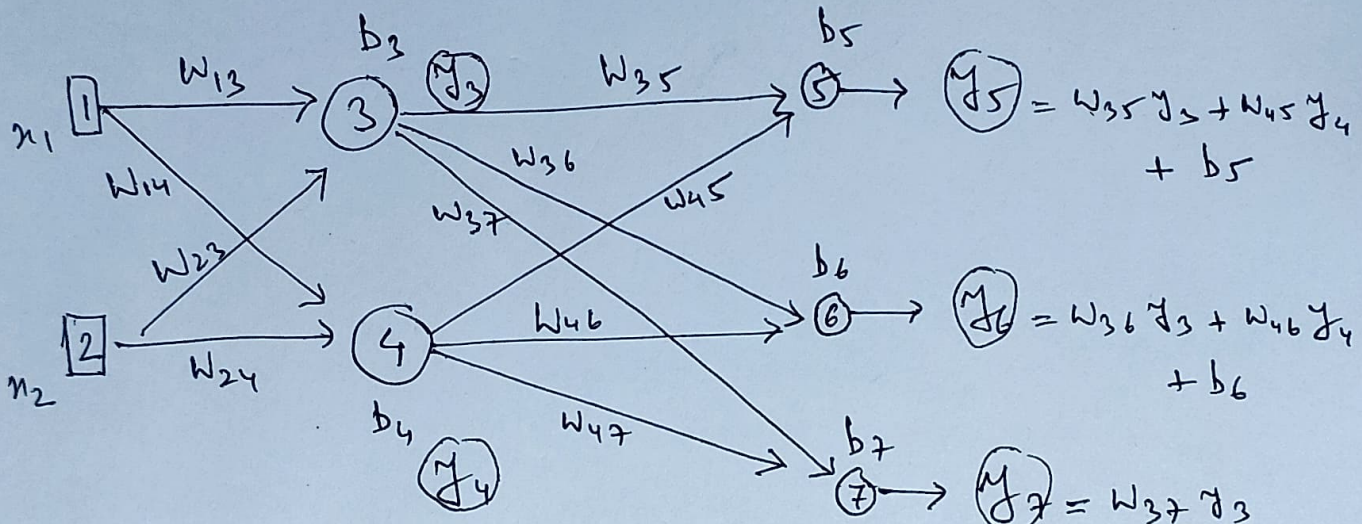
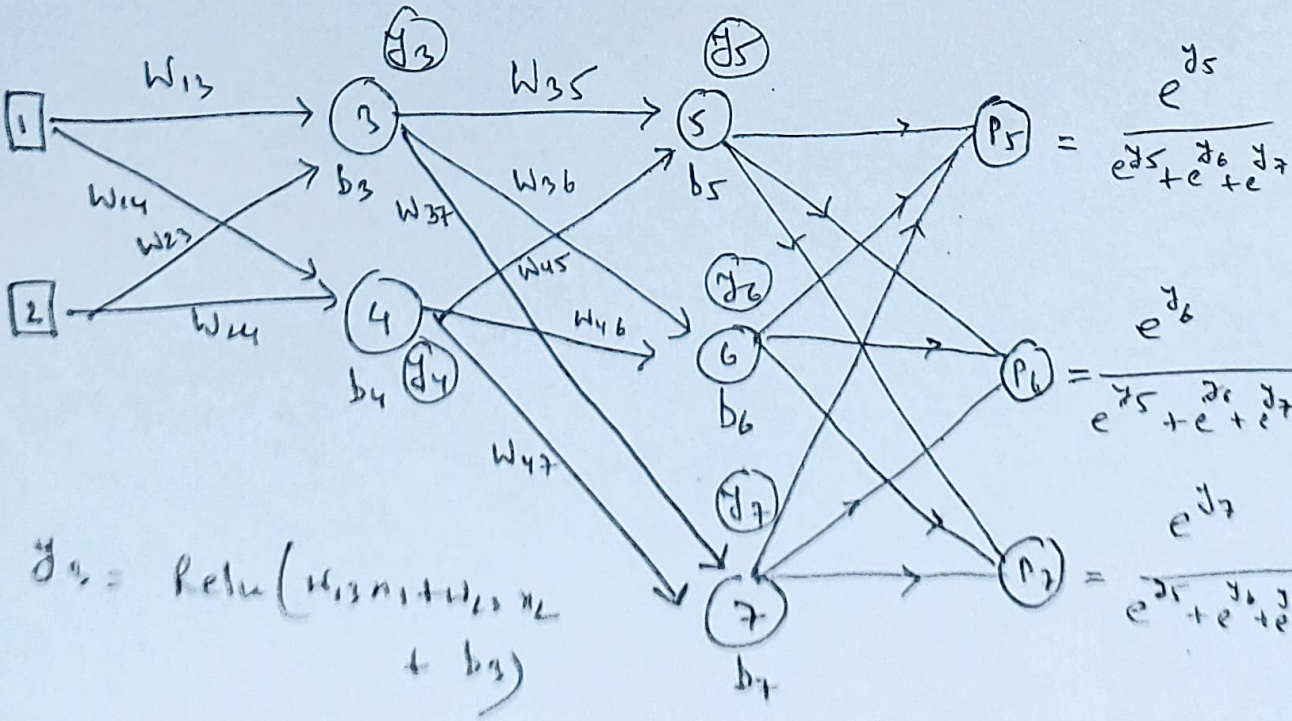


$$y_3 = \text{Relu}(W_{13} x_1 + W_{23} x_2 + b_3)$$

$$y_4 = \text{Relu}(W_{14} x_1 + W_{24} x_2 + b_4)$$



$$y_3 = \text{Relu}(W_{13} x_1 + W_{23} x_2 + b_3)$$

$$y_4 = \text{Relu}(W_{14} x_1 + W_{24} x_2 + b_4)$$

$$y_5 = W_{35} y_3 + W_{45} y_4 + b_5$$

$$y_6 = W_{36} y_3 + W_{46} y_4 + b_6$$

$$y_7 = W_{37} y_3 + W_{47} y_4 + b_7$$

Parameter Update 1

Chowdhury Mofizur Rahman

April 22, 2025

At first start with the output nodes. Define loss function and differentiate loss function with respect to all the output nodes.

$$Loss = \frac{1}{2}\{(Ob_5 - Y_5)^2 + (Ob_6 - Y_6)^2 + (Ob_7 - Y_7)^2\}$$
$$\frac{dLoss}{dY_5} = \frac{1}{2}.2.(Ob_5 - Y_5)(-1) = -Err_5$$
$$\frac{dLoss}{dY_6} = \frac{1}{2}.2.(Ob_6 - Y_6)(-1) = -Err_6$$
$$\frac{dLoss}{dY_7} = \frac{1}{2}.2.(Ob_7 - Y_7)(-1) = -Err_7$$

Write the expressions for output nodes

$$Y_5 = W_{35}Y_3 + W_{45}Y_4 + b_5$$
$$Y_6 = W_{36}Y_3 + W_{46}Y_4 + b_6$$
$$Y_7 = W_{37}Y_3 + W_{47}Y_4 + b_7$$

Now you can differentiate loss function with respect to all the output layer parameters $W_{35}, W_{45}, W_{36}, W_{46}, W_{37}, W_{47}, b_5, b_6, b_7$ using chain rule

$$\frac{dLoss}{dW_{35}} = \frac{dloss}{dY_5}.\frac{dY_5}{dW_{35}} = -Err_5.Y_3$$
$$\frac{dLoss}{dW_{45}} = \frac{dloss}{dY_5}.\frac{dY_5}{dW_{45}} = -Err_5.Y_4$$
$$\frac{dLoss}{dW_{36}} = \frac{dloss}{dY_5}.\frac{dY_5}{dW_{36}} = -Err_6.Y_3$$
$$\frac{dLoss}{dW_{46}} = \frac{dloss}{dY_6}.\frac{dY_6}{dW_{46}} = -Err_6.Y_4$$
$$\frac{dLoss}{dW_{37}} = \frac{dloss}{dY_7}.\frac{dY_7}{dW_{37}} = -Err_7.Y_3$$
$$\frac{dLoss}{dW_{47}} = \frac{dloss}{dY_7}.\frac{dY_7}{dW_{47}} = -Err_7.Y_4$$
$$\frac{dLoss}{db_5} = \frac{dloss}{dY_5}.\frac{dY_5}{db_5} = -Err_5$$
$$\frac{dLoss}{db_6} = \frac{dloss}{dY_6}.\frac{dY_6}{db_6} = -Err_6$$
$$\frac{dLoss}{db_7} = \frac{dloss}{dY_7}.\frac{dY_7}{db_7} = -Err_7$$

Now using gradient descent approach the update equations for all the output layer parameters are:

$$W_{35} = W_{35} + \eta Y_3.Err_5$$
$$W_{45} = W_{45} + \eta Y_4.Err_5$$

$$W_{36} = W_{36} + \eta Y_3.Err_6$$
$$W_{46} = W_{46} + \eta Y_4.Err_6$$

$$W_{37} = W_{37} + \eta Y_3.Err_7$$
$$W_{47} = W_{47} + \eta Y_4.Err_7$$

$$b_5 = b_5 + \eta Err_5$$
$$b_6 = b_6 + \eta Err_6$$
$$b_7 = b_7 + \eta Err_7$$

Now consider the hidden nodes 3 and 4. Write the expressions for Y_3 and Y_4.

$$Y_3 = Relu\{W_{13}.X_1 + W_{23}.X_2 + b_3\}$$
$$Y_4 = Relu\{W_{14}.X_1 + W_{24}.X_2 + b_4\}$$

Now differentiate the Loss function with respect to hidden outputs Y_3 and Y_4. Both nodes 3 and 4 have parents 5, 6 and 7. Following the chain rule

$$\frac{dLoss}{dY_3} = \frac{dloss}{dY_5}.\frac{dY_5}{dY_3} + \frac{dloss}{dY_6}.\frac{dY_6}{dY_3} + \frac{dloss}{dY_7}.\frac{dY_7}{dY_3} = (-Err_5).W_{35} + (-Err_6).W_{36} + (-Err_7).W_{37}$$

$$\frac{dLoss}{dY_4} = \frac{dloss}{dY_5}.\frac{dY_5}{dY_4} + \frac{dloss}{dY_6}.\frac{dY_6}{dY_4} + \frac{dloss}{dY_7}.\frac{dY_7}{dY_4} = (-Err_5).W_{45} + (-Err_6).W_{46} + (-Err_7).W_{47}$$

Now we can differentiate Loss function with respect to all the hidden layer parameters W_{13}, W_{23}, W_{14}, W_{24}, b_3 and b_4

$$\frac{dLoss}{dW_{13}} = \frac{dLoss}{dY_3}.\frac{dY_3}{dW_{13}} = \{(-Err_5).W_{35}+(-Err_6).W_{36}+(-Err_7).W_{37}\}.Relu^{'}(W_{13}.X_1+W_{23}.X_2+b_3).X_1$$

$$\frac{dLoss}{dW_{23}} = \frac{dLoss}{dY_3}.\frac{dY_3}{dW_{23}} = \{(-Err_5).W_{35}+(-Err_6).W_{36}+(-Err_7).W_{37}\}.Relu^{'}(W_{13}.X_1+W_{23}.X_2+b_3).X_1$$

$$\frac{dLoss}{dW_{14}} = \frac{dLoss}{dY_4}.\frac{dY_4}{dW_{14}} = \{(-Err_5).W_{45}+(-Err_6).W_{46}+(-Err_7).W_{47}\}.Relu^{'}(W_{14}.X_1+W_{24}.X_2+b_4).X_1$$

$$\frac{dLoss}{dW_{24}} = \frac{dLoss}{dY_4}.\frac{dY_4}{dW_{24}} = \{(-Err_5).W_{45}+(-Err_6).W_{46}+(-Err_7).W_{47}\}.Relu^{'}(W_{14}.X_1+W_{24}.X_2+b_4).X_2$$

$$\frac{dLoss}{db_3} = \frac{dLoss}{dY_3}.\frac{dY_3}{db_3} = \{(-Err_5).W_{35}+(-Err_6).W_{36}+(-Err_7).W_{37}\}.Relu^{'}(W_{13}.X_1+W_{23}.X_2+b_3)$$

$$\frac{dLoss}{db_4} = \frac{dLoss}{dY_4}.\frac{dY_4}{db_4} = \{(-Err_5).W_{45}+(-Err_6).W_{46}+(-Err_7).W_{47}\}.Relu^{'}(W_{14}.X_1+W_{24}.X_2+b_4)$$

Therefore the updated equations for all the hidden layer parameters are as follows:

$$W_{13} = W_{13}+\eta.(Err_5.W_{35}+Err_6.W_{36}+Err_7.W_{37}\}.Relu^{'}(W_{13}.X_1+W_{23}.X_2+b_3).X_1$$

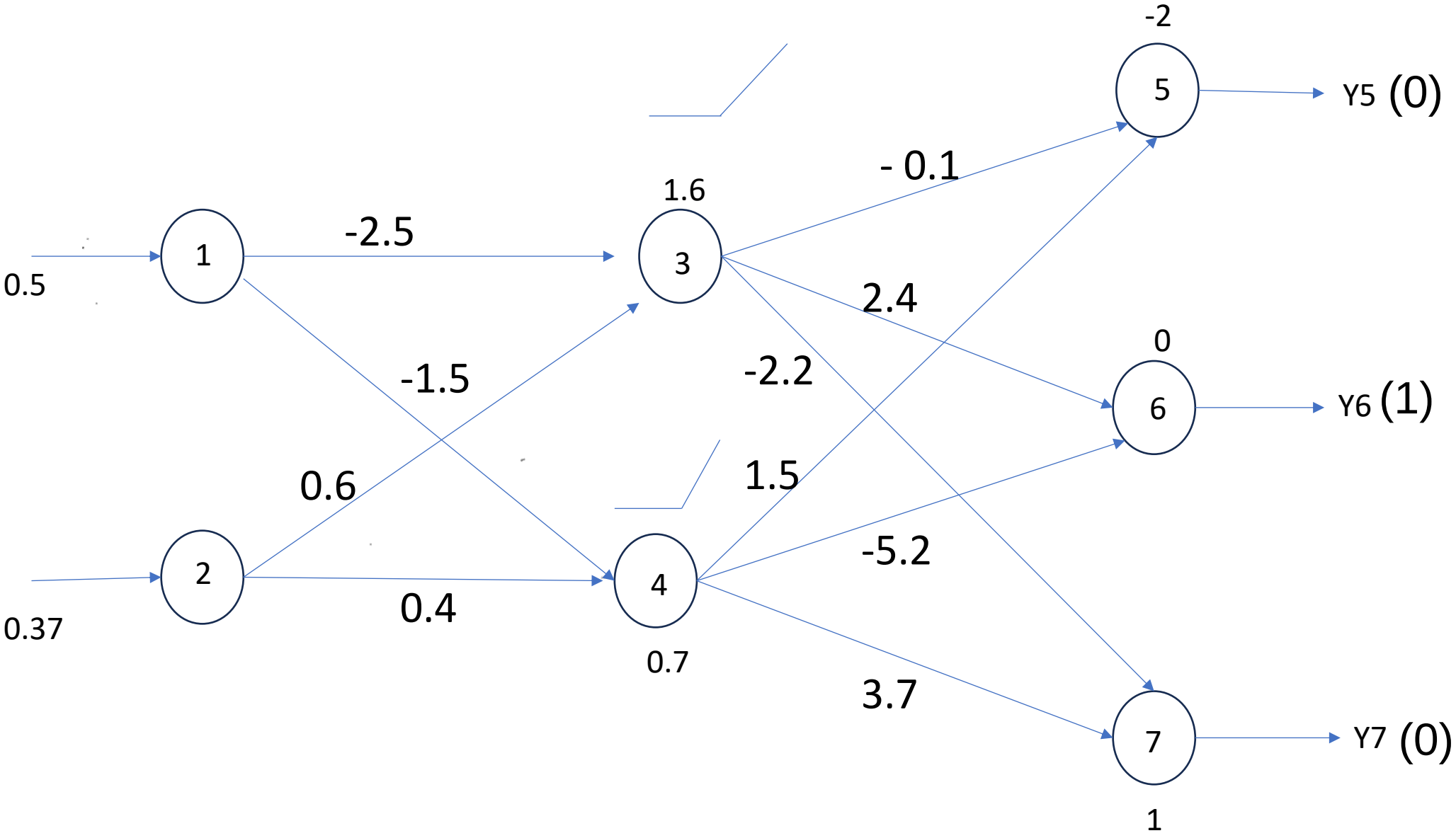
$$W_{23} = W_{23}+\eta.(Err_5.W_{35}+Err_6.W_{36}+Err_7.W_{37}).Relu^{'}(W_{13}.X_1+W_{23}.X_2+b_3).X_2$$

$$W_{14} = W_{14}+\eta.(Err_5.W_{45}+Err_6.W_{46}+Err_7.W_{47}).Relu^{'}(W_{14}.X_1+W_{24}.X_2+b_4).X_1$$

$$W_{24} = dW_{24}+\eta.(Err_5.W_{45}+Err_6.W_{46}+Err_7.W_{47}).Relu^{'}(W_{14}.X_1+W_{24}.X_2+b_4).X_2$$

$$b_3 = b_3 + \eta.(Err_5.W_{35} + Err_6.W_{36} + Err_7.W_{37}).Relu^{'}(W_{13}.X_1 + W_{23}.X_2 + b_3)$$

$$b_4 = b_4 + \eta.(Err_5.W_{45} + Err_6.W_{46} + Err_7.W_{47}).Relu^{'}(W_{14}.X_1 + W_{24}.X_2 + b_4)$$


-2
5
Y5 (0)
- 0.1
1.6
1
-2.5
3
0.5
2.4
0
-1.5
-2.2
6
Y6 (1)
0.6
1.5
2
4
-5.2
0.37
0.4
0.7
3.7
7
Y7 (0)
1

Y3 = Relu(0.5x(-2.5) + 0.6x0.37 + 1.6) = 0.57

Y4 = Relu(0.5x(-1.5) + 0.37x0.4 + 0.7) = 0.1

Y5 = 0.57x(-0.1) + 0.1x1.5 = -1.9

Y6 = 2.4x0.57 + 0.1x(-5.2) + 0 = 0.85

Y7 = 0.57x(-2.2)+ 0.1x3.7 + 1 = 0.12

Err5 = 1.9 Err6 = 0.15 Err6 = -0.12

W35 = -0.1+0.1x0.57x1.9 = 0.01

W46 = -5.2+0.1x0.1x0.15 = -5.19

b5 = -2 +0.1x1.9 = -1.81

b7 = 1+0.1x(-0.12) = 0.99

W13 = -2.5 + 0.1x((-0.1)x1.9+2.4x0.15+(-2.2)x(-0.12)) x Relu'(0.57) x 0.5 = -2.28

W24 = 0.4 + 0.1x(1.5x1.9+(-5.2)x0.15 + 3.7x(-0.12))x Relu'(0.1)x0.37 = 1.0