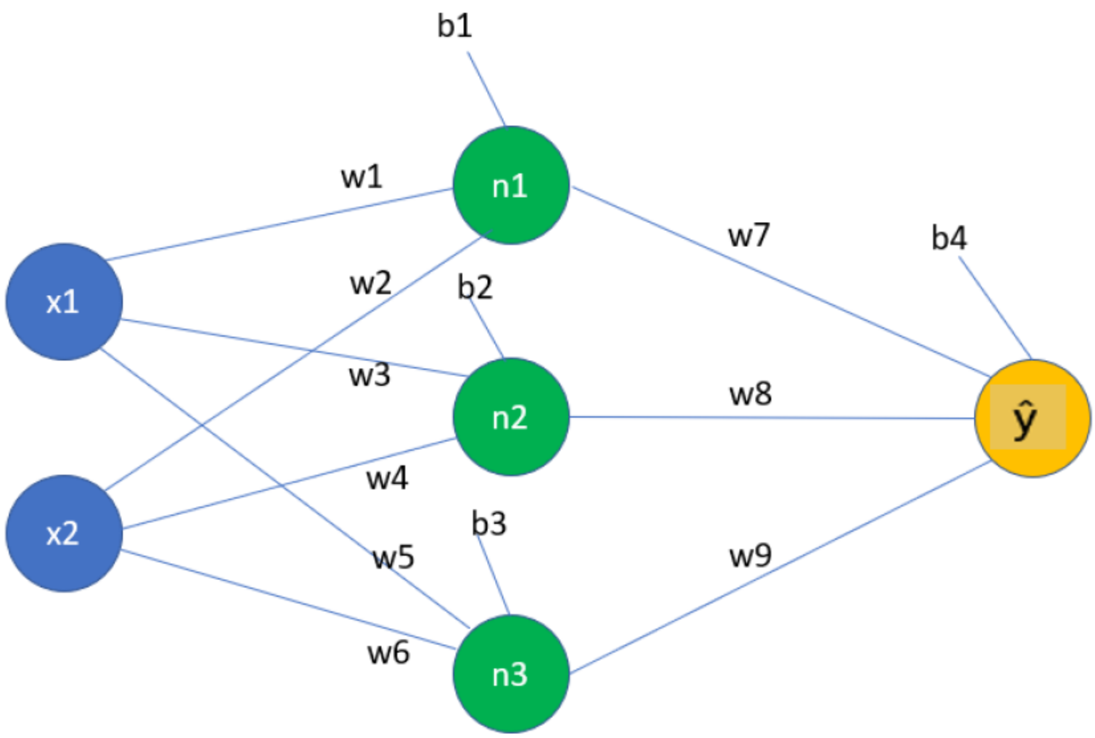
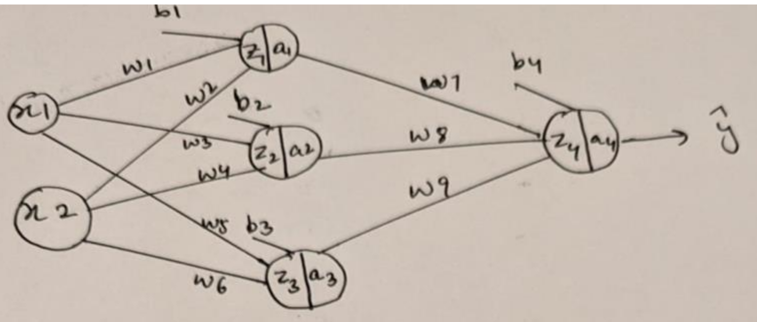
Problem 1 (**8 pts**): The figure below shows a 2-layer, feed-forward neural network with three hidden-layer nodes and one output node. x1 and x2 are the two inputs. For the following questions, assume the learning rate is α = 0.1; activation function = sigmoid; loss function, MSE = $\frac{1}{2}(y - \hat{y})^2$; target output $y = 1$; For instance, the output of n1 equal $sigmoid(w1 * x1 + w2 * x2 \; + b1)$. Compute one step of the backpropagation.

a. Assume x1 = 1, x2 = 0; all weights and biases equal 1. Compute the updated weights for both the hidden layer and output layer. Show all steps in your calculations. (**4 pts**)



Ans: For Easy understanding, I have rounded off the updated weights till 4 or 5 decimal place.

Q1)
a)



Forward Pass

$$z_1 = w_1 x_1 + w_2 x_2 + b_1$$
$$= (1\times1) + (1\times0) + 1$$
$$= 2$$
$$a_1 = \sigma(2) = \frac{1}{1+e^{-2}} = 0.881$$

$$z_2 = w_3 x_1 + w_4 x_2 + b_2$$
$$= (1\times1) + (1\times0) + 1$$
$$= 2$$
$$a_2 = \sigma(2) = \frac{1}{1+e^{-2}} = 0.881$$

$$z_3 = w_5 x_1 + w_6 x_2 + b_3$$
$$= (1\times1) + (1\times0) + 1$$
$$= 2$$
$$a_3 = \sigma(2) = \frac{1}{1+e^{-2}} = 0.881$$

$$z_4 = w_7 a_1 + w_8 a_2 + w_9 a_3 + b_4$$
$$= (0.881\times1) + (0.881\times1) + (0.881\times1) + 1$$
$$= 3.643$$
$$a_4 = \sigma(3.643) = \frac{1}{1+e^{-3.643}} = 0.974$$

For Backpropagation:
Following paths are followed to updated respective weights
For w1: w1 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error
For w2: w2 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error
For w3: w3 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error
For w4: w4 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error
For w5: w5 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error

For w6: w6 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error
For w7: w7----> z4 ------ > a4 ----→ Error
For w8: w8 ----> z4 ------ > a4 ----→ Error
For w9: w9 ----> z4 ------ > a4 ----→ Error
For b1: b1 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error
For b2: b2 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error
For b3: b3 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error
For b4: b4 ----> z4 ------ > a4 ----→ Error

~~Erro~~

Error:

Loss function $= MSE = \frac{1}{2}(y - \hat{y})^2$

$MSE = \frac{1}{2}(1 - 0.974)^2 = \underline{0.00034}$

$\underline{MSE = 0.00034}$

Back propagation

→ for b_4:

$$\frac{\partial E}{\partial b_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial b_4}$$

$$\frac{\partial E}{\partial a_4} = \frac{\partial}{\partial a_4}\left(\tfrac{1}{2}(y - \hat{y})^2\right) = -(y - \hat{y})$$

$$= -(1 - 0.974)$$

$$\boxed{\frac{\partial E}{\partial a_4} = -0.0260}$$

$$\frac{\partial a_4}{\partial z_4} = \frac{\partial}{\partial z_4}\left(1/1+e^{-x}\right) = \sigma(1-\sigma)$$

$$= 0.974 \times (1 - 0.974)$$

$$\boxed{\frac{\partial a_4}{\partial z_4} = 0.024}$$

$$\frac{\partial z_4}{\partial b_4} = \frac{\partial}{\partial b_4}(w_7 a_1 + w_8 a_2 + w_9 a_3 + b_4)$$
$$= 1$$

$$\frac{\partial E}{\partial b_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial b_4}$$
$$= -0.0260 \times 0.024 \times 1$$

$$\frac{\partial E}{\partial b_4} = -0.0006$$

Taking $\eta = 0.1$ (learning rate)

$$b_{4new} = b_4 - \left(\eta \times \frac{\partial E}{\partial b_4}\right)$$
$$= 1 - (0.1 \times -0.0006)$$

$$\boxed{b_{4new} = 1.00006}$$

→ for w_7:

$$\frac{\partial E}{\partial w_7} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4}} \times \frac{\partial z_4}{\partial w_7}$$

this we calculated earlier

$$\frac{\partial z_4}{\partial w_7} = \frac{\partial (a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)}{\partial w_7}$$
$$= a_1$$
$$= 0.881$$

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_7}$$
$$= -0.0260 \times 0.024 \times 0.881$$

$\partial E / \partial w_7 = -0.0005$

$$w_{7new} = w_7 - \eta \times \frac{\partial E}{\partial w_7}$$

$$= 1 - (0.1 \times -0.0005)$$

$$\boxed{w_{7new} = 1.00005}$$

for w_8:

$$\frac{\partial E}{\partial w_8} = \underbrace{\boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4}}}_{\text{this we already know}} \times \frac{\partial z_4}{\partial w_8}$$

$$\frac{\partial z_4}{\partial w_8} = \frac{\partial (a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)}{\partial w_8}$$

$$= a_2 = 0.881$$

$$\frac{\partial E}{\partial w_8} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_8}$$

$$= -0.0260 \times 0.024 \times 0.881$$

$$\frac{\partial E}{\partial w_8} = -0.0005$$

Taking η (learning rate) $= 0.1$

$$w_{8new} = w_8 - \left(\eta \times \frac{\partial E}{\partial w_8}\right)$$

$$= 1 - (0.1 \times -0.0005)$$

$$\boxed{w_{8new} = 1.00005}$$

for w_9:

$$\frac{\partial E}{\partial w_9} = \underbrace{\boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4}}}_{\text{this we already know}} \times \frac{\partial z_4}{\partial w_9}$$

$$\frac{\partial z_4}{\partial w_9} = \frac{\partial}{\partial w_9}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= a_3 = 0.881$$

$$\frac{\partial E}{\partial w_9} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_9}$$

$$= -0.0260 \times 0.024 \times 0.881$$

$$\frac{\partial E}{\partial w_9} = -0.0005$$

Taking η (learning rate) $= 0.1$

$$w_{9new} = w_9 - \left(\eta \times \frac{\partial E}{\partial w_9}\right)$$

$$= 1 - (0.1 \times -0.0005)$$

$$\boxed{w_{9new} = 1.00005}$$

→ for w_1:

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

we already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.0260 \; ; \; \frac{\partial a_4}{\partial z_4} = 0.024$$

$$\frac{\partial z_4}{\partial a_1} = \frac{\partial}{\partial a_1}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= w_7 = 1$$

$$\boxed{\frac{\partial z_4}{\partial a_1} = 1}$$

$$\frac{\partial a_1}{\partial z_1} = \frac{\partial (1/1+e^{-x})}{\partial z_1} = \sigma(1-\sigma)$$

$$= 0.881 \times (1 - 0.88)$$

$$\boxed{\frac{\partial a_1}{\partial z_1} = 0.105}$$

$$\frac{\partial z_1}{\partial w_1} = \frac{\partial}{\partial w_1}(x_1 w_1 + x_2 w_2 + b_1)$$

$$= x_1 = 1$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\frac{\partial E}{\partial w_1} = -6.552 \times 10^{-5}$$

Taking $\eta = 0.1$

$$w_{1new} = w_1 - \left(\eta \times \frac{\partial E}{\partial w_1}\right)$$

$$= 1 - (0.1 \times (-6.552 \times 10^{-5}))$$

$$\boxed{w_{1new} = 1.0000065}$$

for w_2:

$$\frac{\partial E}{\partial w_2} = \underbrace{\boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1}}}_{\text{this we already know}} \times \frac{\partial z_1}{\partial w_2}$$

$$\frac{\partial z_1}{\partial w_2} = \frac{\partial}{\partial w_2}(x_1 w_1 + x_2 w_2 + b_1)$$

$$= x_2 = 0$$

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_2}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 0$$

$$\frac{\partial E}{\partial w_2} = 0$$

Taking η (learning rate) $= 0.1$

$$w_{2new} = w_2 - \left(\eta \times \frac{\partial E}{\partial w_2}\right)$$

$$= 1 - (0.1 \times 0)$$

$$\boxed{w_{2new} = 1}$$

for b1:

$$\frac{\partial E}{\partial b_1} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1}} \times \frac{\partial z_1}{\partial b_1}$$

this we already know

$$\frac{\partial z_1}{\partial b_1} = \frac{\partial}{\partial b_1}(x_1 w_1 + x_2 w_2 + b_1)$$

$$= 1$$

$$\frac{\partial E}{\partial b_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\boxed{\frac{\partial E}{\partial b_1} = -6.552 \times 10^{-5}}$$

Taking $\eta = 0.1$ (learning rate)

$b_1 new = b_1 - (\eta \times \frac{\partial E}{\partial b_1})$

$= 1 - (0.1 \times (-6.552 \times 10^{-5}))$

$$\boxed{b_1 new = 1.0000065}$$

→ for w_3

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_3}$$

we already know the value of

$\frac{\partial E}{\partial a_4} = -0.0260$; $\frac{\partial a_4}{\partial z_4} = 0.024$

$$\frac{\partial z_4}{\partial a_2} = \frac{\partial}{\partial a_2}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$= w_8$

$= 1$

$$\boxed{\frac{\partial z_4}{\partial a_2} = 1}$$

$$\frac{\partial a_2}{\partial z_2} = \frac{\partial}{\partial z_2}(1/1+e^{-x}) = \sigma(1-\sigma)$$

$= 0.881 \times (1 - 0.881)$

$= 0.105$

$$\frac{\partial z_2}{\partial w_3} = \frac{\partial}{\partial w_3}(x_1 w_3 + x_2 w_4 + b_2)$$

$= x_1 = 1$

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_3}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\frac{\partial E}{\partial w_3} = -6.552 \times 10^{-5}$$

Taking η (learning rate) $= 0.1$

$$w_{3new} = w_3 - \left(\eta \times \frac{\partial E}{\partial w_3}\right)$$

$$= 1 - (0.1 \times -6.552 \times 10^{-5})$$

$$\boxed{w_{3new} = 1.0000065}$$

for w_4:

$$\frac{\partial E}{\partial w_4} = \underbrace{\boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2}}}_{\text{this we already know}} \times \frac{\partial z_2}{\partial w_4}$$

$$\frac{\partial z_2}{\partial w_4} = \frac{\partial}{\partial w_4}(x_1 w_3 + x_2 w_4 + b_2)$$

$$= \cancel{x_4} x_2 = 0$$

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_4}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 0$$

$$\frac{\partial E}{\partial w_4} = 0$$

Taking $\eta = 0.1$

$$w_{4new} = w_4 - \left(\eta \times \frac{\partial E}{\partial w_4}\right)$$

$$= 1 - (0.1 \times 0)$$

$$\boxed{w_{4new} = 1}$$

→ for b_2:

$$\frac{\partial E}{\partial b_2} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2}} \times \frac{\partial z_2}{\partial b_2}$$

this we already know

$$\frac{\partial z_2}{\partial b_2} = \frac{\partial}{\partial b_2}(x_1 w_3 + x_2 w_4 + b_2)$$

$$= 1$$

$$\frac{\partial E}{\partial b_2} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_2}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\frac{\partial E}{\partial b_2} = -6.552 \times 10^{-5}$$

Taking η (learning rate) $= 0.1$

$$b_{2new} = b_2 - \left(\eta \times \frac{\partial E}{\partial b_2}\right)$$

$$= 0.1 - (0.1 \times -6.552 \times 10^{-5})$$

$$\boxed{b_{2new} = 1.0000065}$$

→ for w_5:

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_5}$$

we already know the value of;

$$\frac{\partial E}{\partial a_4} = -0.0260 \quad ; \quad \frac{\partial a_4}{\partial z_4} = 0.024$$

$$\frac{\partial z_4}{\partial a_3} = \frac{\partial}{\partial a_3}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= w_9 = 1$$

$$\boxed{\frac{\partial z_4}{\partial a_3} = 1}$$

$$\frac{\partial a_3}{\partial z_3} = \frac{\partial}{\partial z_3}(1/1+e^{-x}) = \sigma(1-\sigma)$$

$$=(1-0.881) \times (0.881)$$

$$\boxed{\frac{\partial a_3}{\partial z_3} = 0.105}$$

$$\frac{\partial z_3}{\partial w_5} = \frac{\partial}{\partial w_5}(x_1 w_5 + x_2 w_6 + b_3)$$

$$= x_1 = 1$$

$$\boxed{\frac{\partial z_3}{\partial w_5} = 1}$$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_5}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\frac{\partial E}{\partial w_5} = -6.552 \times 10^{-5}$$

Taking η (learning rate) = 0.1

$$w_{5new} = w_5 - \left(\eta \times \frac{\partial E}{\partial w_5}\right)$$

$$= 1 - (0.1 \times -6.552 \times 10^{-5})$$

$$\boxed{w_{5new} = 1.0000065}$$

→ For w_6:

$$\frac{\partial E}{\partial w_6} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3}} \times \frac{\partial z_3}{\partial w_6}$$

this we already know

$$\frac{\partial z_3}{\partial w_6} = \frac{\partial}{\partial w_6}(x_1 w_5 + x_2 w_6 + b_3)$$

$$= x_2 = 0$$

$$\frac{\partial E}{\partial w_6} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_6}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 0$$

$$\boxed{\frac{\partial E}{\partial w_6} = 0}$$

Taking η (learning rate) $= 0.1$

$$w_{6new} = w_6 - \left(\eta \times \frac{\partial E}{\partial w_6}\right)$$

$$= 1 - (0.1 \times 0)$$

$$\boxed{w_{6new} = 1}$$

For b_3:

$$\frac{\partial E}{\partial b_3} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3}} \times \frac{\partial z_3}{\partial b_3}$$

this we already know

$$\frac{\partial z_3}{\partial b_3} = \frac{\partial (x_1 w_5 + x_2 w_6 + b_3)}{\partial b_3}$$

$$= 1$$

$$\frac{\partial E}{\partial b_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial b_3}$$

$$= -0.0260 \times 0.024 \times 1 \times 0.105 \times 1$$

$$\frac{\partial E}{\partial b_3} = -6.552 \times 10^{-5}$$

Taking $\eta = 0.1$

$$b_{3new} = b_3 - \left(\eta \times \frac{\partial E}{\partial b_3}\right)$$

$$= 1 - (0.1 \times -6.552 \times 10^{-5})$$

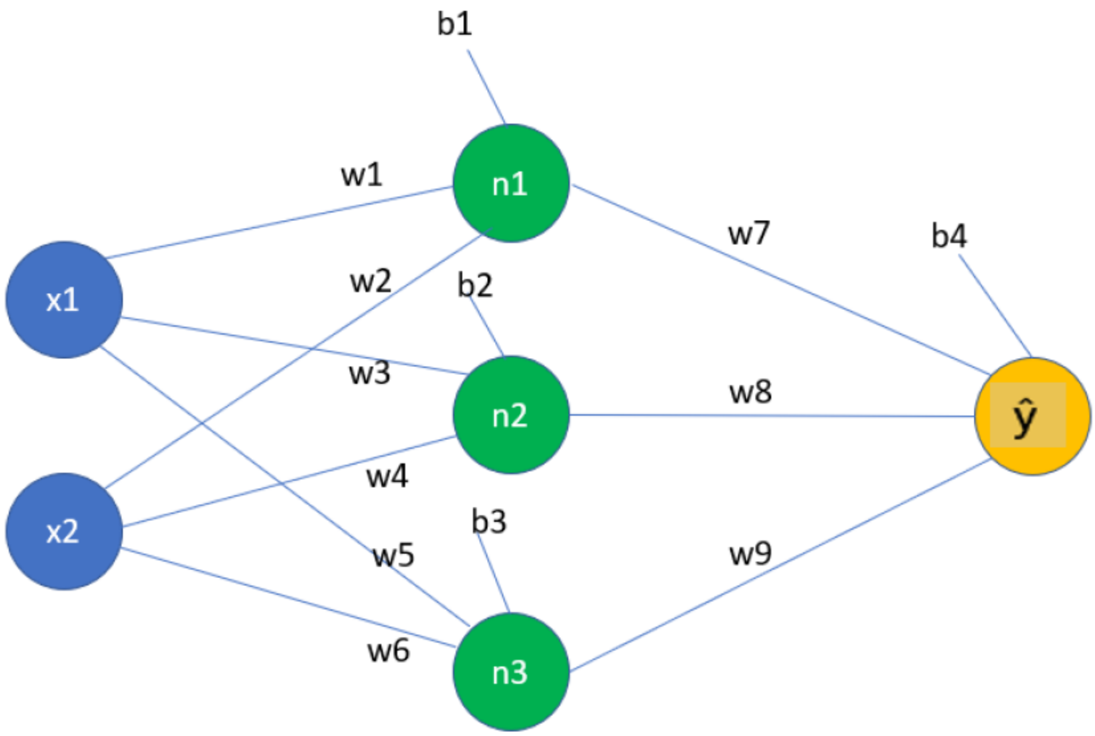
$$\boxed{b_{3new} = 1.0000065}$$

Therefore updated weights after one iteration

Old weight	New weight
$w_1 = 1$	$w_1 = 1.0000065$
$w_2 = 1$	$w_2 = 1$
$w_3 = 1$	$w_3 = 1.0000065$
$w_4 = 1$	$w_4 = 1$
$w_5 = 1$	$w_5 = 1.0000065$
$w_6 = 1$	$w_6 = 1$
$w_7 = 1$	$w_7 = 1.00005$
$w_8 = 1$	$w_8 = 1.00005$
$w_9 = 1$	$w_9 = 1.00005$
$b_1 = 1$	$b_1 = 1.0000065$
$b_2 = 1$	$b_2 = 1.0000065$
$b_3 = 1$	$b_3 = 1.0000065$
$b_4 = 1$	$b_4 = 1.00006$

Hence, from these updated weights we can see that when weight and bias were initialised as 1, it will lead to vanishing gradients in back-propagation. Due to this, the model will have difficulty in learning and will take time to converge or may not converge at all. Therefore, it is recommended to take weight initialization techniques such as Xavier or glorot distribution to avoid vanishing gradient.

b. Assume $x1 = 1$, $x2 = 0$; $w1 = w3 = w5 = 0.5$; $w2 = w4 = w6 = -0.25$; $w7= 1$, $w8 = -1$, $w9 =0$ and biases weights equal 0.1. Compute the updated weights for both the hidden layer and output layer. Show all steps in your calculations. (**4 pts**)



Ans: <u>For Easy understanding, I have rounded off the updated weights till 4 or 5 decimal place.</u>

For Backpropagation:

Following paths are followed to updated respective weights

For w1: w1 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error

For w2: w2 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error

For w3: w3 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error

For w4: w4 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error

For w5: w5 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error

For w6: w6 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error

For w7: w7----> z4 ------ > a4 ----→ Error

For w8: w8 ----> z4 ------ > a4 ----→ Error

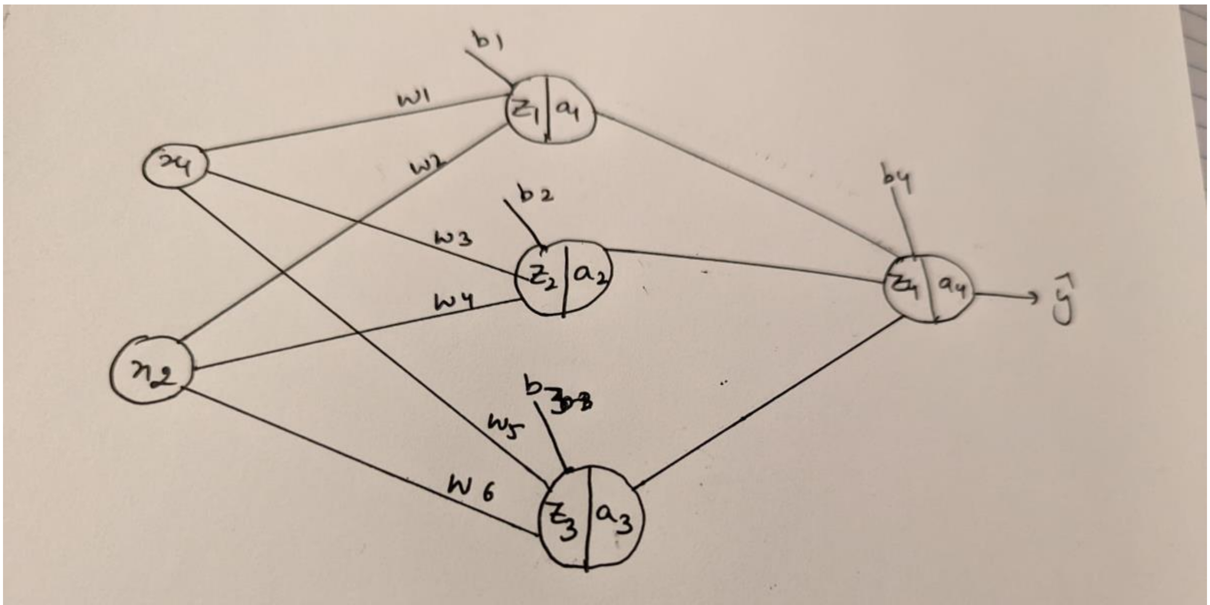
For w9: w9 ----> z4 ------ > a4 ----→ Error

For b1: b1 ----> z1 ----- > a1.------ > z4 ------ > a4 ----→ Error

For b2: b2 ----> z2 ----- > a2.------ > z4 ------ > a4 ----→ Error

For b3: b3 ----> z3 ----- > a3.------ > z4 ------ > a4 ----→ Error

For b4: b4 ----> z4 ------ > a4 ----→ Error



$x_1 = 1, \quad x_2 = 0$

$w1 = w3 = w5 = 0.5$

$w2 = w4 = w6 = -0.25$

$w7 = 1, \quad w8 = -1, w9 = 0$

biases $= 0.1$

→ Forward pass:

$$Z1 = w1x1 + w2x2 + b1$$

$$= (0.5 \times 1) + (-0.25 \times 0) + 0.1$$

$$Z_1 = 0.6$$

$$a_1 = \sigma(0.6) = \frac{1}{1 + e^{-0.6}} = 0.646$$

$\underline{z_2} = w_3 x_1 + w_4 x_2 + b_2$

$= (0.5 \times 1) + (-0.25 \times 0) + 0.1$

$z_2 = 0.6$

$a_2 = \sigma(0.6) = \frac{1}{1+e^{-0.6}} = 0.646$

$\underline{z_3} = w_5 x_1 + w_6 x_2 + b_3$

$= (0.5 \times 1) + (-0.25 \times 0) + 0.1$

$z_3 = 0.6$

$a_3 = \sigma(0.6) = \frac{1}{1+e^{-0.6}} = 0.646$

$\underline{z_4} = w_7 a_1 + w_8 a_2 + w_9 a_3 + b_4$

$= (1 \times 0.646) + (-1 \times 0.646) + (0 \times 0.646) + 0.1$

$= 0.1$

$a_4 = \sigma(0.1) = \frac{1}{1+e^{-0.1}} = 0.525$

<u>ERROR (MSE):</u>

Loss function $= MSE = \frac{1}{2}(y - \hat{y})^2$

$MSE = \frac{1}{2}(1 - 0.525)^2$

$= \frac{1}{2}(0.2256) = \underline{0.113}$

Back propagation

→ for b_4 :

$$\frac{\partial E}{\partial b_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial b_4}$$

$$\frac{\partial E}{\partial a_4} = \frac{\partial}{\partial a_4}\left(\tfrac{1}{2}(y-\hat{y})^2\right) = -(y-\hat{y})$$

$$= -(1-0.525)$$

$$\boxed{\frac{\partial E}{\partial a_4} = -0.475}$$

$$\frac{\partial a_4}{\partial z_4} = \frac{\partial}{\partial z_4}\left(1/1+e^{-x}\right) = (\sigma_{z_4})(1-\sigma_{z_4})$$

$$= (0.525) \times (1-0.525)$$

$$= 0.525 \times 0.475$$

$$\boxed{\frac{\partial a_4}{\partial z_4} = 0.250}$$

$$\frac{\partial z_4}{\partial b_4} = \frac{\partial}{\partial b_4}\left(w_7 a_1 + w_8 a_2 + w_9 a_3 + b_4\right)$$

$$\boxed{\frac{\partial z_4}{\partial b_4} = 1}$$

Hence, $$\frac{\partial E}{\partial b_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial b_4}$$

$$= -0.475 \times 0.250 \times 1$$

$$= -0.119$$

Taking η (learning rate) = 0.1

$$b_{4new} = b_4 - \eta\left(\frac{\partial E}{\partial b_4}\right)$$

$$= 0.1 - (0.1 \times -0.119)$$

$$\boxed{b_{4new} = 0.112}$$

for w_7:

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_7}$$

we already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.475 \text{ and } \frac{\partial a_4}{\partial z_4} = 0.250$$

$$\frac{\partial z_4}{\partial w_7} = \frac{\partial}{\partial w_7}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= a_1 = 0.646$$

$$\boxed{\frac{\partial z_4}{\partial w_7} = 0.646}$$

Taking η (learning rate) = 0.1

$$\frac{\partial E}{\partial w_7} = -0.475 \times 0.250 \times 0.646$$

$$= -0.077$$

$$\frac{\partial E}{\partial w_7} = -0.077$$

$$w_{7new} = w_7 - (\eta)\left(\frac{\partial E}{\partial w_7}\right)$$

$$= 1 - (0.1 \times -0.077)$$

$$= 1.0077$$

$$\boxed{w_{7new} = 1.0077}$$

→ for w_8:

$$\frac{\partial E}{\partial w_8} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_8}$$

we already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.4751 \text{ and } \frac{\partial a_4}{\partial z_4} = 0.250$$

$$\frac{\partial z_4}{\partial w_8} = \frac{\partial (a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)}{\partial w_8}$$

$$= a_2 = 0.646$$

$$\frac{\partial E}{\partial w_8} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_8}$$

$$= -0.4751 \times 0.250 \times 0.646$$

$$= -0.077$$

Taking η (learning rate) = 0.1

$$w_{8new} = w_8 - \eta\left(\frac{\partial E}{\partial w_8}\right)$$

$$= 1 - ((0.1) \times (-0.077))$$

$$\boxed{w_{8new} = -0.992}$$

→ for w_9:

$$\frac{\partial E}{\partial w_9} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_9}$$

we already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.4751 \quad \text{and} \quad \frac{\partial a_4}{\partial z_4} = 0.250$$

$$\frac{\partial z_4}{\partial w_9} = \frac{\partial}{\partial w_9}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= a_3 = 0.646$$

$$\frac{\partial E}{\partial w_9} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial w_9}$$

$$= -0.4751 \times 0.250 \times 0.646$$

$$\frac{\partial E}{\partial w_9} = -0.077$$

Taking η (learning rate) $= 0.1$

$$w_{9new} = w_9 - \eta\left(\frac{\partial E}{\partial w_9}\right)$$

$$= 0 - (0.1 \times -0.077)$$

$$\boxed{w_{9new} = 0.0077}$$

for w1:

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

we already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.4751 \quad \text{and} \quad \frac{\partial a_4}{\partial z_4} = 0.250$$

$$\frac{\partial z_4}{\partial a_1} = \frac{\partial}{\partial a_1}(a_1 w_7 + a_2 w_8 + a_3 w_9 + b_4)$$

$$= w_7 = 1$$

$$\boxed{\frac{\partial z_4}{\partial a_1} = 1}$$

$$\frac{\partial a_1}{\partial z_1} = \frac{\partial}{\partial z_1}(1/1+e^{-x}) = \sigma(1-\sigma)$$

$$= 0.646 \times (1 - 0.646)$$

$$= 0.229$$

$$\frac{\partial z_1}{\partial w_1} = \frac{\partial}{\partial w_1}(w_1 x_1 + w_2 x_2 + b_1)$$

$$= x_1 = 1$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

$$= -0.4751 \times 0.250 \times 1 \times 0.229 \times 1$$

$$= -0.027$$

Taking η (learning rate) $= 0.1$

$$w_{1new} = w_1 - \left(\eta \times \frac{\partial E}{\partial w_1}\right)$$

$$= 0.5 - (0.1 \times -0.027)$$

$$\boxed{w_{1new} = 0.5027}$$

→ For W_2 :

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_2}$$

We already know the value of,

$$\frac{\partial E}{\partial a_4} = -0.4751 \; ; \; \frac{d a_4}{\partial z_4} = 0.250$$

$$\frac{\partial z_4}{\partial a_1} = 1 \; ; \; \frac{\partial a_1}{\partial z_1} = 0.229$$

$$\frac{\partial z_1}{\partial w_2} = \frac{\partial (x_1 w_1 + x_2 w_2 + b_1)}{\partial w_2}$$

$$= x_2 = 0$$

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_2}$$

$$= -0.4751 \times 0.250 \times 1 \times 0.229 \times 0$$

$$= 0$$

Taking $\eta = 0.1$

$$W_{2new} = W_2 - \eta \times \frac{\partial E}{\partial w_2}$$

$$= -0.25 - 0$$

$$\boxed{W_{2new} = -0.25}$$

for $b1$

$$\frac{\partial E}{\partial b_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}$$

we already know the value of, $\frac{\partial z_4}{\partial a_1} = 1$

$\frac{\partial E}{\partial a_4} = -0.4751$ $\quad \frac{\partial a_4}{\partial z_4} = 0.250$

$$\frac{\partial a_1}{\partial z_1} = 0.229$$

$$\frac{\partial z_1}{\partial b_1} = \frac{\partial (w_1 x_1 + w_2 x_2 + b_1)}{\partial b_1}$$

$$= 1$$

$$\frac{\partial E}{\partial b_1} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}$$

$$= -0.4751 \times 0.250 \times 1 \times 0.229 \times 1$$

$$= -0.027$$

$$b_{1new} = b_1 - \eta \times \frac{\partial E}{\partial b_1}$$

$$= 0.1 - (0.1 \times -0.027)$$

$$\boxed{b_{1new} = 0.1027}$$

→ for w_3:

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_3}$$

we already know the value of,

$\frac{\partial E}{\partial a_4} = -0.4751$ and $\frac{\partial a_4}{\partial z_4} = 0.250$

$$\frac{\partial z_4}{\partial a_2} = \frac{\partial}{\partial a_2}(w_7 a_1 + w_8 a_2 + w_9 a_3 + b_4)$$

$$= w_8 = -1$$

$$\frac{\partial a_2}{\partial z_2} = \frac{\partial}{\partial z_2}(1/1+e^{-x}) = \sigma(1-\sigma)$$

$$= 0.646 \times (1-0.646)$$

$$\frac{\partial a_2}{\partial z_2} = 0.229$$

$$\frac{\partial z_2}{\partial w_3} = \frac{\partial}{\partial w_3}(x_1 w_3 + x_2 w_4 + b_2)$$

$$= x_1 = 1$$

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_3}$$

$$= -0.4751 \times 0.250 \times -1 \times 0.229 \times 1$$

$$= 0.0271$$

Taking the η (learning rate) $= 0.1$

$$w_{3new} = w_3 - \eta\left(\frac{\partial E}{\partial w_3}\right)$$

$$= 0.5 - (0.1 \times 0.0271)$$

$$= 0.5 - 0.00271$$

$$\boxed{w_{3new} = 0.4973}$$

→ **for w_4:**

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_4}$$

We already know the value of,

$\frac{\partial E}{\partial a_4} = -0.4751$; $\frac{\partial a_4}{\partial z_4} = 0.250$; $\frac{\partial z_4}{\partial a_2} = -1$

$\frac{\partial a_2}{\partial z_2} = 0.229$:

$$\frac{\partial z_2}{\partial w_4} = \frac{\partial (x_1 w_3 + x_2 w_4 + b_2)}{\partial w_4}$$

$$= x_2 = 0$$

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_4}$$

$$= -0.4751 \times 0.250 \times -1 \times 0.229 \times 0$$

$$\frac{\partial E}{\partial w_4} = 0$$

Taking $\eta = 0.1$,

$$w_{4new} = w_4 - \eta \times \frac{\partial E}{\partial w_4}$$

$$= -0.25 - 0$$

$$\boxed{w_{4new} = -0.25}$$

→ For b_2:

$$\frac{\partial E}{\partial b_2} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2}} \times \frac{\partial z_2}{\partial b_2}$$

this we already know

$$\frac{\partial z_2}{\partial b_2} = \frac{\partial (x_1 w_3 + x_2 w_4 + b_2)}{\partial b_2}$$

$$= 1$$

$$\frac{\partial E}{\partial b_2} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_2}$$

$$= -0.4751 \times 0.250 \times -1 \times 0.229 \times 1$$

$$\frac{\partial E}{\partial b_2} = 0.0271$$

Taking $\eta = 0.1$,

$$b_{2new} = b_2 - \left(\eta \times \frac{\partial E}{\partial b_2}\right)$$

$$= 0.1 - (0.1 \times 0.0271)$$

$$\boxed{b_{2new} = 0.0973}$$

→ For w5:

$$\frac{\partial E}{\partial w_5} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4}} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_5}$$

this we already know

$$\frac{\partial z_4}{\partial a_3} = \frac{\partial}{\partial a_3}(a_1 w_7 + a_2 w_8 + a_3 w_9)$$

$$= w_9 = 0$$

$$\frac{\partial a_3}{\partial z_3} = \frac{\partial}{\partial z_3}(1/1+e^{-x}) = \sigma(1-\sigma) = 0.646(1-0.646)$$

$$= 0.229$$

$$\frac{\partial z_3}{\partial w_5} = \frac{\partial}{\partial w_5}(x_1 w_5 + x_2 w_6 + b_3)$$

$$= x_1 = 1$$

$$\frac{\partial E}{\partial w_5} = -0.4751 \times 0.250 \times 0 \times 0.229 \times 1$$

$$= 0$$

Taking $\eta = 0.1$

$$w_{5new} = w_5 - \left(\eta \times \frac{\partial E}{\partial w_5}\right)$$

$$= 0.5 - (0.1 \times 0)$$

$$\boxed{w_{5new} = 0.5}$$

→ For w_6 :

$$\frac{\partial E}{\partial w_6} = \boxed{\frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3}} \times \frac{\partial z_3}{\partial w_6}$$

this we already know

$$\frac{\partial z_3}{\partial w_6} = \frac{\partial(x_1 w_5 + x_2 w_6 + b_3)}{\partial w_6}$$

$$= x_2 = 0$$

$$\frac{\partial E}{\partial w_6} = -0.4751 \times 0.250 \times 0 \times 0.229 \times 0$$

$$= 0$$

Taking $\eta = 0.1$

$$w_{6new} = w_6 - \left(\eta \times \frac{\partial E}{\partial w_6}\right)$$

$$= -0.25 - 0$$

$$\boxed{w_{6new} = -0.25}$$

→ For b_3 :

$$\frac{\partial E}{\partial b_3} = \frac{\partial E}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \times \frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial b_3}$$

this is 0 as w_9 is 0, calculated above.

$$\frac{\partial E}{\partial b_3} = 0$$

Taking $\eta = 0.1$,

$$b_{3\,new} = 0.1 - \left(\eta \times \frac{\partial E}{\partial b_3}\right)$$

$$\boxed{b_{3\,new} = 0.1}$$

→ updated weights after one Iteration

Old weight	New weights
$w_1 = 0.5$	$w_1 = 0.5027$
$w_2 = -0.25$	$w_2 = -0.25$
$w_3 = 0.5$	$w_3 = 0.4973$
$w_4 = -0.25$	$w_4 = -0.25$
$w_5 = 0.5$	$w_5 = 0.5$
$w_6 = -0.25$	$w_6 = -0.25$
$w_7 = 1$	$w_7 = 1.0077$
$w_8 = -1$	$w_8 = -0.992$
$w_9 = 0$	$w_9 = 0.0077$
$b_1 = 0.1$	$b_1 = 0.1027$
$b_2 = 0.1$	$b_2 = 0.0973$
$b_3 = 0.1$	$b_3 = 0.1$
$b_4 = 0.1$	$b_4 = 0.112$

Hence, as one of the input was zero, w2 , w4 and w6 doesn't change at all. Also the weights are updated very little. Due to this, it is recommended to use He distribution, glorot or xavier distribution which helps in converging to the global minima faster. Also, learning rate can be adjusted accordingly.

Coding!!!

Problem 2 **(10 pts):** We will develop an Artificial Neural Networks using MNIST digit data, where we will train an ANN model and then classify new instances. You can directly download the data using scikit learn. The dataset currently contains 10 classes. You should split the data into train and test data, where train data should be used for only training the model. You should select any random two classes' data for develop a binary classification. For instance, you can select 5 and 6.

Ans: Please check the notebook with detailed description of the model and results along with codes.

Summary of the models is given below

Questions	Test Accuracy	Test loss	Precision	Recall	F1 Score
Question 2 Part A (Without Early Stopping)	98.69%	0.0448	This was not calculated as predictions for 5 and 6 digit are made uisng Early Stopping Method	This was not calculated as predictions for 5 and 6 digit are made uisng Early Stopping Method	This was not calculated as predictions for 5 and 6 digit are made uisng Early Stopping Method
Question 2 Part A (With Early Stopping)	98.815	0.0507	98.37%	99.35%	98.86%
Question 2 Part B (Random Normal Weights, Early Stopping)	98.21%	0.056	97.91%	98.65%	98.28%
Question 2 Part B (He Norm Weights, Early Stopping)	98.48%	0.048	98.65%	98.42%	98.53%
Question 2 Part B (Glorot Uniform Weights, Early Stopping)	98.63%	0.046	98.54%	98.83%	98.68%
Question 3 (With Early Stopping)	97%	0.112	96.97%	96.97%	96.96%

From the above figure we can see the test accuracy for different models. In ques 2 part A, when the model was trained without early stopping test accuracy was slightly low as compared to when early stopping method was used.
For Ques2 Part B, we can see the test accuracy was slightly higher for Glorot Distribution.
For Ques 3, we were able to correctly predict numbers using softmax for multi class classification