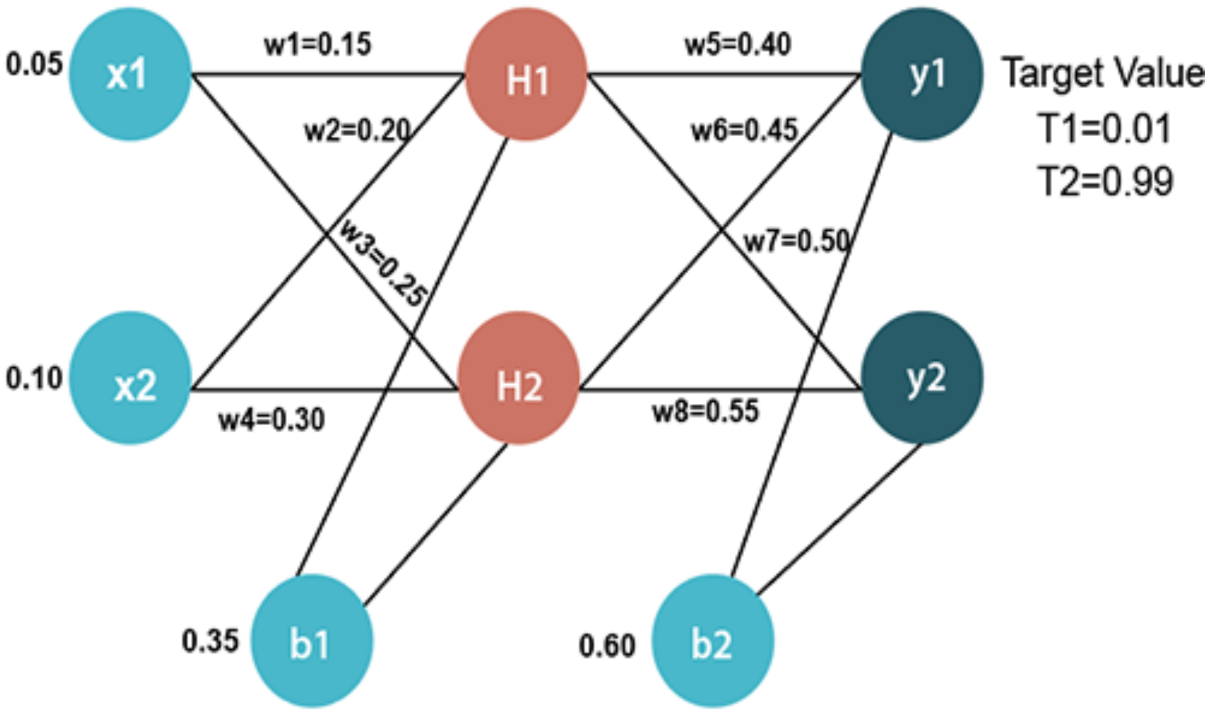
Backpropagation



Input values

X1=0.05

X2=0.10

Initial weight

W1=0.15	w5=0.40
W2=0.20	w6=0.45
W3=0.25	w7=0.50
W4=0.30	w8=0.55

Bias Values

b1=0.35 b2=0.60

Target Values

T1=0.01

T2=0.99

Now, we first calculate the values of H1 and H2 by a forward pass.
Forward Pass
To find the value of H1 we first multiply the input value from the weights as
$H1=x1\times w_1+x2\times w_2+b1$
$H1=0.05\times 0.15+0.10\times 0.20+0.35$
$H1=0.3775$

To calculate the final result of H1, we performed the sigmoid function as

$$\mathrm{H1_{final}} = \frac{1}{1+\frac{1}{e^{H1}}}$$

$$\mathrm{H1_{final}} = \frac{1}{1+\frac{1}{e^{0.3775}}}$$

$$\mathbf{H1_{final} = 0.593269992}$$

We will calculate the value of H2 in the same way as H1
$H2=x1\times w_3+x2\times w_4+b1$
$H2=0.05\times 0.25+0.10\times 0.30+0.35$
$H2=0.3925$
To calculate the final result of H1, we performed the sigmoid function as

$$\mathrm{H2_{final}} = \frac{1}{1+\frac{1}{e^{H2}}}$$

$$\mathrm{H2_{final}} = \frac{1}{1+\frac{1}{e^{0.3925}}}$$

$$\mathbf{H2_{final} = 0.596884378}$$

Now, we calculate the values of y1 and y2 in the same way as we calculate the H1 and H2.
To find the value of y1, we first multiply the input value i.e., the outcome of H1 and H2 from the weights as
$y1=H1\times w_5+H2\times w_6+b2$
$y1=0.593269992\times 0.40+0.596884378\times 0.45+0.60$
$y1=1.10590597$

To calculate the final result of y1 we performed the sigmoid function as

$$y1_{\text{final}} = \frac{1}{1 + \frac{1}{e^{y1}}}$$

$$y1_{\text{final}} = \frac{1}{1 + \frac{1}{e^{1.10590597}}}$$

$$\mathbf{y1_{final} = 0.75136507}$$

We will calculate the value of y2 in the same way as y1

$y2=H1\times w_7+H2\times w_8+b2$

$y2=0.593269992\times 0.50+0.596884378\times 0.55+0.60$

y2=1.2249214

To calculate the final result of H1, we performed the sigmoid function as

$$y2_{\text{final}} = \frac{1}{1 + \frac{1}{e^{y2}}}$$

$$y2_{\text{final}} = \frac{1}{1 + \frac{1}{e^{1.2249214}}}$$

$$\mathbf{y2_{final} = 0.772928465}$$

Our target values are 0.01 and 0.99. Our y1 and y2 value is not matched with our target values T1 and T2.

Now, we will find the **total error**, which is simply the difference between the outputs from the target outputs. The total error is calculated as

$$E_{\text{total}} = \sum \frac{1}{2}(\text{target} - \text{output})^2$$

So, the total error is

$$= \frac{1}{2}(\mathrm{t1} - \mathrm{y1_{final}})^2 + \frac{1}{2}(\mathrm{T2} - \mathrm{y2_{final}})^2$$

$$= \frac{1}{2}(0.01 - 0.75136507)^2 + \frac{1}{2}(0.99 - 0.772928465)^2$$

$$= 0.274811084 + 0.0235600257$$

$$\mathbf{E_{total} = 0.29837111}$$

Now, we will backpropagate this error to update the weights using a backward pass.
Backward pass at the output layer
To update the weight, we calculate the error correspond to each weight with the help of a total error. The error on weight w is calculated by differentiating total error with respect to w.

$$\mathrm{Error_w} = \frac{\partial \mathrm{E_{total}}}{\partial \mathrm{w}}$$

We perform backward process so first consider the last weight w5 as

$$\mathrm{Error_{w5}} = \frac{\partial \mathrm{E_{total}}}{\partial \mathrm{w5}} \ldots \ldots \ldots . . (1)$$

$$\mathrm{E_{total}} = \frac{1}{2}(\mathrm{T1} - \mathrm{y1_{final}})^2 + \frac{1}{2}(\mathrm{T2} - \mathrm{y2_{final}})^2 \ldots \ldots \ldots . . (2)$$

From equation two, it is clear that we cannot partially differentiate it with respect to w5 because there is no any w5. We split equation one into multiple terms so that we can easily differentiate it with respect to w5 as

$$\frac{\partial \mathrm{E_{total}}}{\partial \mathrm{w5}} = \frac{\partial \mathrm{E_{total}}}{\partial \mathrm{y1_{final}}} \times \frac{\partial \mathrm{y1\_final}}{\partial \mathrm{y1}} \times \frac{\partial \mathrm{y1}}{\partial \mathrm{w5}} \ldots \ldots \ldots \ldots \ (3)$$

Now, we calculate each term one by one to differentiate E_{total} with respect to w5 as

$$\frac{\partial E_{total}}{\partial y1_{final}} = \frac{\partial(\frac{1}{2}(T1 - y1_{final})^2 + \frac{1}{2}(T2 - y2_{final})^2)}{\partial y1_{final}}$$

$$= 2 \times \frac{1}{2} \times (T1 - y1_{final})^{2-1} \times (-1) + 0$$

$$= -(T1 - y1_{final})$$

$$= -(0.01 - 0.75136507)$$

$$\boldsymbol{\frac{\partial E_{total}}{\partial y1_{final}} = 0.74136507} \dots\dots\dots. \mathbf{(4)}$$

$$y1_{final} = \frac{1}{1 + e^{-y1}} \dots\dots\dots\dots\dots\dots \ (5)$$

$$\frac{\partial y1_{final}}{\partial y1} = \frac{\partial(\frac{1}{1 + e^{-y1}})}{\partial y1}$$

$$= \frac{e^{-y1}}{(1 + e^{-y1})^2}$$

$$= e^{-y1} \times (y1_{final})^2 \dots\dots\dots\dots\dots (6)$$

$$y1_{final} = \frac{1}{1 + e^{-y1}}$$

$$e^{-y1} = \frac{1 - y1_{final}}{y1_{final}} \dots\dots\dots\dots\dots (7)$$

Putting the value of e^{-y} in equation (5)

$$= \frac{1 - y1_{final}}{y1_{final}} \times (y1_{final})^2$$

$$= y1_{final} \times (1 - y1_{final})$$

$$= 0.75136507 \times (1 - 0.75136507)$$

$$\mathbf{\frac{\partial y1_{final}}{\partial y1} = 0.186815602 \ldots\ldots\ldots (8)}$$

$$y1 = H1_{final} \times w5 + H2_{final} \times w6 + b2 \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

$$\frac{\partial y1}{\partial w5} = \frac{\partial(H1_{final} \times w5 + H2_{final} \times w6 + b2)}{\partial w5}$$

$$= H1_{final}$$

$$\mathbf{\frac{\partial y1}{\partial w5} = 0.596884378 \ldots\ldots\ldots (10)}$$

So, we put the values of $\frac{\partial E_{total}}{\partial y1_{final}}, \frac{\partial y1_{final}}{\partial y1}$, and $\frac{\partial y1}{\partial w5}$ in equation no (3) to find the final result.

$$\frac{\partial E_{total}}{\partial w5} = \frac{\partial E_{total}}{\partial y1_{final}} \times \frac{\partial y1_{final}}{\partial y1} \times \frac{\partial y1}{\partial w5}$$

$$= 0.74136507 \times 0.186815602 \times 0.593269992$$

$$\mathbf{Error_{w5} = \frac{\partial E_{total}}{\partial w5} = 0.0821670407 \ldots\ldots\ldots (11)}$$

Now, we will calculate the updated weight $w5_{new}$ with the help of the following formula

$$w5_{new} = w5 - \eta \times \frac{\partial E_{total}}{\partial w5} \text{ Here, } \eta = \text{learning rate} = 0.5$$

$$= 0.4 - 0.5 \times 0.0821670407$$

$$\mathbf{w5_{new} = 0.35891648 \ldots\ldots\ldots (12)}$$

In the same way, we calculate $w6_{new}$,$w7_{new}$, and $w8_{new}$ and this will give us the following values

$w5_{new}$=0.35891648
$w6_{new}$=408666186

$w7_{new}$=0.511301270
$w8_{new}$=0.561370121

Backward pass at Hidden layer

Now, we will backpropagate to our hidden layer and update the weight w1, w2, w3, and w4 as we have done with w5, w6, w7, and w8 weights.

We will calculate the error at w1 as

$$\text{Error}_{w1} = \frac{\partial E_{total}}{\partial w1}$$

$$E_{total} = \frac{1}{2}(T1 - y1_{final})^2 + \frac{1}{2}(T2 - y2_{final})^2$$

From equation (2), it is clear that we cannot partially differentiate it with respect to w1 because there is no any w1. We split equation (1) into multiple terms so that we can easily differentiate it with respect to w1 as

$$\frac{\partial E_{total}}{\partial w1} = \frac{\partial E_{total}}{\partial H1_{final}} \times \frac{\partial H1\_final}{\partial H1} \times \frac{\partial H1}{\partial w1} \dots\dots\dots\dots \; (13)$$

Now, we calculate each term one by one to differentiate E_{total} with respect to w1 as

$$\frac{\partial E_{total}}{\partial H1_{final}} = \frac{\partial(\frac{1}{2}(T1 - y1_{final})^2 + \frac{1}{2}(T2 - y2_{final})^2)}{\partial H1} \dots\dots\dots\dots (14)$$

We again split this because there is no any $H1^{final}$ term in E^{toatal} as

$$\frac{\partial E_{total}}{\partial H1_{final}} = \frac{\partial E_1}{\partial H1_{final}} + \frac{\partial E_2}{\partial H1_{final}} \dots\dots\dots (15)$$

$\frac{\partial E_1}{\partial H1_{final}}$ and $\frac{\partial E_2}{\partial H1_{final}}$ will again split because in E1 and E2 there is no H1 term.

Splitting is done as

$$\frac{\partial E_1}{\partial H1_{final}} = \frac{\partial E_1}{\partial y1} \times \frac{\partial y1}{\partial H1_{final}} \dots\dots\dots (16)$$

$$\frac{\partial E_2}{\partial H1_{final}} = \frac{\partial E_2}{\partial y2} \times \frac{\partial y2}{\partial H1\_final} \dots\dots\dots (17)$$

We again Split both $\frac{\partial E_1}{\partial y1}$ and $\frac{\partial E_2}{\partial y2}$ because there is no any y1 and y2 term in E1 and E2.

We split it as

$$\frac{\partial E_1}{\partial y1} = \frac{\partial E_1}{\partial y1_{fianl}} \times \frac{\partial y1_{final}}{\partial y1} \ldots\ldots\ldots (18)$$

$$\frac{\partial E_2}{\partial y2} = \frac{\partial E_2}{\partial y2_{fianl}} \times \frac{\partial y2_{final}}{\partial y2} \ldots\ldots\ldots (19)$$

Now, we find the value of $\frac{\partial E_1}{\partial y1}$ and $\frac{\partial E_2}{\partial y2}$ by putting values in equation (18) and (19) as
From equation (18)

$$\frac{\partial E_1}{\partial y1} = \frac{\partial E_1}{\partial y1_{fianl}} \times \frac{\partial y1_{final}}{\partial y1}$$

$$= \frac{\partial(\frac{1}{2}(T1 - y1_{final})^2)}{\partial y1_{final}} \times \frac{\partial y1_{final}}{\partial y1}$$

$$= 2 \times \frac{1}{2}(T1 - y1_{final}) \times (-1) \times \frac{\partial y1_{final}}{\partial y1}$$

From equation (8)

$$= 2 \times \frac{1}{2}(0.01 - 0.75136507) \times (-1) \times 0.186815602$$

$$\boldsymbol{\frac{\partial E_1}{\partial y1} = 0.138498562} \ldots\ldots\ldots (20)$$

From equation (19)

$$\frac{\partial E_2}{\partial y2} = \frac{\partial E_2}{\partial y2_{fianl}} \times \frac{\partial y2_{final}}{\partial y2}$$

$$= \frac{\partial(\frac{1}{2}(T2 - y2_{final})^2)}{\partial y2_{final}} \times \frac{\partial y2_{final}}{\partial y2}$$

$$= 2 \times \frac{1}{2}(T2 - y2_{final}) \times (-1) \times \frac{\partial y2_{final}}{\partial y2} \dots \dots \dots (21)$$

$$y2_{final} = \frac{1}{1 + e^{-y2}} \dots \dots \dots \dots \dots \dots \; (22)$$

$$\frac{\partial y2_{fianl}}{\partial y2} = \frac{\partial(\frac{1}{1 + e^{-y2}})}{\partial y2}$$

$$= \frac{e^{-y2}}{(1 + e^{-y2})^2}$$

$$= e^{-y2} \times (y2_{final})^2 \dots \dots \dots \dots \dots (23)$$

$$y2_{final} = \frac{1}{1 + e^{-y2}}$$

$$e^{-y2} = \frac{1 - y2_{final}}{y2_{final}} \dots \dots \dots \dots \dots (24)$$

Putting the value of e^{-y2} in equation (23)

$$= \frac{1 - y2_{final}}{y2_{final}} \times (y2_{final})^2$$

$$= y2_{final} \times (1 - y2_{final})$$

$$= 0.772928465 \times (1 - 0.772928465)$$

$$\boldsymbol{\frac{\partial y2_{fianl}}{\partial y2} = 0.175510053 \dots \dots \dots . (25)}$$

From equation (21)

$$= 2 \times \frac{1}{2}(0.99 - 0.772928465) \times (-1) \times\ 0.175510053$$

$$\frac{\partial \mathbf{E_1}}{\partial \mathbf{y1}} = -\mathbf{0.0380982366126414} \ldots \ldots \ldots (\mathbf{26})$$

Now from equation (16) and (17)

$$\frac{\partial \mathrm{E}_1}{\partial \mathrm{H1}_{\mathrm{final}}} = \frac{\partial \mathrm{E}_1}{\partial \mathrm{y1}} \times \frac{\partial \mathrm{y1}}{\partial \mathrm{H1}_{\mathrm{final}}}$$

$$= 0.138498562 \times \frac{\partial(\mathrm{H1}_{\mathrm{final}} \times \mathrm{w}_5 + \mathrm{H2}_{\mathrm{final}} \times \mathrm{w}_6 + \mathrm{b2})}{\partial \mathrm{H1}_{\mathrm{final}}}$$

$$= 0.138498562 \times \frac{\partial(\mathrm{H1}_{\mathrm{final}} \times \mathrm{w}_5 + \mathrm{H2}_{\mathrm{final}} \times \mathrm{w}_6 + \mathrm{b2})}{\partial \mathrm{H1}_{\mathrm{final}}}$$

$$= 0.138498562 \times \mathrm{w5}$$

$$= 0.138498562 \times 0.40$$

$$\frac{\partial \mathbf{E_1}}{\partial \mathbf{H1_{final}}} = \mathbf{0.0553994248} \ldots \ldots \ldots (\mathbf{27})$$

$$\frac{\partial \mathrm{E}_2}{\partial \mathrm{H1}_{\mathrm{final}}} = \frac{\partial \mathrm{E}_2}{\partial \mathrm{y2}} \times \frac{\partial \mathrm{y2}}{\partial \mathrm{H1\_final}}$$

$$= -0.0380982366126414 \times \frac{\partial(\mathrm{H1}_{\mathrm{final}} \times \mathrm{w}_7 + \mathrm{H2}_{\mathrm{final}} \times \mathrm{w}_8 + \mathrm{b2})}{\partial \mathrm{H1}_{\mathrm{final}}}$$

$$= -0.0380982366126414 \times \mathrm{w7}$$

$$= -0.0380982366126414 \times 0.50$$

$$\frac{\partial \mathbf{E_2}}{\partial \mathbf{H1_{final}}} = -\mathbf{0.0190491183063207} \ldots \ldots \ldots (\mathbf{28})$$

Put the value of $\frac{\partial \mathrm{E}_1}{\partial \mathrm{H1}_{\mathrm{final}}}$ and $\frac{\partial \mathrm{E}_2}{\partial \mathrm{H1}_{\mathrm{final}}}$ in equation (15) as

$$\frac{\partial E_{total}}{\partial H1_{final}} = \frac{\partial E_1}{\partial H1_{final}} + \frac{\partial E_2}{\partial H1_{final}}$$

$$= 0.0553994248 + (-0.0190491183063207)$$

$$\boldsymbol{\frac{\partial E_{total}}{\partial H1_{final}} = 0.0364908241736793 \dots\dots\dots (29)}$$

We have $\frac{\partial E_{total}}{\partial H1_{final}}$; we need to figure out $\frac{\partial H1\_final}{\partial H1}$, $\frac{\partial H1}{\partial w1}$ as

$$\frac{\partial H1_{final}}{\partial H1} = \frac{\partial(\frac{1}{1+e^{-H1}})}{\partial H1}$$

$$= \frac{e^{-H1}}{(1+e^{-H1})^2}$$

$$e^{-H1} \times (H1_{final})^2 \dots\dots\dots (30)$$

$$H1_{final} = \frac{1}{1+e^{-H1}}$$

$$e^{-H1} = \frac{1 - H1_{final}}{H1_{final}} \dots\dots\dots\dots\dots (31)$$

Putting the value of e^{-H1} in equation (30)

$$= \frac{1 - H1_{final}}{H1_{final}} \times (H1_{final})^2$$

$$= H1_{final} \times (1 - H1_{final})$$

$$= 0.593269992 \times (1 - 0.593269992)$$

$$\boldsymbol{\frac{\partial H1_{final}}{\partial H1} = 0.241300708592319 9}$$

We calculate the partial derivative of the total net input to H1 with respect to w1 the same as we did for the output neuron:

$$\mathrm{H1} = \mathrm{H1}_{\mathrm{final}} \times \mathrm{w5} + \mathrm{H2}_{\mathrm{final}} \times \mathrm{w6} + \mathrm{b2} \ldots \ldots \ldots \ldots \ldots \ldots \quad (32)$$

$$\frac{\partial \mathrm{y1}}{\partial \mathrm{w1}} = \frac{\partial(\mathrm{x1} \times \mathrm{w1} + \mathrm{x2} \times \mathrm{w3} + \mathrm{b1} \times 1)}{\partial \mathrm{w1}}$$

$$= \mathrm{x1}$$

$$\frac{\partial \mathbf{H1}}{\partial \mathbf{w1}} = \mathbf{0.05} \ldots \ldots \ldots (\mathbf{33})$$

So, we put the values of $\frac{\partial \mathrm{E}_{\mathrm{total}}}{\partial \mathrm{H1}_{\mathrm{final}}}, \frac{\partial \mathrm{H1}_{\mathrm{final}}}{\partial \mathrm{H1}}$, and $\frac{\partial \mathrm{H1}}{\partial \mathrm{w1}}$ in equation (13) to find the final result.

$$\frac{\partial \mathrm{E}_{\mathrm{total}}}{\partial \mathrm{w1}} = \frac{\partial \mathrm{E}_{\mathrm{total}}}{\partial \mathrm{H1}_{\mathrm{final}}} \times \frac{\partial \mathrm{H1}_{\mathrm{final}}}{\partial \mathrm{H1}} \times \frac{\partial \mathrm{H1}}{\partial \mathrm{w1}}$$

$$= 0.0364908241736793 \times 0.2413007085923199 \times 0.05$$

$$\mathbf{Error_{w1}} = \frac{\partial \mathbf{E_{total}}}{\partial \mathbf{w1}} = \mathbf{0.000438568} \ldots \ldots \ldots . (\mathbf{34})$$

Now, we will calculate the updated weight $w1_{new}$ with the help of the following formula

$$\mathrm{w1}_{\mathrm{new}} = \mathrm{w1} - \eta \times \frac{\partial \mathrm{E}_{\mathrm{total}}}{\partial \mathrm{w1}} \text{ Here } \eta = \text{learning rate} = 0.5$$

$$= 0.15 - 0.5 \times 0.000438568$$

$$\mathbf{w1_{new}} = \mathbf{0.149780716} \ldots \ldots \ldots (\mathbf{35})$$

In the same way, we calculate $w2_{new}$,$w3_{new}$, and w4 and this will give us the following values

$w1_{new}$=0.149780716
$w2_{new}$=0.19956143
$w3_{new}$=0.24975114
$w4_{new}$=0.29950229

We have updated all the weights. We found the error 0.298371109 on the network when we fed forward the 0.05 and 0.1 inputs. In the first round of Backpropagation, the total error is down to 0.291027924. After repeating this process 10,000, the total error is down to 0.0000351085. At this point, the outputs neurons generate 0.159121960 and 0.984065734 i.e., nearby our target value when we feed forward the 0.05 and 0.1.