Value

$\boxed{}$ $\boxed{0-0.25}$ * $\boxed{0-0.25}$ * $\boxed{0-0.25}$ ⇒ Smaller value ⇓ $\boxed{0.000001}$

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

$$O_{31} = \sigma(\underbrace{w_3 * O_{21} + b_3}_{\to z})$$

$z = w_3 * O_{21} + b_3$

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

$$\frac{\partial O_{31}}{\partial O_{21}} = \frac{\partial(\overbrace{\sigma(z)}^{\to \text{Sigmoid}})}{\partial(z)} * \frac{\partial z}{\partial O_{21}}$$

{Chain Rule}.

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

$\Downarrow$

Derivative of Sigmoid

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

* To fix this problem Researchers started exploring other Activation function

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