# MACHINE INTELLIGENCE

## Artificial Neural Network

**K.S.Srinivas**
Department of Computer Science and Engineering

# MACHINE INTELLIGENCE

## Artificial Neural Network

**K.S.Srinivas**

Department of Computer Science and Engineering

## Back Propogation Example

**STEPS:**

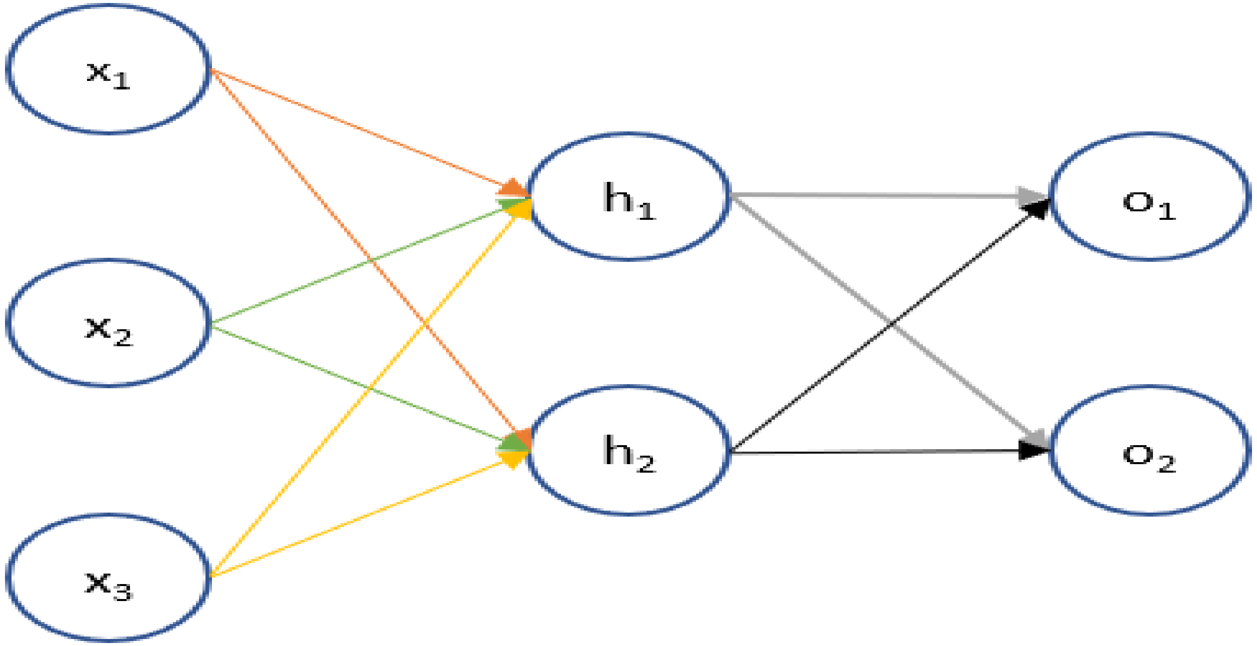
(1) Initialize weights for the parameters we want to train

(2) Forward propagate through the network to get the output values

(3) Define the error or cost function and its first derivatives

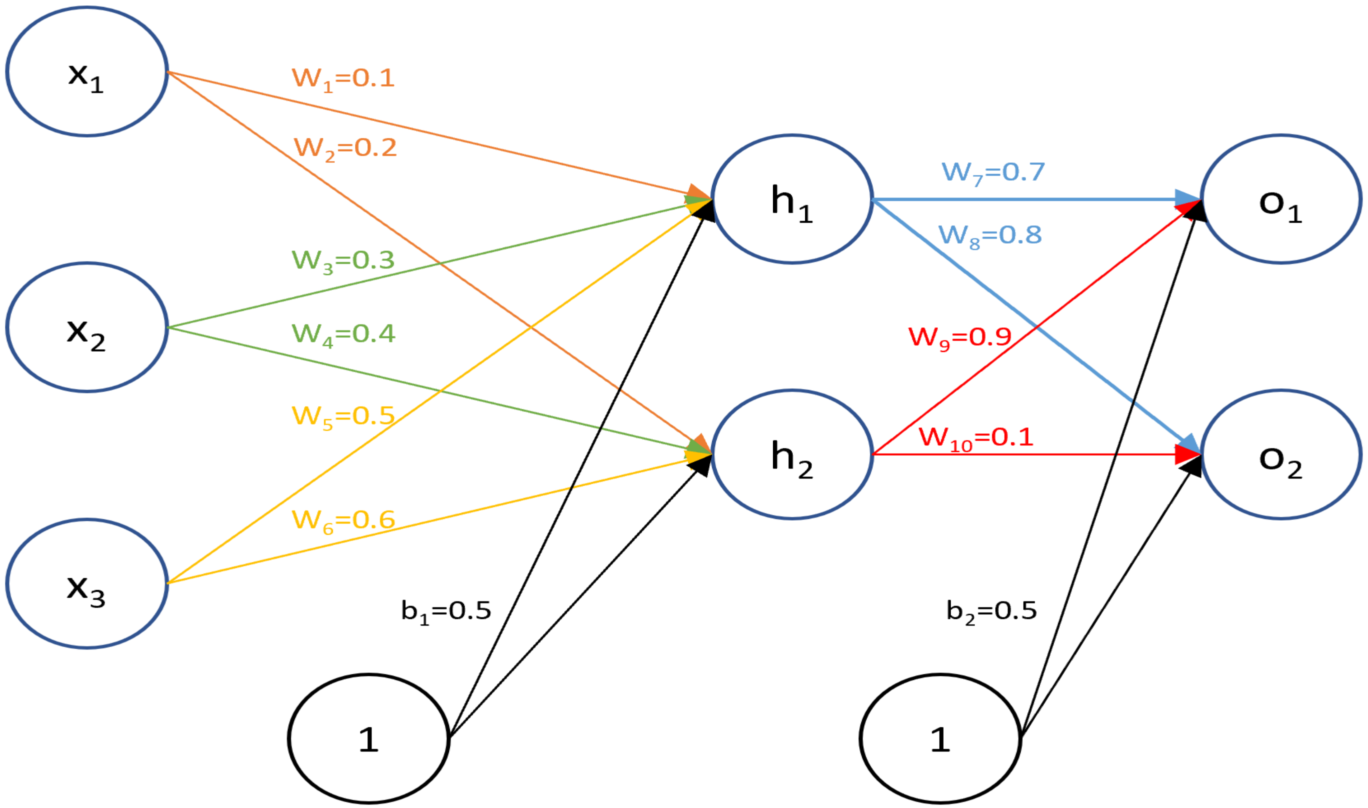
(4) Backpropagate through the network to determine the error derivatives

(5) Update the parameter estimates using the error derivative and the current value



## Back Propogation Example : Step1

The input and target values for this problem are $x_1 = 1, x_2 = 4, x_3 = 5$ and $t_1 = 0.1, t_2 = 0.05$. I will initialize weights as shown in the diagram below. Generally, you will assign them randomly but for illustration purposes, I've chosen these numbers.



## Step2 : Forward propagate through the network to get the output values

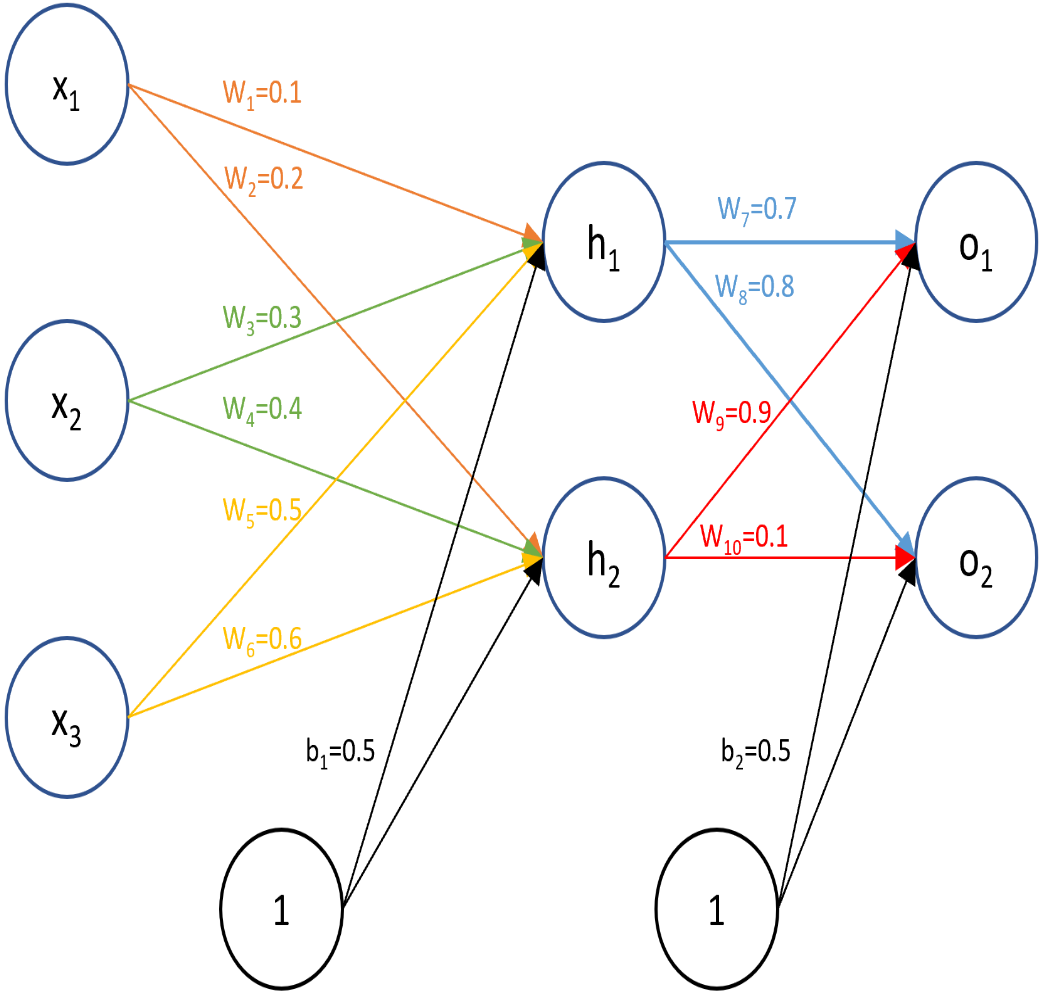
For the input and output layer, Let us use the convention of denoting

$z_{h_1}, z_{h_2}, z_{o_1},$ and $z_{o_2}$

to denote the value before the activation function is applied

$h_1, h_2, o_1,$ and $o_2$

to denote the values after application of the activation function.




Input to hidden layer

$$w_1x_1 + w_3x_2 + w_5x_3 + b_1 = z_{h_1}$$

$$w_2x_1 + w_4x_2 + w_6x_3 + b_1 = z_{h_2}$$

$$h_1 = \sigma(z_{h_1})$$

$$h_2 = \sigma(z_{h_2})$$

Hidden layer to output layer

$$w_7h_1 + w_9h_2 + b_2 = z_{o_1}$$

$$w_8h_1 + w_{10}h_2 + b_2 = z_{o_2}$$

$$o_1 = \sigma(z_{o_1})$$

$$o_2 = \sigma(z_{o_2})$$

## **Step2:** Forward propagate through the network to get the output values

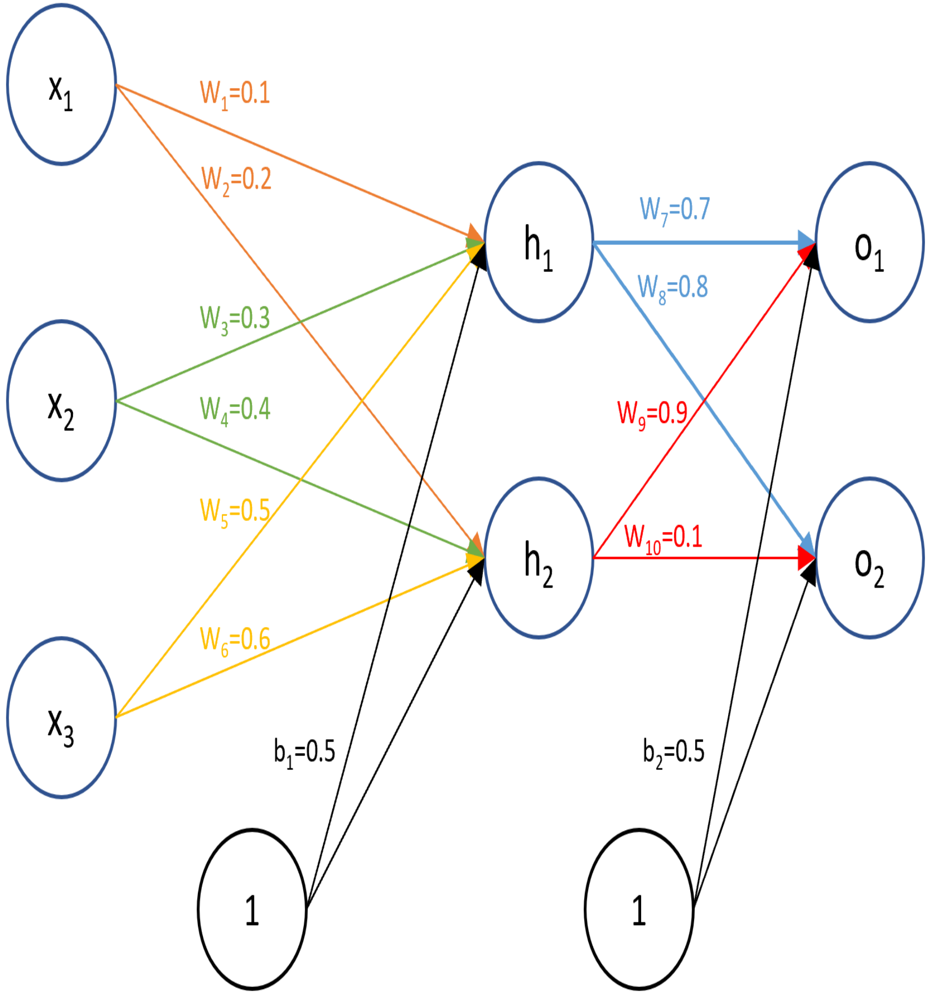
We can use the formulas discussed to forward propagate through the network. Note: I've shown up to four decimal places below but maintained all decimals in actual calculations.

$$w_1x_1 + w_3x_2 + w_5x_3 + b_1 = z_{h_1} = 0.1(1) + 0.3(4) + 0.5(5) + 0.5 = 4.3$$
$$h_1 = \sigma(z_{h_1}) = \sigma(4.3) = 0.9866$$

$$w_2x_1 + w_4x_2 + w_6x_3 + b_1 = z_{h_2} = 0.2(1) + 0.4(4) + 0.6(5) + 0.5 = 5.3$$
$$h_2 = \sigma(z_{h_2}) = \sigma(5.3) = 0.9950$$

$$w_7h_1 + w_9h_2 + b_2 = z_{o_1} = 0.7(0.9866) + 0.9(0.9950) + 0.5 = 2.0862$$
$$o_1 = \sigma(z_{o_1}) = \sigma(2.0862) = 0.8896$$

$$w_8h_1 + w_{10}h_2 + b_2 = z_{o_2} = 0.8(0.9866) + 0.1(0.9950) + 0.5 = 1.3888$$
$$o_2 = \sigma(z_{o_2}) = \sigma(1.3888) = 0.8004$$



## Step3 : Define the error or cost function and its first derivatives

We now define the sum of squares error using the target values and the results from the last layer from forward propagation.

$t_1 = 0.1, t_2 = 0.05$.

$$w_7 h_1 + w_9 h_2 + b_2 = z_{o_1} = 0.7(0.9866) + 0.9(0.9950) + 0.5 = 2.0862$$
$$o_1 = \sigma(z_{o_1}) = \sigma(2.0862) = 0.8896$$

$$w_8 h_1 + w_{10} h_2 + b_2 = z_{o_2} = 0.8(0.9866) + 0.1(0.9950) + 0.5 = 1.3888$$
$$o_2 = \sigma(z_{o_2}) = \sigma(1.3888) = 0.8004$$

$$E = \frac{1}{2}[(o_1 - t_1)^2 + (o_2 - t_2)^2$$

$$\frac{dE}{d_{o_1}} = o_1 - t_1$$

$$\frac{dE}{d_{o_2}} = o_2 - t_2$$

**Complete the Calculation…..**

$h_1 = 0.9866 \quad h_2 = 0.9950 \quad o_1 = 0.8896 \quad o_2 = 0.8004$

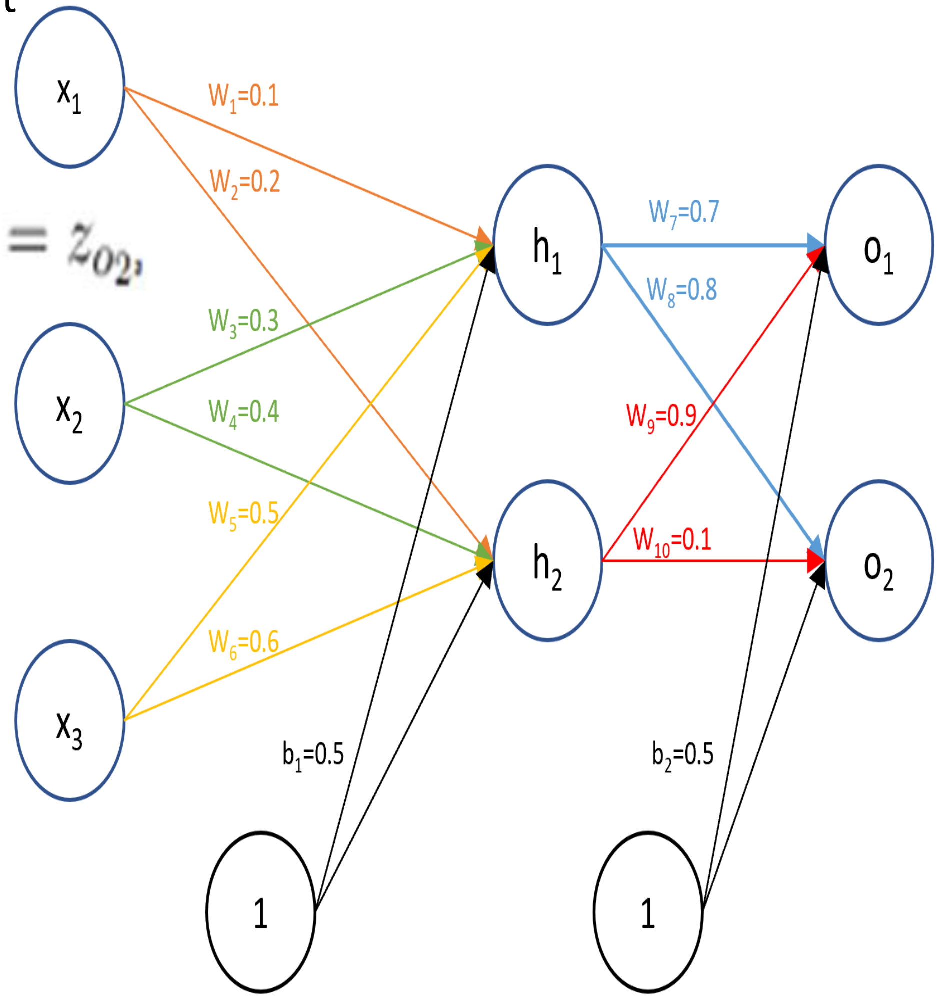
## Step3 : Define the error or cost function and its first derivatives

We are now ready to backpropagate through the network to compute all the error derivatives with respect to the parameters. Note that although there will be many long formulas, we are not doing anything fancy here. We are just using the basic principles of calculus such as the chain rule.

Also, given that $w_7 h_1 + w_9 h_2 + b_2 = z_{o_1}$ and $w_8 h_1 + w_{10} h_2 + b_2 = z_{o_2}$,

we have $\frac{dz_{o_1}}{dw_7} = h_1, \frac{dz_{o_2}}{dw_8} = h_1,$

$\frac{dz_{o_1}}{dw_9} = h_2, \frac{dz_{o_2}}{dw_{10}} = h_2, \frac{dz_{o_1}}{db_2} = 1$, and $\frac{dz_{o_2}}{db_2} = 1.$



$h_1 = 0.9866 \quad h_2 = 0.9950 \quad o_1 = 0.8896 \quad o_2 = 0.8004$

## Step4: Backpropagate through the network to determine the error derivatives

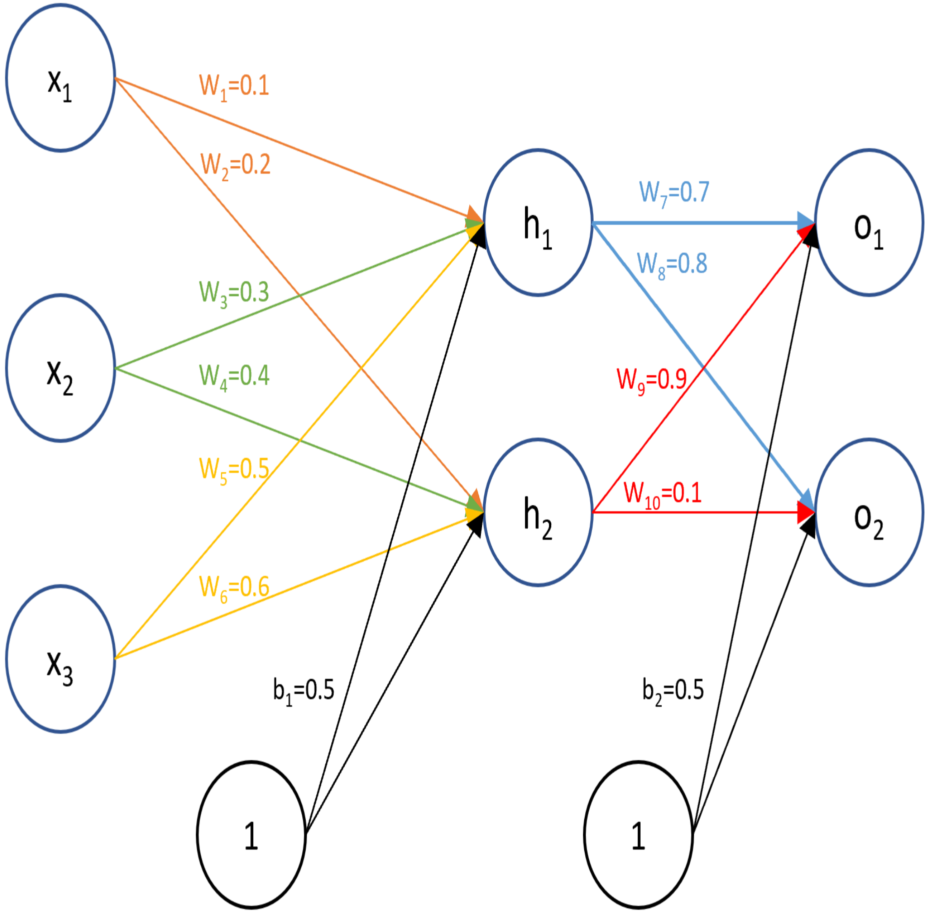
We are now ready to calculate $\frac{dE}{dw_7}, \frac{dE}{dw_8}, \frac{dE}{dw_9}$, and $\frac{dE}{dw_{10}}$ using the derivatives we have already discussed.

$$\frac{dE}{dw_7} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dw_7}$$

$$\frac{dE}{dw_7} = (o_1 - t_1)(o_1(1 - o_1))h_1$$

$$\frac{dE}{dw_7} = (0.8896 - 0.1)(0.8896(1 - 0.8896))(0.9866)$$

$$\frac{dE}{dw_7} = 0.0765$$



## Step4

$h_1 = 0.9866 \quad h_2 = 0.9950 \quad o_1 = 0.8896 \quad o_2 = 0.8004$

$$\frac{dE}{dw_8} = \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dw_8}$$

$$\frac{dE}{dw_8} = (0.7504)(0.1598)(0.9866)$$

$$\frac{dE}{dw_8} = 0.1183$$

$$\frac{dE}{dw_9} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dw_9}$$

$$\frac{dE}{dw_9} = (0.7896)(0.0983)(0.9950)$$

$$\frac{dE}{dw_9} = 0.0772$$

$$\frac{dE}{dw_{10}} = \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dw_{10}}$$

$$\frac{dE}{dw_{10}} = (0.7504)(0.1598)(0.9950)$$

$$\frac{dE}{dw_{10}} = 0.1193$$

$$\frac{dE}{dw_7} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dw_7}$$

$$\frac{dE}{dw_7} = (o_1 - t_1)(o_1(1 - o_1))h_1$$

we have $\frac{dz_{o_1}}{dw_7} = h_1, \frac{dz_{o_2}}{dw_8} = h_1,$

$$\frac{dz_{o_1}}{dw_9} = h_2, \frac{dz_{o_2}}{dw_{10}} = h_2, \frac{dz_{o_1}}{db_2} = 1, \text{and } \frac{dz_{o_2}}{db_2} = 1.$$

## Step4

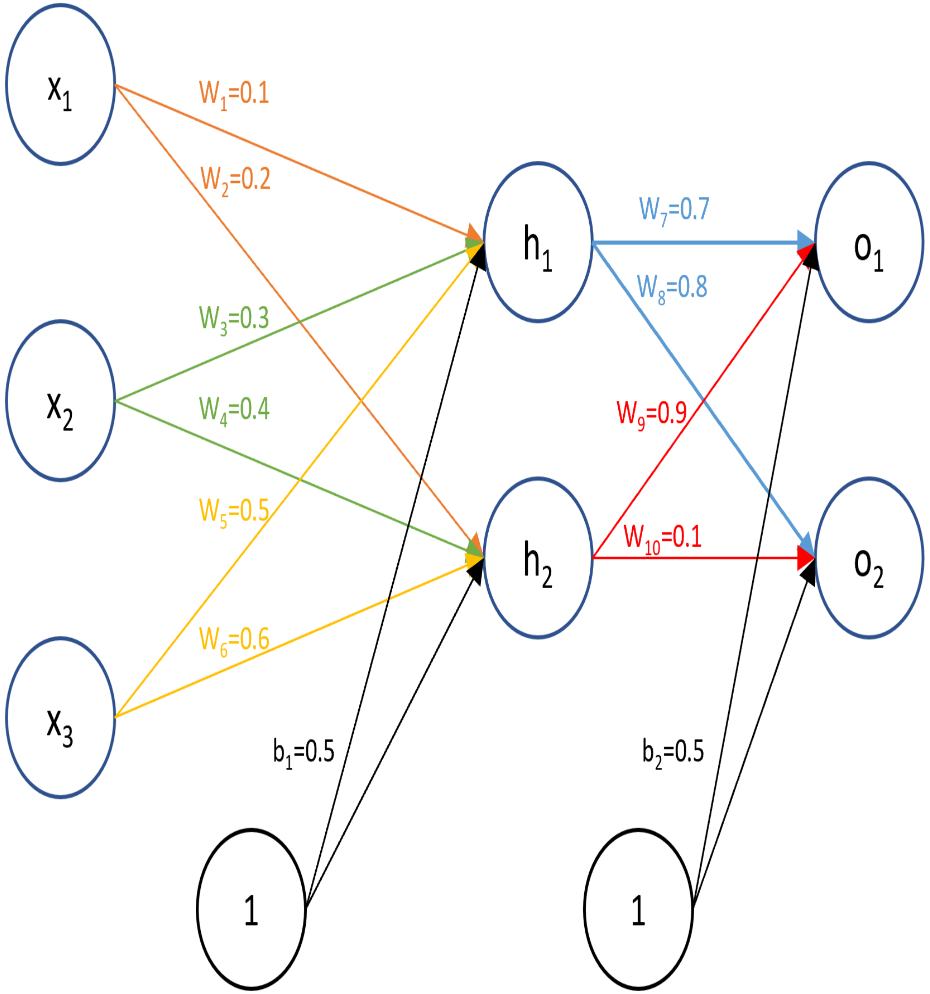
$h_1 = 0.9866 \quad h_2 = 0.9950 \quad o_1 = 0.8896 \quad o_2 = 0.8004$

The error derivative of $b_2$ is a little bit more involved since changes to $b_2$ affect the error through both $o_1$ and $o_2$.

$$\frac{dE}{db_2} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{db_2} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{db_2}$$

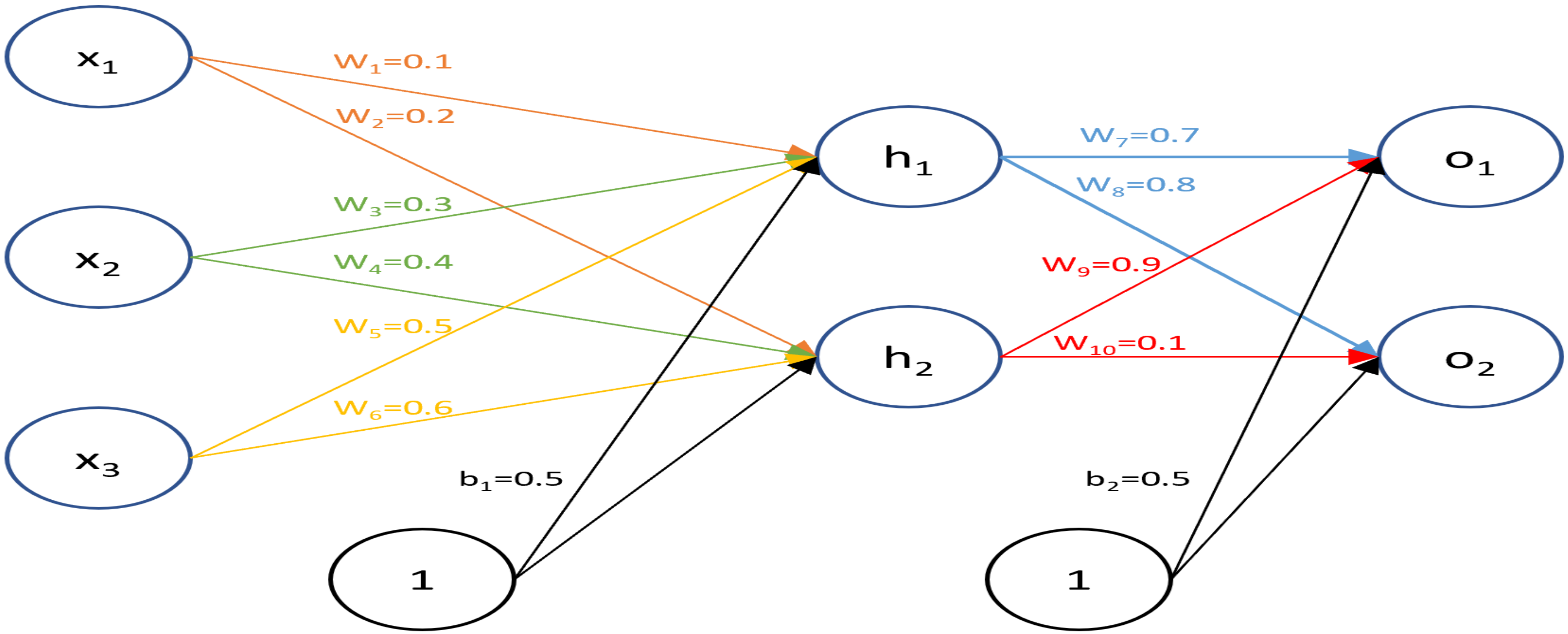
$$\frac{dE}{db_2} = (0.7896)(0.0983)(1) + (0.7504)(0.1598)(1)$$

$$\frac{dE}{db_2} = 0.1975$$



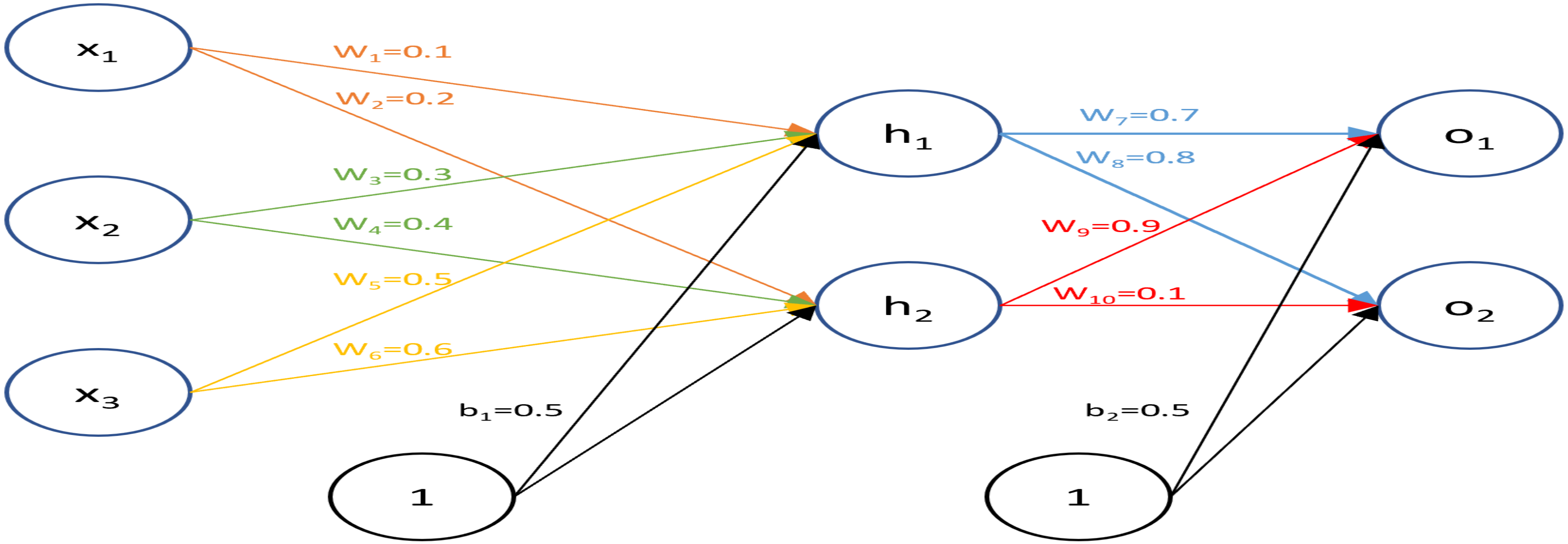
## Step4

To summarize, we have computed numerical values for the error derivatives with respect to $w_7, w_8, w_9, w_{10}$, and $b_2$. We will now backpropagate one layer to compute the error derivatives of the parameters connecting the input layer to the hidden layer. These error derivatives are $\frac{dE}{dw_1}, \frac{dE}{dw_2}, \frac{dE}{dw_3}, \frac{dE}{dw_4}, \frac{dE}{dw_5}, \frac{dE}{dw_6}$, and $\frac{dE}{db_1}$.



# MACHINE INTELLIGENCE

## Step4



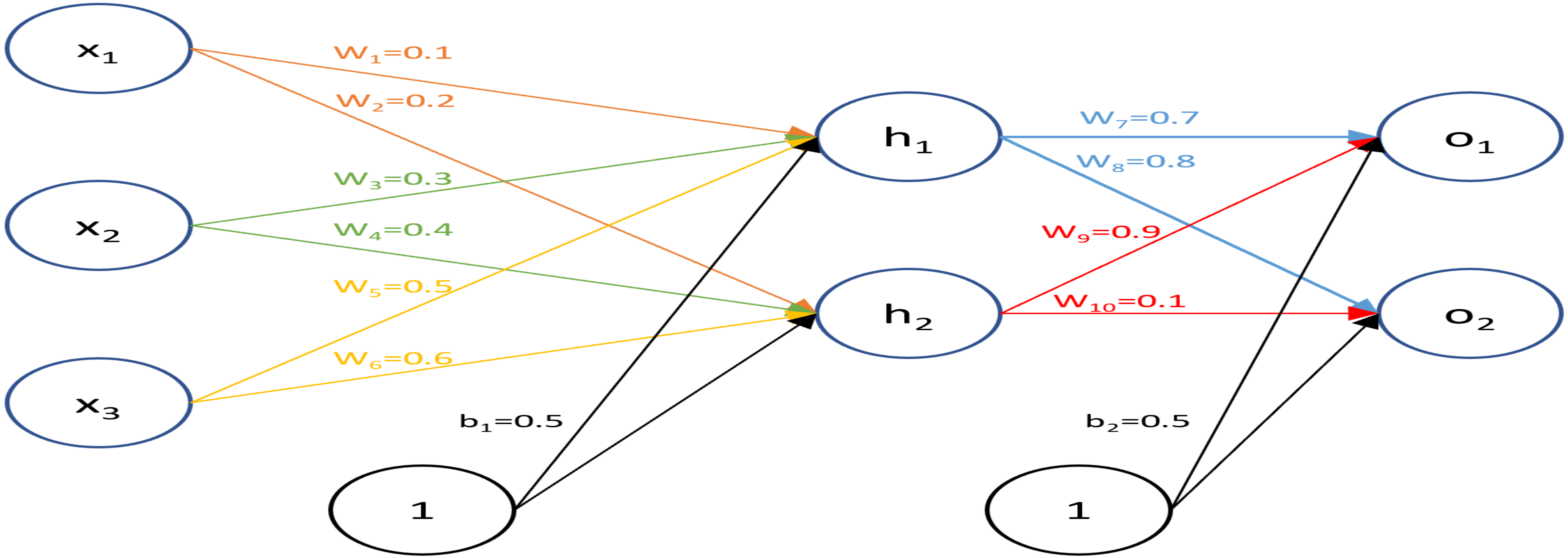

I will calculate $\frac{dE}{dw_1}$, $\frac{dE}{dw_3}$, and $\frac{dE}{dw_5}$ first since they all flow through the $h_1$ node.

$$\frac{dE}{dw_1} = \frac{dE}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{dw_1}$$

## Step4

The calculation of the first term on the right hand side of the equation above is a bit more involved than previous calculations since $h_1$ affects the error through both $o_1$ and $o_2$.

$$\frac{dE}{dh_1} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_1} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_1}$$



## Step4

$$\frac{dE}{dh_1} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_1} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_1}$$

PES UNIVERSITY ONLINE

Now I will proceed with the numerical values for the error derivatives above. These derivatives have already been calculated above or are similar in style to those calculated above. If anything is unclear, please leave a comment.

$$\frac{dE}{dh_1} = (0.7896)(0.0983)(0.7) + (0.7504)(0.1598)(0.8) = 0.1502$$

Plugging the above into the formula for $\frac{dE}{dw_1}$, we get

$$\frac{dE}{dw_1} = (0.1502)(0.0132)(1) = 0.0020$$

$$\frac{dE}{dw_1} = \frac{dE}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{dw_1}$$

The calculations for $\frac{dE}{dw_3}$ and $\frac{dE}{dw_5}$ are below

$$\frac{dE}{dw_3} = \frac{dE}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{dw_3}$$

$$\frac{dE}{dw_3} = (0.1502)(0.0132)(4) = 0.0079$$

$$\frac{dE}{dw_5} = \frac{dE}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{dw_5}$$

$$\frac{dE}{dw_5} = (0.1502)(0.0132)(5) = 0.0099$$

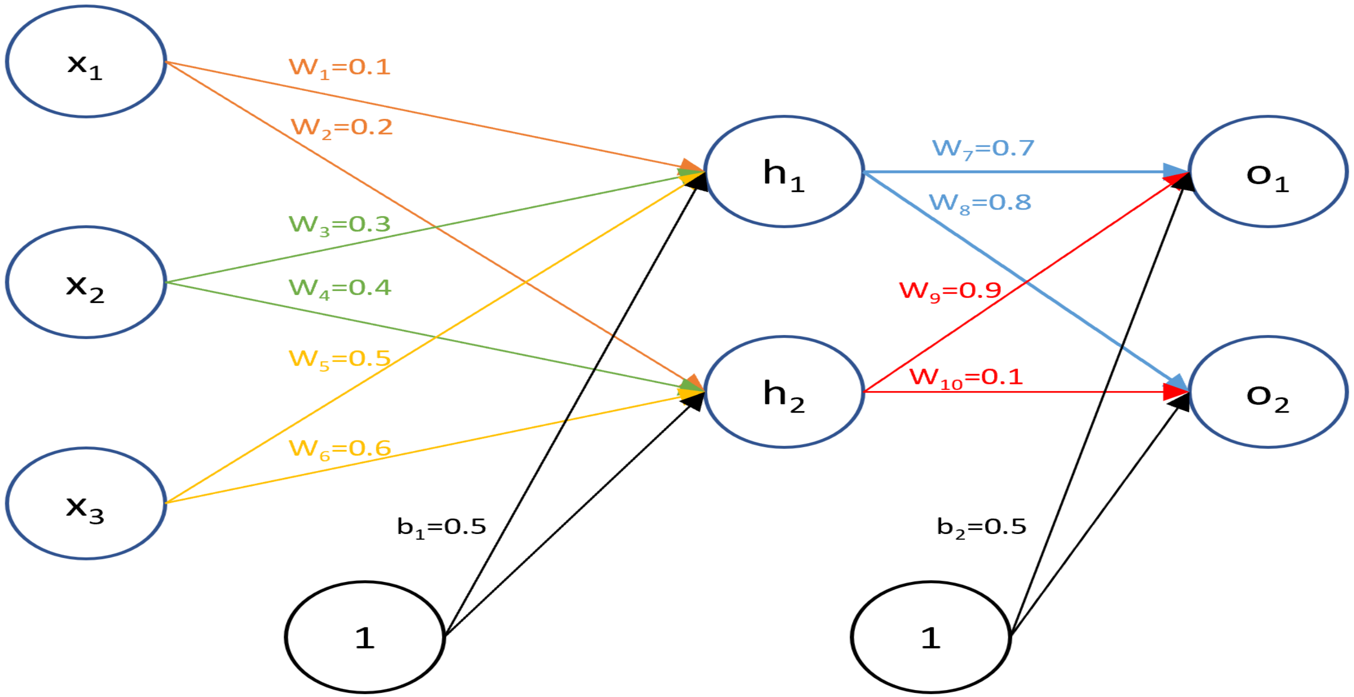
## Step4

I will now calculate $\frac{dE}{dw_2}$, $\frac{dE}{dw_4}$, and $\frac{dE}{dw_6}$ since they all flow through the $h_2$ node.

$$\frac{dE}{dw_2} = \frac{dE}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{dw_2}$$

The calculation of the first term on the right hand side of the equation above is a bit more involved since $h_2$ affects the error through both $o_1$ and $o_2$.

$$\frac{dE}{dh_2} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_2} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_2}$$



## Step4

I will now calculate $\frac{dE}{dw_2}$, $\frac{dE}{dw_4}$, and $\frac{dE}{dw_6}$ since they all flow through the $h_2$ node.

$$\frac{dE}{dw_2} = \frac{dE}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{dw_2}$$

The calculation of the first term on the right hand side of the equation above is a bit more involved since $h_2$ affects the error through both $o_1$ and $o_2$.

$$\frac{dE}{dh_2} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_2} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_2}$$

$$\frac{dE}{dh_2} = (0.7896)(0.0983)(0.9) + (0.7504)(0.1598)(0.1) = 0.0818$$

Plugging the above into the formula for $\frac{dE}{dw_2}$, we get

$$\frac{dE}{dw_2} = (0.0818)(0.0049)(1) = 0.0004$$

## Step4

The calculations for $\frac{dE}{dw_4}$ and $\frac{dE}{dw_6}$ are below

$$\frac{dE}{dw_4} = \frac{dE}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{dw_4}$$

$$\frac{dE}{dw_4} = (0.0818)(0.0049)(4) = 0.0016$$

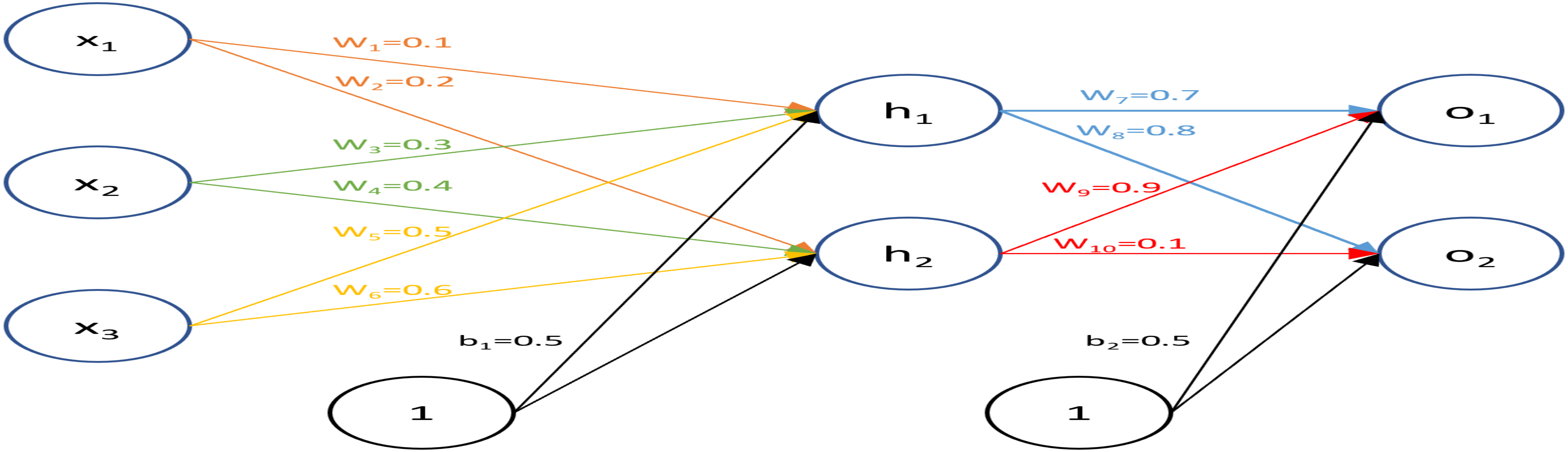
$$\frac{dE}{dw_6} = \frac{dE}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{dw_6}$$

$$\frac{dE}{dw_6} = (0.0818)(0.0049)(5) = 0.0020$$

## Step4

The final error derivative we have to calculate is $\frac{dE}{db_1}$, which is done next

$$\frac{dE}{db_1} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{db_1} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{db_1}$$



## Step4

The final error derivative we have to calculate is $\frac{dE}{db_1}$, which is done next

$$\frac{dE}{db_1} = \frac{dE}{do_1}\frac{do_1}{dz_{o_1}}\frac{dz_{o_1}}{dh_1}\frac{dh_1}{dz_{h_1}}\frac{dz_{h_1}}{db_1} + \frac{dE}{do_2}\frac{do_2}{dz_{o_2}}\frac{dz_{o_2}}{dh_2}\frac{dh_2}{dz_{h_2}}\frac{dz_{h_2}}{db_1}$$

$$\frac{dE}{db_1} = (0.7896)(0.0983)(0.7)(0.0132)(1) + (0.7504)(0.1598)(0.1)(0.0049)(1) = 0.0008$$

## Step5

We now have all the error derivatives and we're ready to make the parameter updates after the first iteration of backpropagation. We will use the learning rate of $\alpha = 0.01$

$$w_1 := w_1 - \alpha \frac{dE}{dw_1} = 0.1 - (0.01)(0.0020) = 0.1000$$

$$w_2 := w_2 - \alpha \frac{dE}{dw_2} = 0.2 - (0.01)(0.0004) = 0.2000$$

$$w_3 := w_3 - \alpha \frac{dE}{dw_3} = 0.3 - (0.01)(0.0079) = 0.2999$$

$$w_4 := w_4 - \alpha \frac{dE}{dw_4} = 0.4 - (0.01)(0.0016) = 0.4000$$

$$w_5 := w_5 - \alpha \frac{dE}{dw_5} = 0.5 - (0.01)(0.0099) = 0.4999$$

$$w_6 := w_6 - \alpha \frac{dE}{dw_6} = 0.6 - (0.01)(0.0020) = 0.6000$$

$$w_7 := w_7 - \alpha \frac{dE}{dw_7} = 0.7 - (0.01)(0.0765) = 0.6992$$

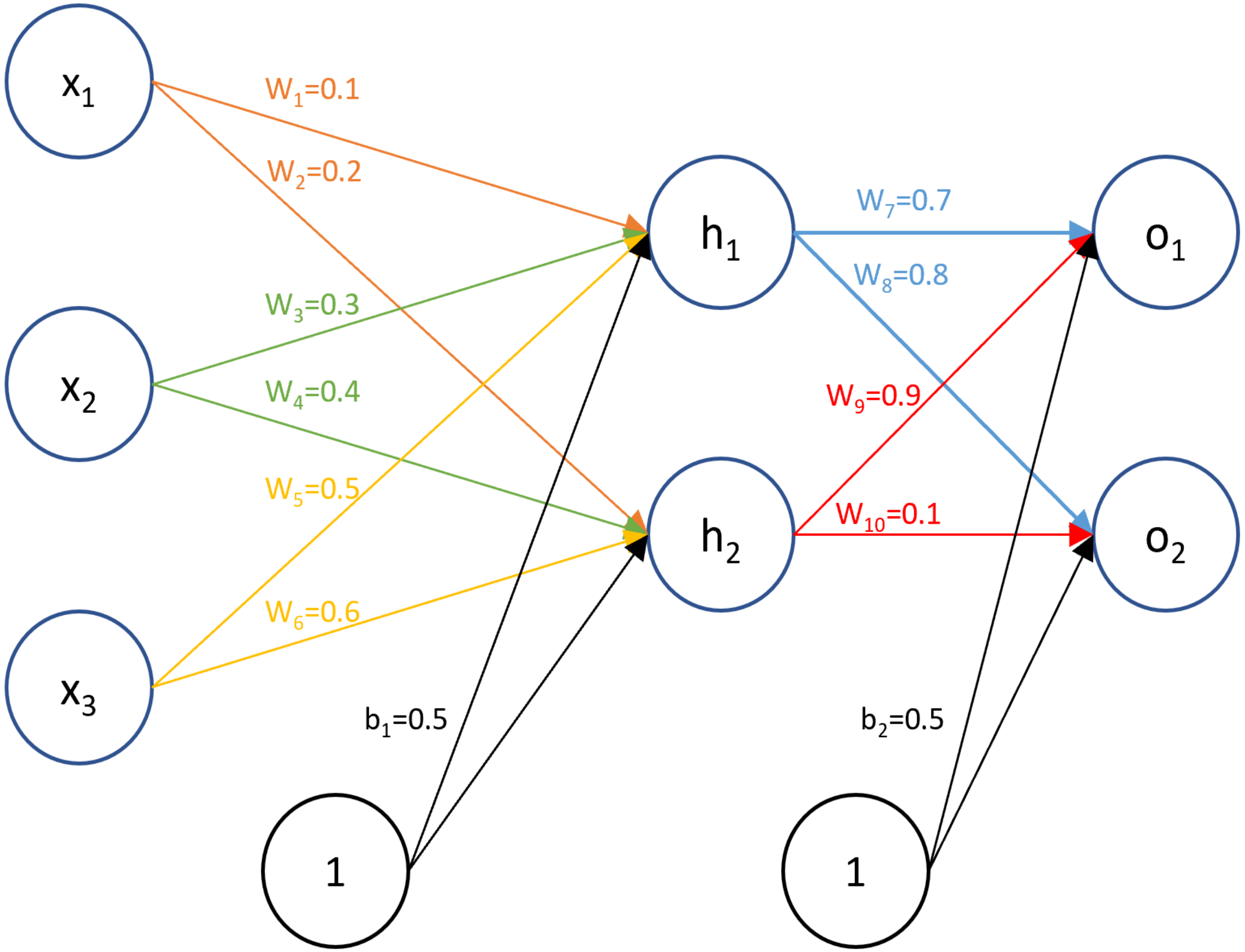
$$w_8 := w_8 - \alpha \frac{dE}{dw_8} = 0.8 - (0.01)(0.1183) = 0.7988$$

$$w_9 := w_9 - \alpha \frac{dE}{dw_9} = 0.9 - (0.01)(0.0772) = 0.8992$$

$$w_{10} := w_{10} - \alpha \frac{dE}{dw_{10}} = 0.1 - (0.01)(0.1193) = 0.0988$$

$$b_1 := b_1 - \alpha \frac{dE}{db_1} = 0.5 - (0.01)(0.0008) = 0.5000$$

$$b_2 := b_2 - \alpha \frac{dE}{db_2} = 0.5 - (0.01)(0.1975) = 0.4980$$



## Example

We repeat that over and over many times until the error goes down and the parameter estimates stabilize or converge to some values.

## Example

## Resource

https://www.anotsorandomwalk.com/backpropagation-example-with-numbers-step-by-step/

# THANK YOU

**N Mehala**
**mehala@pes.edu**
+91 80 2672 1983 Extn 701