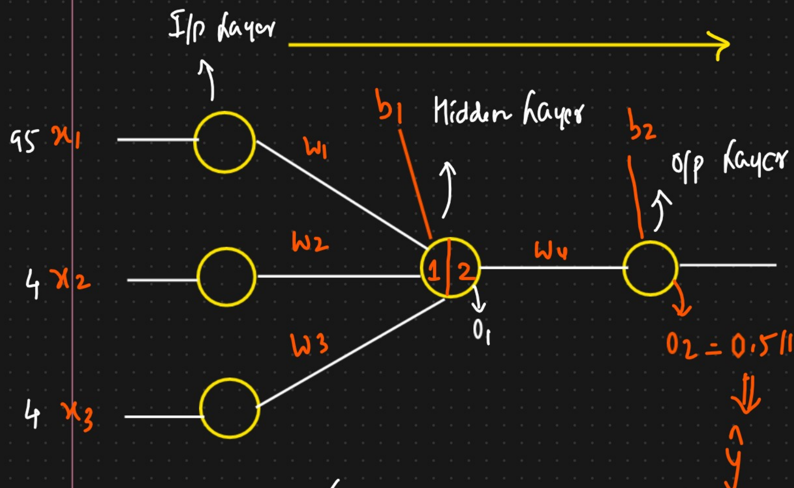
Multi Layered Perceptron Model [Artificial Neural N/W]

1. Forward Propogation
2. Backward Propogation → Geoffrey Hinton →
3. Activation Function
4. Optimizers
5. Loss functions

Dataset

x_1 IQ	x_2 Study Hours	x_3 Play Hours	O/P P/F
→ 95	4	4	1
→ 100	5	2	1
95	2	7	0

1. Forward Propogation In Multi Layered NN



Loss $[y - \hat{y}] \approx 0$

$w_1 \quad w_3 \quad b_1$

$[0.01, 0.02, 0.03] \quad [0.001]$

Step 1

$Z = x_1 w_1 + x_2 w_2 + x_3 w_3 + b_1$

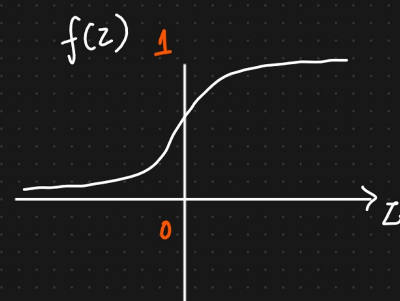
$$\boxed{Z = \sum_{i=1}^{w} w_i^T x_i + b}$$

$Z = 95 \times 0.01 + 4 \times 0.02 + 4 \times 0.03 + 0.01$

$Z = 1.151$

Step 2 [Activation Function] ⇒ 0 to 1

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



$f(z) = \frac{1}{1+e^{-1.151}} = 0.759 = O_1$

O/p Layer

$w_4 = [0.02] \quad b_2 = 0.03$

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

$$= 0.759 * 0.02 + 0.03$$

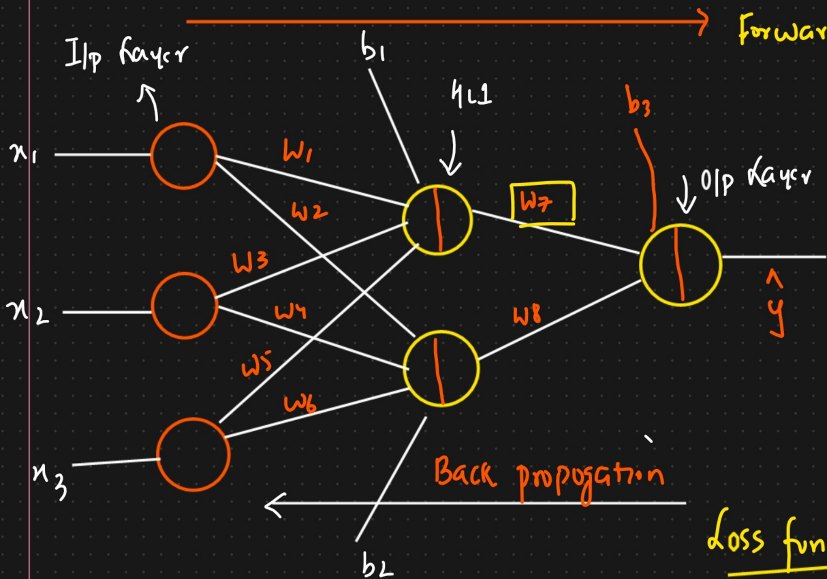
$$= 0.04518$$

Step 2 Activation(z) $\Rightarrow \dfrac{1}{1+e^{-(0.04518)}} = 0.511$

$O_2 = 0.511$

② Backward Propogation And Weight Updation Formula

Forward Propogation



Simple Linear Regression

Error

$$\text{Loss}\left[(y-\hat{y})^2\right] \downarrow\downarrow \approx 0$$

$\Downarrow$

Optimizer

Loss function

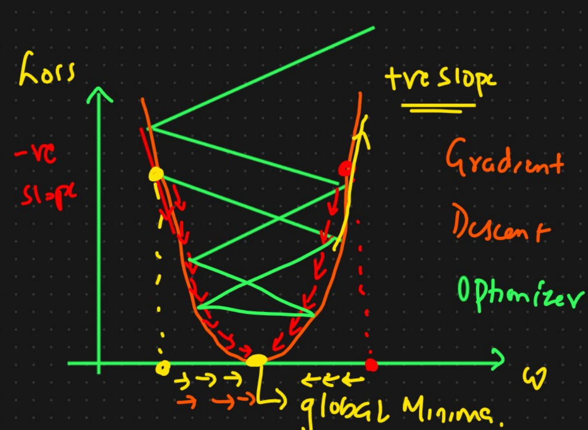
Regression	Classification
① MSE	① Binary Cross Entropy
② MAE	② Categorical Cross Entropy
③ Huber Loss	

Weight Updation Formula – Back Propogation

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

Slope

Learning Rate



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

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

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

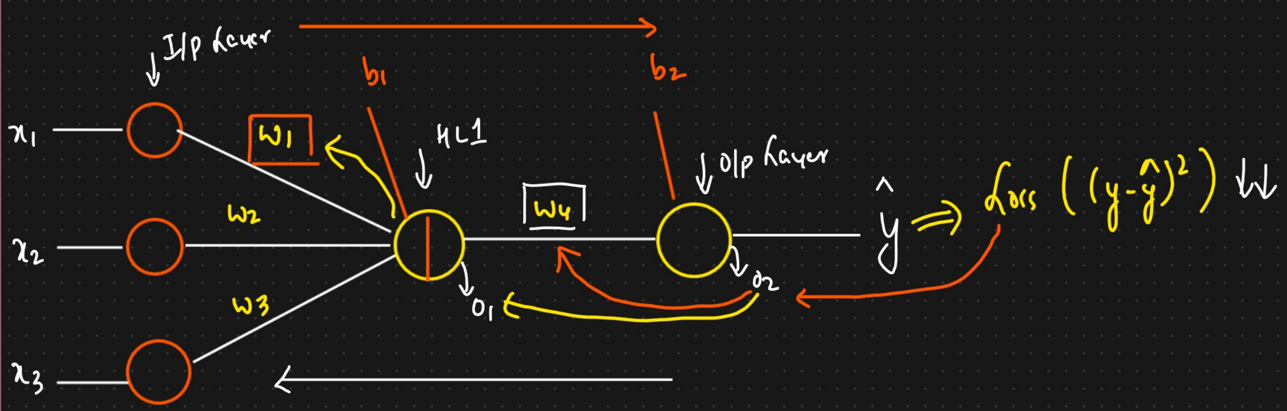
$W_{new} = W_{old} + (+ve)$

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

When W reaches global Minima

$\boxed{W_{new} = W_{old}}$ $\Rightarrow$ loss ↓↓ minimal

③ Chain Rule of Derivative



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

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

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

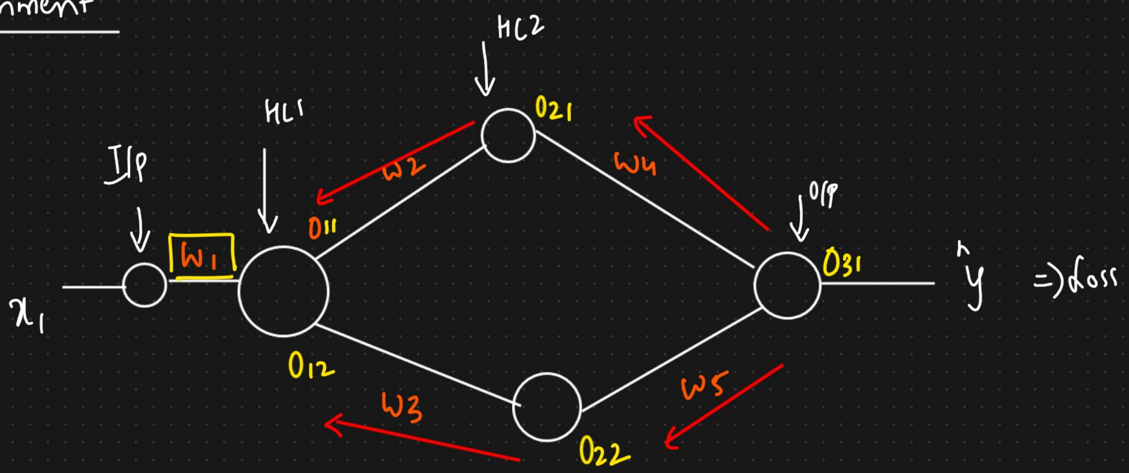
$$\boxed{b_{2new} = b_{2old} - \eta \frac{\partial L}{\partial b_{2old}}}$$

$$\frac{\partial L}{\partial W_{4old}} = \frac{\partial L}{\partial O_2} * \frac{\partial O_2}{\partial W_{4old}}$$ {Chain Rule}.

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

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

Assignment



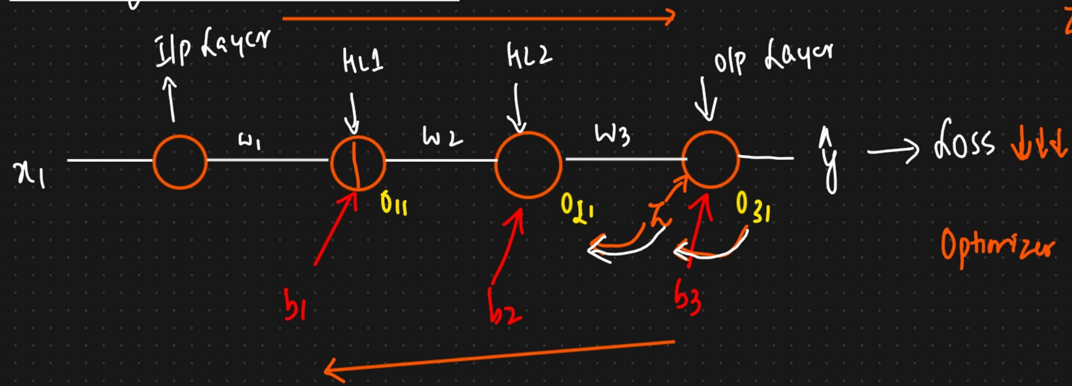
$$w_{1new} = w_{1old} - \eta \boxed{\frac{\partial L}{\partial w_{1old}}}$$

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

$$+$$

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

④ Vanishing Gradient Problem



Back propogation

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

$\sigma(z) = 0$ to 1.

⇓

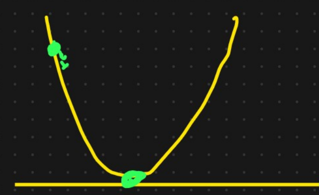
Sigmoid

⊛ Sigmoid Activation Function

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

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

$\partial(\sigma(z)) = 0$ to 0.25



0.01

⇑

Learning Rate

$\boxed{0.01}$

$$W_{1new} = W_{1old} - \eta \left\{ \boxed{\frac{\partial L}{\partial W_{old}}} \right. \Rightarrow \boxed{W_{1new} \approx W_{1old}} \Leftarrow$$

0.25 * 0.05 * 0.05 * 0.0025 ≈ Small value

$$\frac{\partial L}{\partial W_{1old}} = \frac{\partial L}{\partial O_{31}} * \boxed{\frac{\partial O_{31}}{\partial O_{21}}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial W_{11}} \Rightarrow \text{Chain Rule}$$

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

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

$$\frac{\partial O_{31}}{\partial O_{21}} = \frac{\partial(\sigma(z))}{\partial z} * \frac{\partial z}{\partial O_{21}} \quad \{\text{Chain Rule}\}.$$

⇒ Sigmoid

$$= \underline{\underline{0 \text{ to } 0.25}} * \frac{\partial((w_3 * O_{21}) + b_3)}{\partial O_{21}}$$

$$\boxed{\frac{\partial O_{31}}{\partial O_{21}} = 0 \text{ to } 0.25 * w_3} \Rightarrow \text{Small value}$$

Different Activation

1. Tanh
2. Relu
3. Parametric Relu