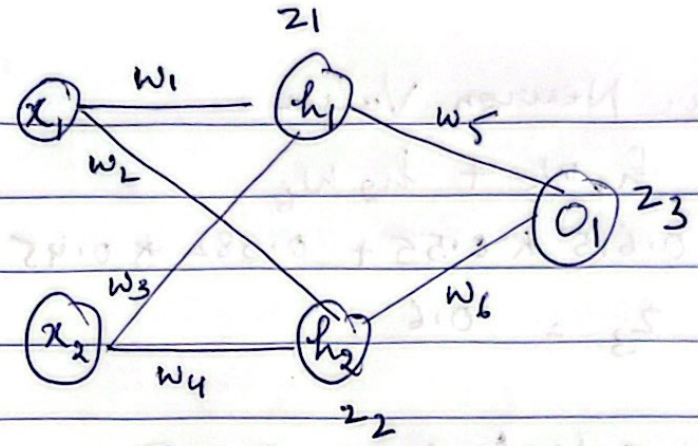

Sigmoid Function as activation function

MSE = loss function

learning Rate = 0.1

Initial weights

$W_1 = 0.7$, $W_2 = 0.3$, $W_3 = 0.4$, $W_4 = 0.6$, $W_5 = 0.55$

$W_6 = 0.45$

$x_1 = 0.5$ $x_2 = 0.3$

Feed Forward Step :

$$z_1 = x_1 w_1 + x_2 w_3 = 0.5 \times 0.7 + 0.3 \times 0.4 = 0.47$$

hidden layer node value $h_1 = \text{sigmoid}(z_1) = \dfrac{1}{1+e^{-0.47}}$

$$h_1 = 0.615$$

$$z_2 = x_1 w_2 + x_2 w_4 = 0.5 \times 0.3 + 0.3 \times 0.6 = 0.33$$

$$h_2 = \text{sigmoid}(z_2)$$

$$= \frac{1}{1+e^{-0.33}} = 0.582$$

Output layer Neuron Value.

$$z_3 = h_1 w_5 + h_2 w_6.$$
$$= 0.615 * 0.55 + 0.582 * 0.45$$
$$z_3 = 0.6$$

$$\therefore O_1 = \text{Sigmoid}(z_3)$$
$$= \frac{1}{1+e^{-0.6}} = 0.645$$

Calculating loss :-

$$\text{MSE}, \; C = \frac{1}{1} \sum_{i}^{1} (y_1 - \hat{y}_1)^2$$

Since Only one output neuron. therefore Only 1 in denominator

$$\left\{ C = \frac{1}{n} \sum_{i}^{n} (y_i - \hat{y}_i)^2 \right.$$

$$C = \frac{1}{1} \sum_{i}^{1} (1 - 0.645)^2$$

$$C = 0.126$$

Back propagation :- (1st layer).

Updating the weight of w_5

New weight = current weight + $(-\nabla C) \times$ learning rate

$$= \text{current weight} + \left(-\frac{\partial C}{\partial w_5}\right) \times \text{learning rate}$$

$$\frac{\partial c}{\partial w_5} = \frac{\partial c}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_5} \quad \{\text{chain rule}\}$$

$$c = (y - o_1)^2$$

$$\frac{\partial c}{\partial o_1} = \frac{\partial}{\partial o_1}(y - o_1)^2$$

$$= -2(y - o_1)$$

$$\frac{\partial c}{\partial o_1} = -2(1 - 0.645) = -0.71$$

$$\frac{\partial o_1}{\partial z_3} = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) \quad \{\text{derivative of Sigmoid function}\}$$

$$= \sigma(z)(1 - \sigma(z))$$

$$\frac{\partial o_1}{\partial z_3} = 0.645(1 - 0.645) = 0.229$$

$$z_3 = h_1 w_5 + h_2 w_6$$

$$\frac{\partial z_3}{\partial w_5} = h_1 = 0.615$$

$$\frac{\partial c}{\partial w_5} = -0.71 * 0.229 * 0.615 = -0.01$$

New weight w_5 = $0.55 + 0.01 * 0.1$

$= 0.551$

$$w_5^* = 0.551$$

New weight of w_6.

$$\frac{\partial C}{\partial w_6} = \frac{\partial C}{\partial o_1} * \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_6}$$

$z_3 = h_1 w_5 + h_2 w_6 = h_2 = 0.582$

$$\frac{\partial z_3}{\partial w_6} = -0.71 * 0.229 * 0.582 = -0.095$$

New weight of w_6

$$w_6^* = 0.45 + 0.095 \times 0.1$$
$$= 0.4595$$

Now we move to 2nd layer.

New weight of w_1 = Current weight + $(-\nabla C) \times$ learning rate

w_1^*

$$\frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial o_1} * \frac{\partial o_1}{\partial z_3} * \frac{\partial z_3}{\partial h_1} * \frac{\partial h_1}{\partial z_1} * \frac{\partial z_1}{\partial w_1}$$

$z_3 = h_1 w_5 + h_2 w_6$

$$\frac{\partial z_3}{\partial h_1} = w_5 = 0.55$$

$$\frac{\partial h_1}{\partial z_1} = \frac{\partial}{\partial z_1}\left(\frac{1}{1+e^{-h_1}}\right) \quad \left\{ \because \frac{d\sigma(x)}{d(x)} = \sigma(x) \cdot (1-\sigma(x)) \right\}$$

$$\frac{\partial h_1}{\partial z_1} = 0.615(1-0.615) = 0.237$$

$$z_1 = x_1 w_1 + x_2 w_3$$

$$\frac{\partial z_1}{\partial w_1} = x_1 = 0.5$$

$$\frac{\partial C}{\partial w_1} = -0.71 * 0.229 * 0.55 * 0.237 * 0.5$$

$$\frac{\partial C}{\partial w_1} = -0.0106$$

$$w_1^* = 0.7 + 0.0106 \times 0.1$$

$$w_1^* = 0.7010$$

New weight of w_2

$$w_2^* = w_2 + \frac{\partial C}{\partial w_2} \times 0.1$$

$$\frac{\partial C}{\partial w_2} = \frac{\partial C}{\partial o_1} * \frac{\partial o_1}{\partial z_3} * \frac{\partial z_3}{\partial h_2} * \frac{\partial h_2}{\partial z_2} * \frac{\partial z_2}{\partial w_2}$$

$$\frac{\partial z_3}{\partial h_2} = \frac{\partial}{\partial h_2}(h_1 w_5 + h_2 w_6) = w_6 = 0.45$$

$$\frac{\partial h_2}{\partial z_2} = h_2(1-h_2) = 0.243$$

$$\frac{\partial z_2}{\partial w_2} = \frac{\partial}{\partial w_2}(x_1 w_2 + x_2 w_4) = x_1 = 0.5$$

$$\frac{\partial C}{\partial w_2} = -0.71 * 0.229 * 0.45 * 0.243 * 0.5 = -0.009$$

$w_2^* = 0.30009.$

New weight of w_3, w_3^*

$$w_3^* = w_3 + (-\nabla C)\, 0.1$$

$$\frac{\partial C}{\partial w_3} = \frac{\partial C}{\partial o_1} * \frac{\partial o_1}{\partial z_3} * \frac{\partial z_3}{\partial h_1} * \frac{\partial h_1}{\partial z_1} * \frac{\partial z_1}{\partial w_3}$$

$$\frac{\partial z_3}{\partial h_1} = \frac{\partial}{\partial h_1}\left(h_1 w_5 + h_2 w_6\right) = w_5 = 0.55$$

$$\frac{\partial h_1}{\partial z_1} = \frac{\partial}{\partial z_1}\left(\frac{1}{1+e^{-z_1}}\right) = h_1(1-h_1) = 0.237$$

$$\frac{\partial z_1}{\partial w_3} = \frac{\partial}{\partial w_3}\left(x_1 w_1 + x_2 w_3\right) = x_2 = 0.5$$

$$\frac{\partial C}{\partial w_3} = -0.71 * 0.229 * 0.55 * 0.237 * 0.5 = -0.011$$

$$\Rightarrow w_3^* = 0.4 + 0.011 * 0.1 = 0.4011$$

New weight of w_4, w_4^*

$$w_4^* = w_4 + (-\nabla C)\, 0.1.$$

$$\frac{\partial C}{\partial w_4} = \frac{\partial C}{\partial o_1} \frac{\partial o_1}{\partial z_3} \frac{\partial z_3}{\partial h_2} \frac{\partial h_2}{\partial z_2} \frac{\partial z_2}{\partial w_4}.$$

$$\frac{\partial z_3}{\partial h_2} = \frac{\partial}{\partial h_2}(h_1 w_5 + h_2 w_6) = w_6 = 0.45$$

$$\frac{\partial h_2}{\partial z_2} = h_2(1-h_2) = 0.243$$

$$\frac{\partial z_2}{\partial w_4} = \frac{\partial}{\partial w_4}(x_1 w_2 + x_2 w_4) = x_2 = 0.3$$

$$\Rightarrow \frac{\partial c}{\partial w_4} = -0.71 * 0.229 * 0.45 * 0.243 * 0.3 = -0.05$$

$$\Rightarrow w_4^* = 0.6 + 0.05 \times 0.1 = 0.6005$$

Check if Neural Network error has reduced

$$z_1 = x_1 w_1^* + x_2 w_3^* = 0.5 * 0.7010 + 0.3 * 0.4011$$

$$z_1 = 0.4708$$

$$z_2 = x_1 w_2^* + x_2 w_4^* = 0.5 * 0.3009 + 0.3 * 0.6005$$

$$z_2 = 0.3306$$

$$h_1 = \text{sigmoid}(z_1) = \frac{1}{1+e^{-0.4708}} = 0.6156$$

$$h_2 = \text{sigmoid}(z_2) = \frac{1}{1+e^{-0.3306}} = 0.5819$$

$$z_3 = h_1 w_5^* + h_2 w_6^*$$

$$z_3 = 0.6156 + 0.551 + 0.5819 \times 0.4595 = 0.6066$$

$$O_1 = \text{sigmoid}(z_3) = \frac{1}{1+e^{-0.6066}}$$

$$O_1 = 0.6472$$

$$\text{loss} = C = \frac{1}{1}\sum_{i}^{1}(y_1-\hat{y}_1)^2$$

$$= (1-0.6472)^2$$

$$C = 0.1245$$

Which is lower then previous loss of 0.126.
We continue to repeat same steps until we reach minima value for loss.