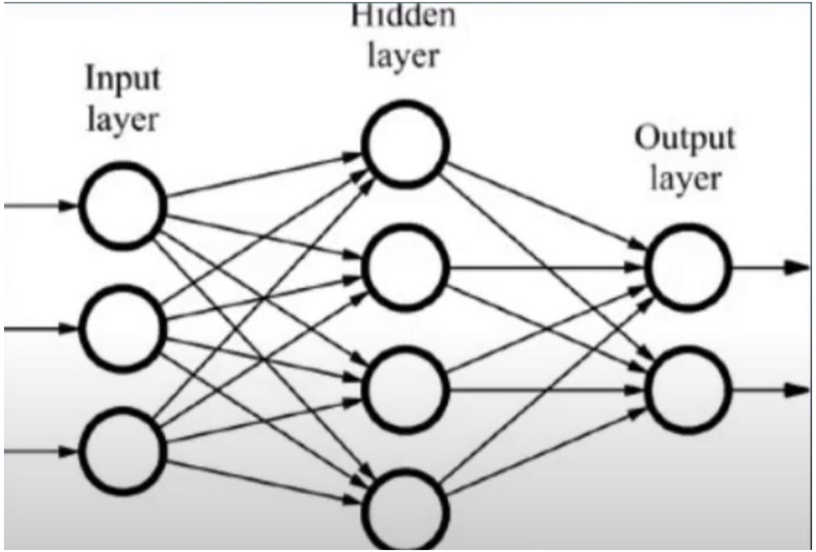
Supervised Learning
Multilayer preceptron
Backpropagation

Natthariya Laopracha

Computer Science of Mahasarakham University

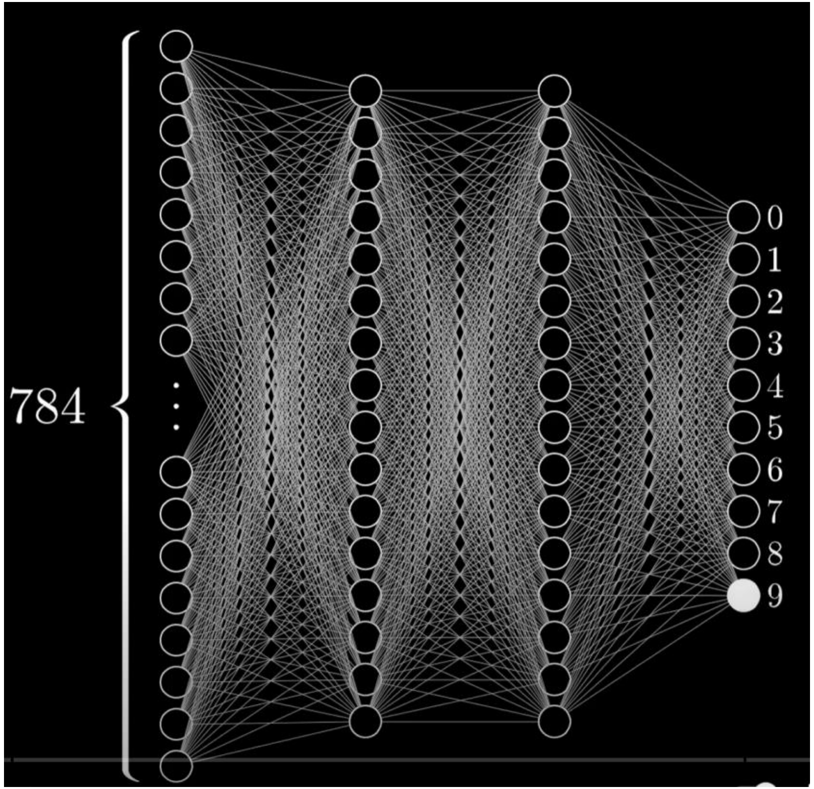
Multilayer perceptron



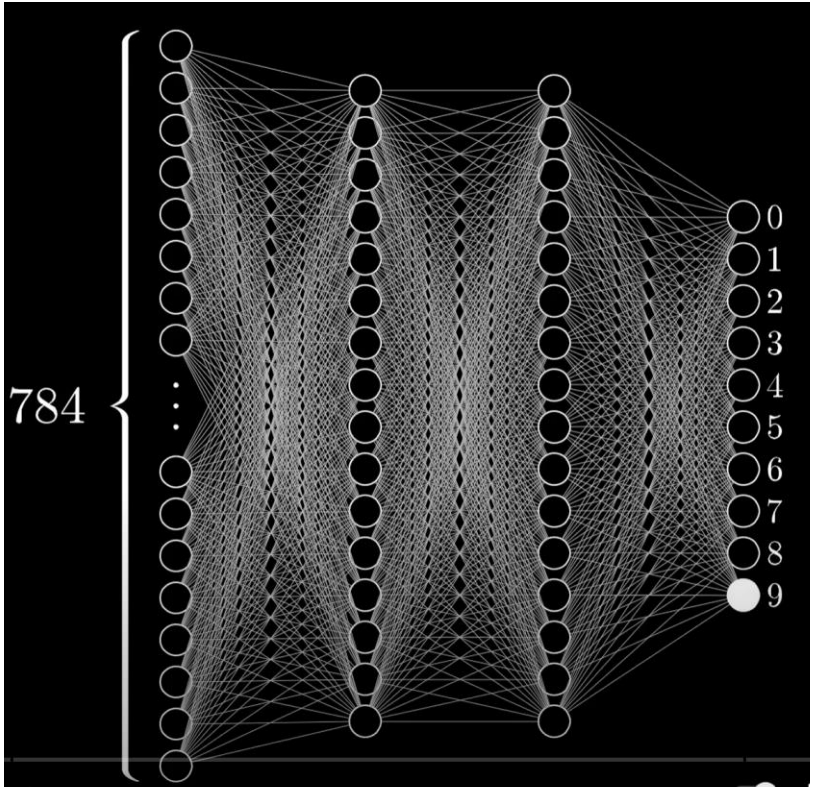
1. The input layer is connected to the hidden layers.
2. The output layer is connected to the output layer by means of interconnection weights.
3. The architecture of back propagation resembles a multi-layered feed forward network.
4. The increasing the number of hidden layers results in the computational complexity of the network.

As a result, the time taken for convergence and to minimize the error may be very high.

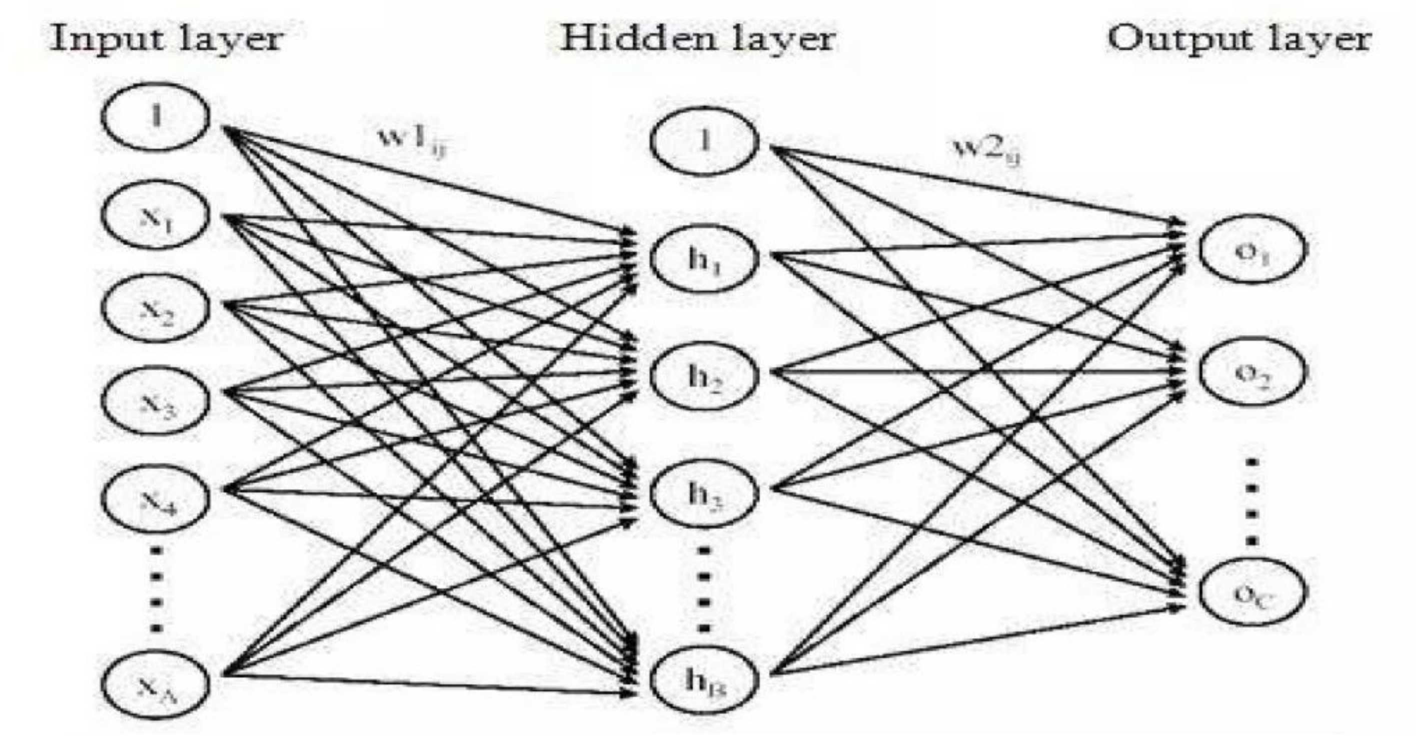
Multilayer perceptron
Backpropation for classification



Multilayer perceptron
Backpropation for classification



Multilayer perceptron
Backpropation Computation

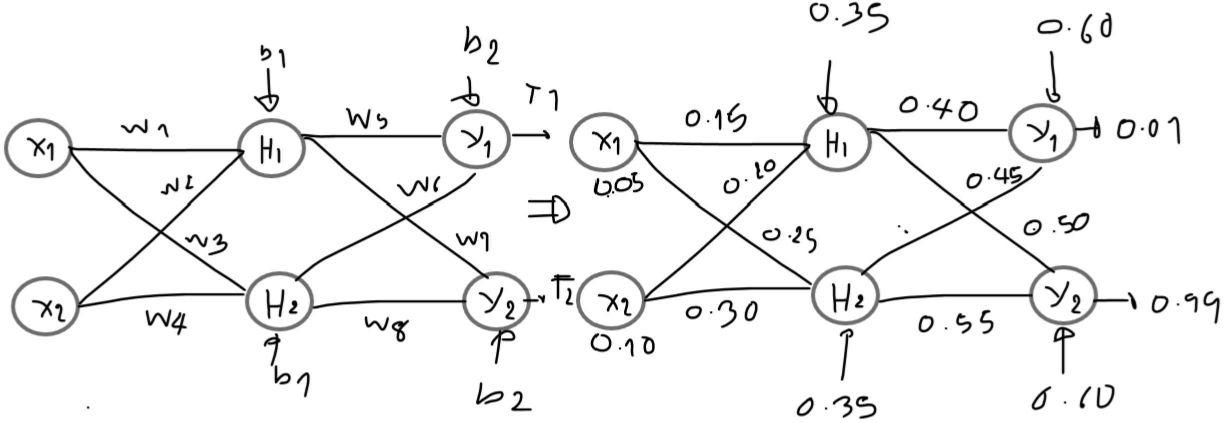


Multilayer perceptron Backpropation Training Algorithm

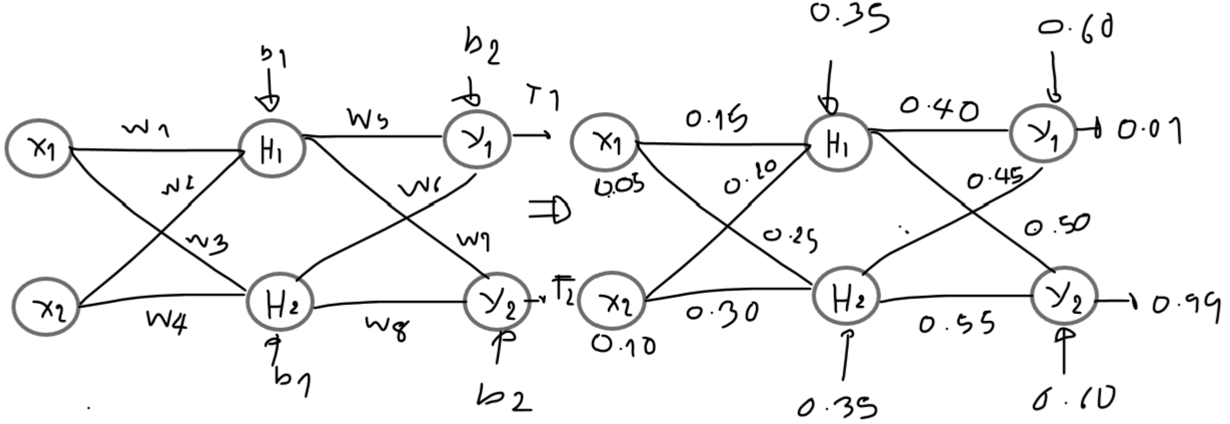
- **Initialization of weights-** some small random values are assigned.
- **Feed forward-** each input unit (X) receives an input signal and transmits this signal to each of the hidden units $Z_1, Z_2, \ldots\ldots Z_n$. Each hidden unit then calculates the activation function and sends its signal Z_i to each output unit. The output unit calculates the activation function to form the response of the given input pattern.
- **Back propagation of errors-** each output unit compares activation Y_k with its target value T_K to determine the associated error for that unit. Based on the error, the factor $\delta_K(k=1,\ldots\ldots,m)$ is computed and is used to distribute the error at output unit Y_k back to all units in the previous layer. Similarly, the factor $\delta_j(j=1,\ldots.,p)$ is compared for each hidden unit Z_j.
- **Updation of the weights and biases.**

Multilayer perceptron
Backpropation Training Algorithm

1. Initialization of weights



2. Forward

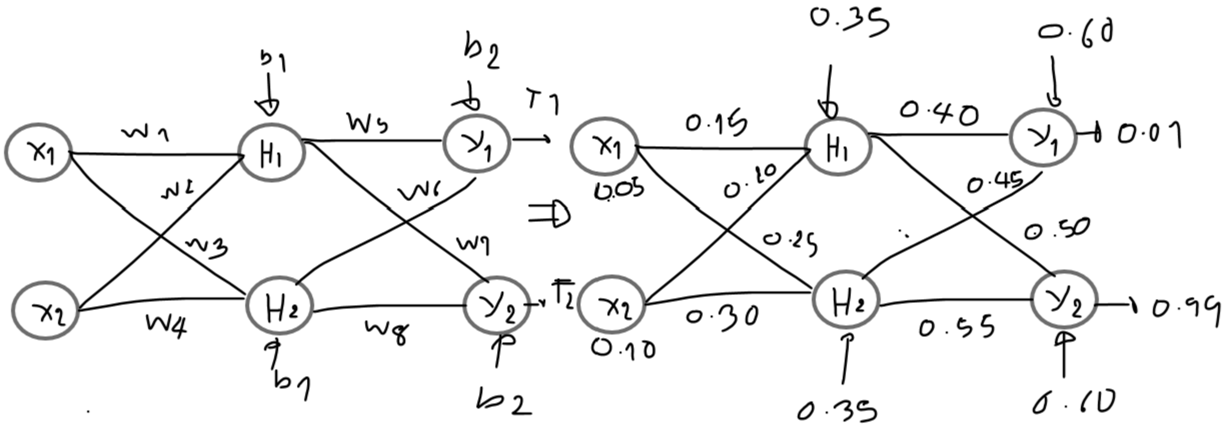


$$H_1 = x_1 * w_1 + x_2 * w_2 + b_1$$

$$= 0.05 * 0.15 + 0.10 * 0.20 + 0.35 = 0.3775$$

$$f(H_1) = \frac{1}{1+e^{-H_1}} = \frac{1}{1+e^{-0.3775}} = 0.59326992$$

2. Forward (cont.)



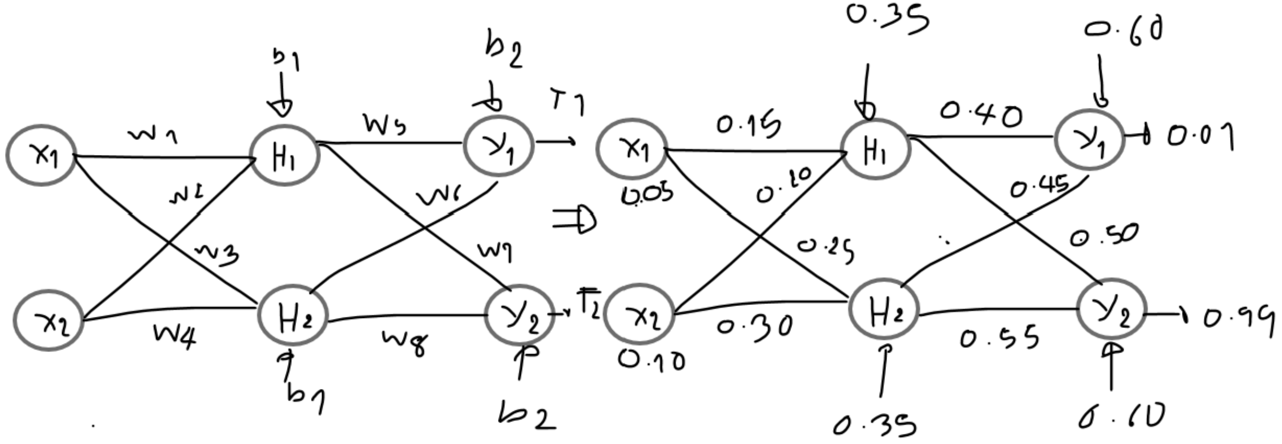
$$H_2 = x_1 * W_3 + x_2 * W_4 + b$$

$$= 0.05 \times 0.25 + 0.10 \times 0.30 + 0.35$$

$$= 0.3925$$

$$f(H_2) = 0.596884$$

2. Forward (cont.)

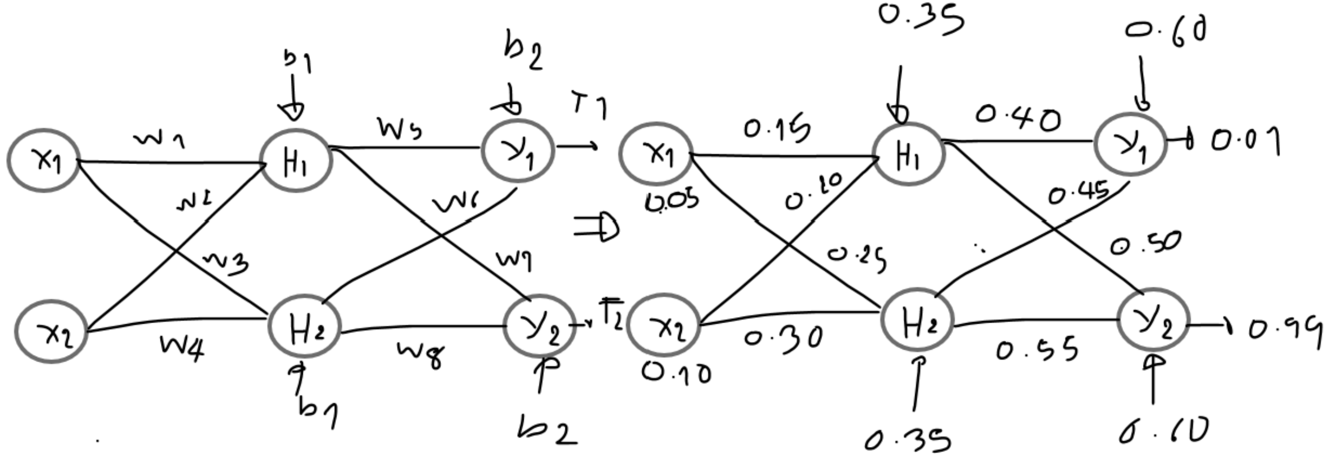


$$y_1 = f(H_1) * w_5 + f(H_2) * w_6 + b_2$$

$$= 0.5932 \times 0.40 + 0.596684 \times 0.45 + 0.60 = 1.10590$$

$$f(y_1) = \frac{1}{1+e^{-y_1}} = \frac{1}{1+e^{-1.10590}} = 0.731363$$

2. Forward (cont.)



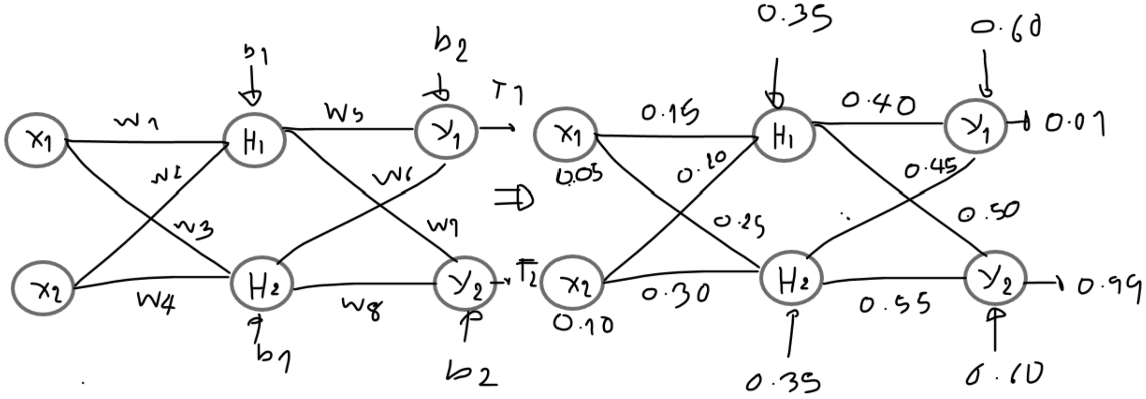
$$y_2 = f(H_1) * W_7 + f(H_2) * W_8 + b_2$$

$$= 0.59326 * 0.55 + 0.596884 * 0.55 + 0.60$$

$$= 1.224914$$

$$f(y_2) = \frac{1}{1+e^{-y_2}} = \frac{1}{1+e^{-1.224914}} = 0.772927$$

2. Forward (cont.)



Total error

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

$$= E_1 + E_2$$

$$E_1 = \frac{1}{2}(T_1 - f(y_1))^2$$

$$E_2 = \frac{1}{2}(T_2 - f(y_2))^2$$

$$= \frac{1}{2}(0.01 - 0.75136507)^2 + \frac{1}{2}(0.99 - 0.772))^2$$

$$= 0.274811083 + 0.0235 = 0.298371109$$

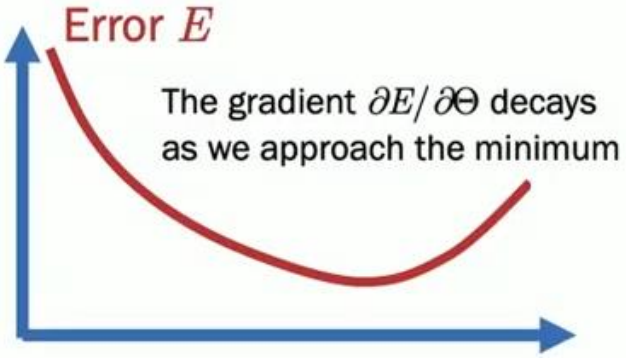
2. Backward to update weights

Chain rule

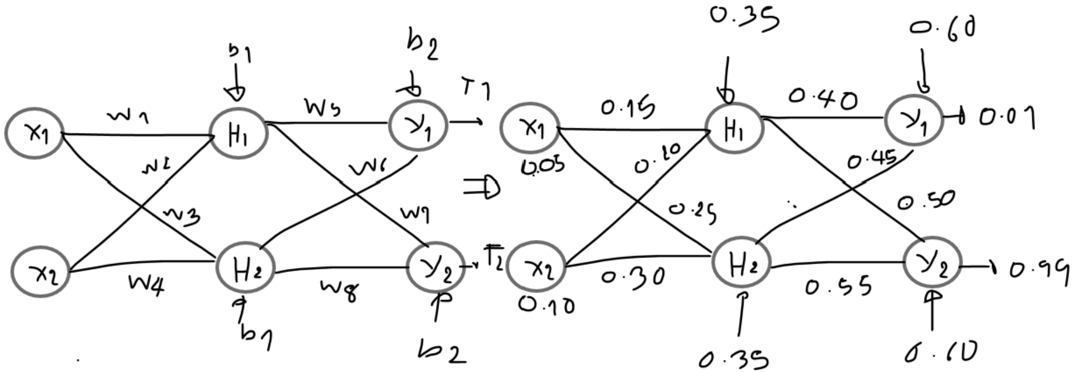
$$z = f(y)$$
$$y = f(x)$$
$$\frac{dz}{dy} = \frac{dz}{\cancel{dx}} \times \frac{\cancel{dx}}{dy}$$

Gradient

- Approximate model parameters Θ such that the prediction error E becomes minimized at optimal values Θ^*
- **Algorithm:**
  1. Set the initial value of the model parameters Θ
  2. Update the model parameters Θ w.r.t. the gradient $\partial E / \partial \Theta$ (slope) at the current point
  3. Repeat step 2 until the model parameters Θ converge (the gradient is close to zero)



2. Backward to update weights



Consider w5 compute at w5

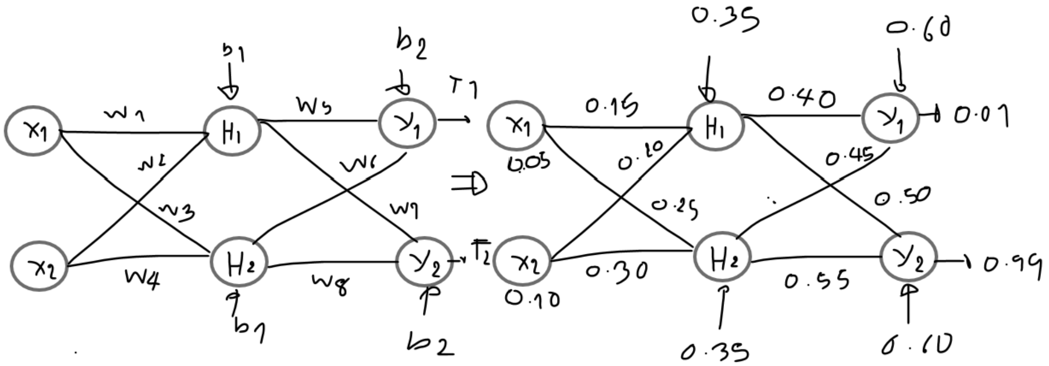
$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_5}$$

$$E_{total} = \frac{1}{2}(T_1 - f(y_1))^2 + \frac{1}{2}(T_2 - f(y_2))^2$$

$$\frac{\partial E_{total}}{\partial f(y_1)} = 2 * \frac{1}{2}(T_1 - f(y_1))^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial f(y_1)} = -(T_1 - f(y_1)) \quad = -(0.01 - 0.75136507) = 0.74136507$$

2. Backward to update weights



Consider w5 compute at w5 (cont.)

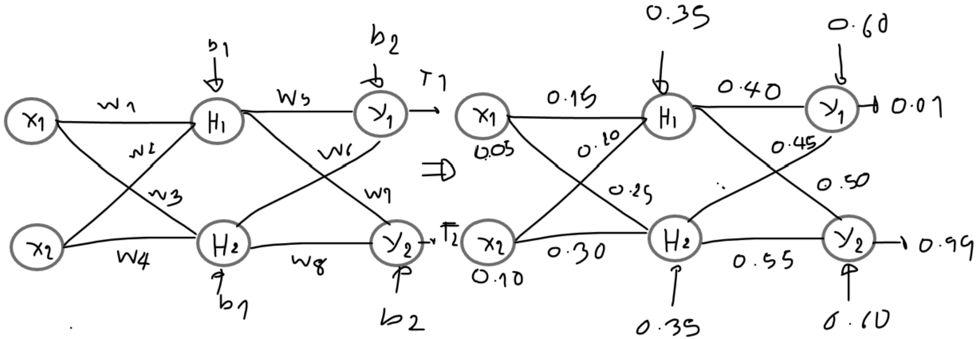
$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_5}$$

$$\frac{\partial f(y_1)}{\partial y_1} = f(y1)\big(1 - f(y1)\big) = 0.75136507(1 - 0.75136507) = 0.186815602$$

$$\frac{\partial y_1}{\partial w_5} = 1 * f(H_1) * w_5^{1-1} + 0 + 0 = f(H_1) = 0.59326992$$

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_5} = 0.74136507 * 0.186815602 * 0.59326992 = 0.082167041$$

2. Backward to update weights



Consider w5 compute at w5 (cont.)

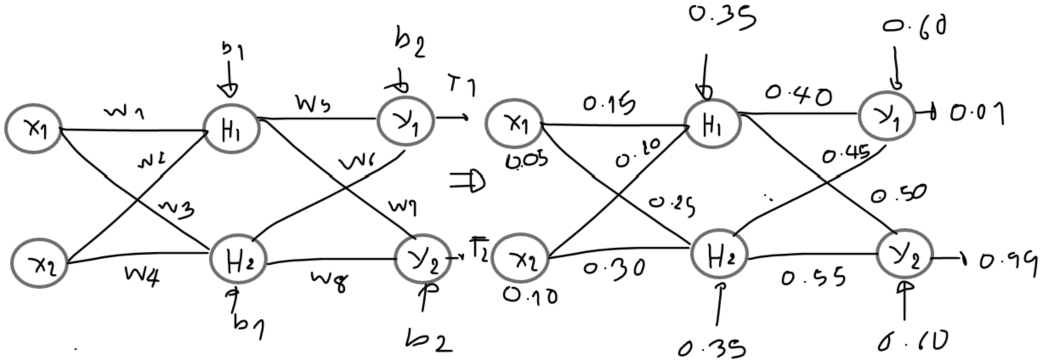
Update w5

$$W_5 = W_5 - n * \frac{\partial E_{total}}{\partial w_5}$$

Where n is learning rate =0.5

=0.4-0.5*0.82167041

2. Backward to update weights



Consider w6 compute at w6

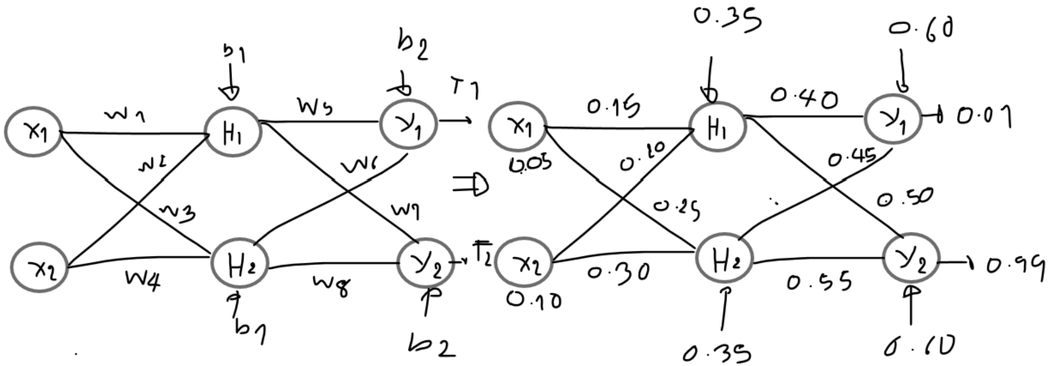
$$\frac{\partial E_{total}}{\partial w_6} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_6}$$

$$E_{total} = \frac{1}{2}(T_1 - f(y_1))^2 + \frac{1}{2}(T_2 - f(y_2))^2$$

$$\frac{\partial E_{total}}{\partial f(y_1)} = 2 * \frac{1}{2}(T_1 - f(y_1))^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial f(y_1)} = -(T_1 - f(y_1)) \quad = -(0.01 - 0.75136507) = 0.74136507$$

2. Backward to update weights



Consider w6 compute at w6 (cont.)

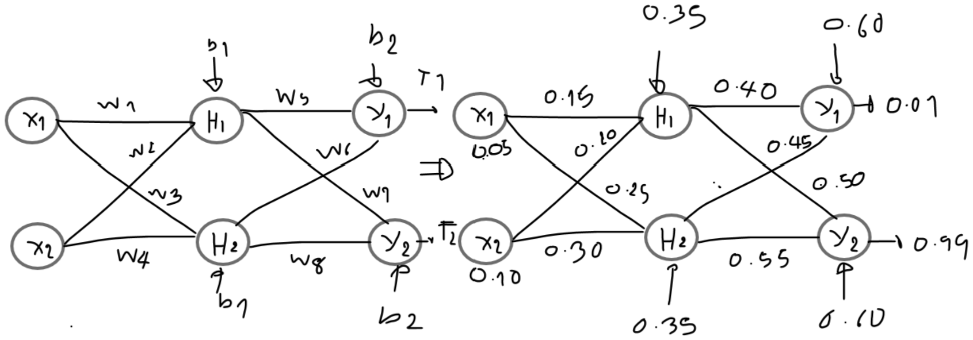
$$\frac{\partial E_{total}}{\partial w_6} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_6}$$

$$\frac{\partial f(y_1)}{\partial y_1} = f(y1)\big(1 - f(y1)\big) = 0.75136507(1 - 0.75136507) = 0.186815602$$

$$\frac{\partial y_1}{\partial w_6} = 1 * f(H_2) * w_6^{1-1} + 0 + 0 = f(H_2) = 0.596684$$

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} * \frac{\partial y_1}{\partial w_5} = 0.74136507*0.186815602*0.596884=0.082668$$

2. Backward to update weights



Consider w6 compute at w6 (cont.)

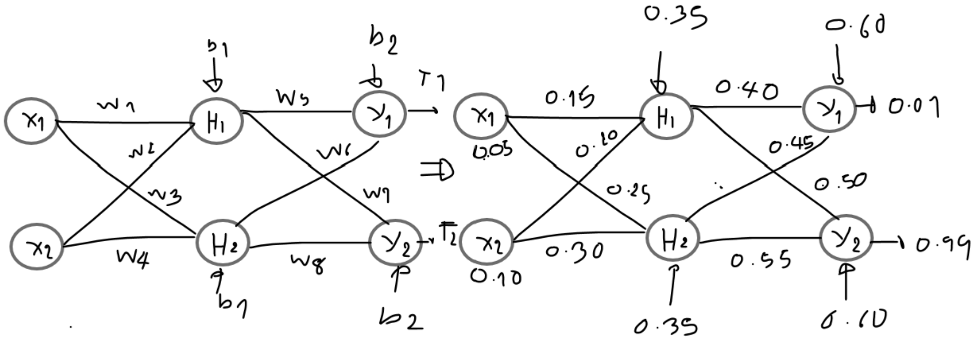
Update w6

$$W_6 = W_6 - n * \frac{\partial E_{total}}{\partial w_6}$$

Where n is learning rate =0.5

=0.45-0.5*0.82668
=0.408666

2. Backward to update weights



Consider w7 compute at w7

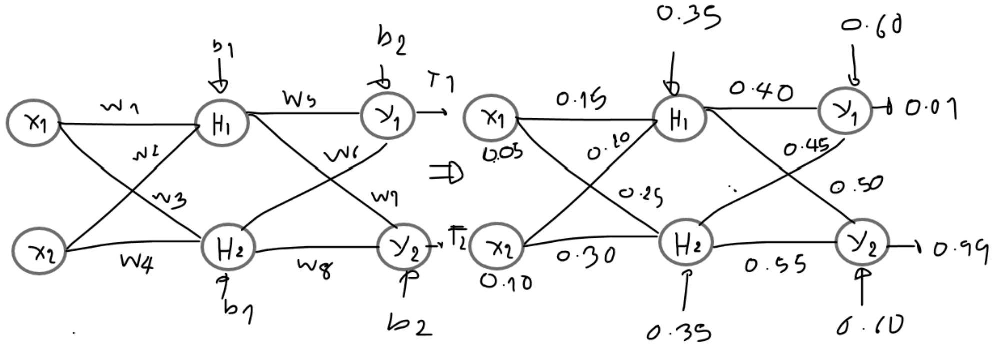
$$\frac{\partial E_{total}}{\partial w_7} = \frac{\partial E_{total}}{\partial f(y_2)} * \frac{\partial f(y_2)}{\partial y_2} * \frac{\partial y_2}{\partial w_7}$$

$$E_{total} = \frac{1}{2}(T_1 - f(y_1))^2 + \frac{1}{2}(T_2 - f(y_2))^2$$

$$\frac{\partial E_{total}}{\partial f(y_2)} = 2 * \frac{1}{2}(T_2 - f(y_2))^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial f(y_2)} = -(T_2 - f(y_2)) \quad = -(0.99 - 0.772927) = -0.21707$$

2. Backward to update weights



Consider w7 compute at w7 (cont.)

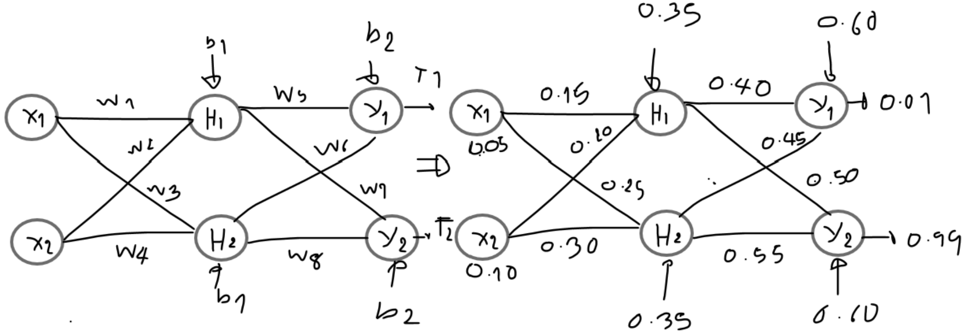
$$\frac{\partial E_{total}}{\partial w_7} = \frac{\partial E_{total}}{\partial f(y_2)} * \frac{\partial f(y_2)}{\partial y_2} * \frac{\partial y_2}{\partial w_7}$$

$$\frac{\partial f(y_2)}{\partial y_2} = f(y2)\big(1 - f(y2)\big) = 0.772927(1 - 0.772927) = 0.175511$$

$$\frac{\partial y_2}{\partial w_7} = 1 * f(H_1) * w_7^{1-1} + 0 + 0 = f(H_1) = 0.59327$$

$$\frac{\partial E_{total}}{\partial w_7} = \frac{\partial E_{total}}{\partial f(y_2)} * \frac{\partial f(y_2)}{\partial y_2} * \frac{\partial y_2}{\partial w_7} = -0.21707*0.175511*0.59327=-0.0226$$

2. Backward to update weights



Consider w7 compute at w7 (cont.)

Update w7

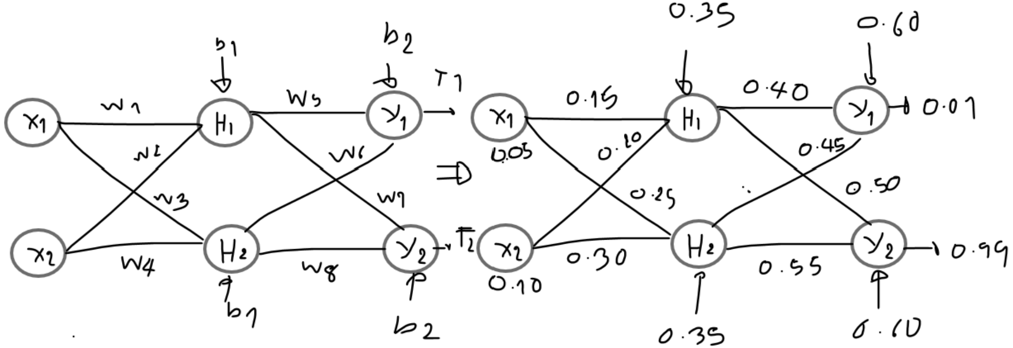
$$W_7 = W_7 - n * \frac{\partial E_{total}}{\partial w_6}$$

Where n is learning rate =0.5

=0.50-(0.5*(- 0.0226))
=0.5113

solve w8 (answer is 0.5613)

2. Backward to update weights



Hidden layer Update weights (w1,w2,w3,w4)
consider w1

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial f(H_1)} * \frac{\partial f(H_1)}{\partial H_1} * \frac{\partial H_1}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial f(H_1)} = \frac{\partial E_1}{\partial f(H_1)} + \frac{\partial E_2}{\partial f(H_1)}$$

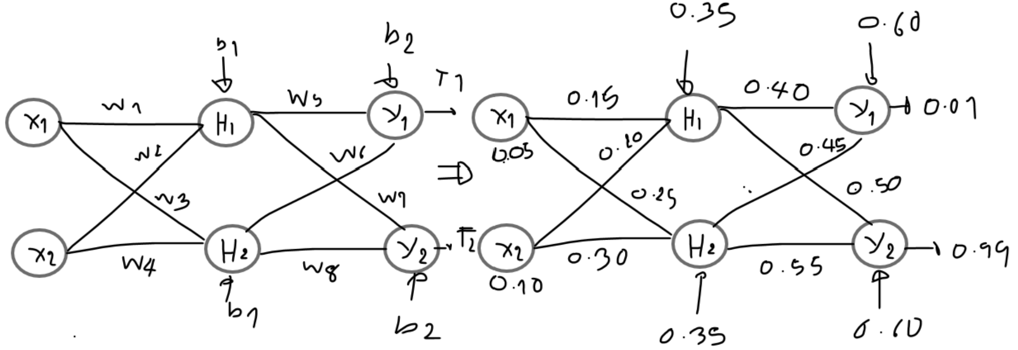
$$\frac{\partial E_1}{\partial f(H_1)} = \frac{\partial E_1}{\partial y_1} * \frac{\partial y_1}{\partial f(H_1)}$$

$$\frac{\partial E_1}{\partial y_1} = \frac{\partial E_1}{\partial f(y_1)} * \frac{\partial f(y_1)}{\partial y_1} = 0.74136507 * 0.186815602 = 0.138498562$$

$$\frac{\partial y_1}{\partial f(H_1)} = w_5 = 0.40$$

$$\frac{\partial E_1}{\partial f(H_1)} = 0.138498562 * 0.40 = 0.055399425$$

2. Backward to update weights



Hidden layer Update weights (w1,w2,w3,w4)

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial f(H_1)} * \frac{\partial f(H_1)}{\partial H_1} * \frac{\partial H_1}{\partial w_1}$$

$$\frac{\partial y_2}{\partial f(H_1)} = w_7 = 0.5$$

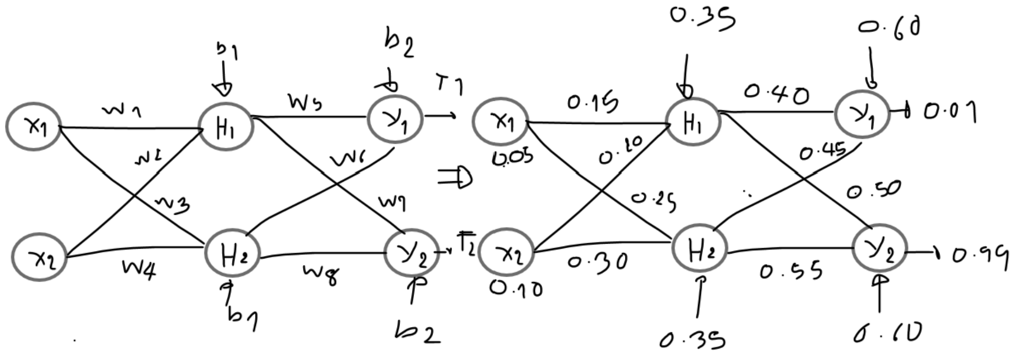
$$\frac{\partial E_{total}}{\partial f(H_1)} = \frac{\partial E_1}{\partial f(H_1)} + \frac{\partial E_2}{\partial f(H_1)}$$

$$\frac{\partial E_2}{\partial f(H_2)} = -0.0381 * 0.50 = -0.01905$$

$$\frac{\partial E_2}{\partial f(H_1)} = \frac{\partial E_2}{\partial y_2} * \frac{\partial y_2}{\partial f(H_1)}$$

$$\frac{\partial E_2}{\partial y_2} = \frac{\partial E_2}{\partial f(y_2)} * \frac{\partial f(y_2)}{\partial y_2} = -0.21707 * 0.175511 = -0.0381$$

2. Backward to update weights



Hidden layer Update weights (w1,w2,w3,w4)

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial f(H_1)} * \frac{\partial f(H_1)}{\partial H_1} * \frac{\partial H_1}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial f(H_1)} = \frac{\partial E_1}{\partial f(H_1)} + \frac{\partial E_2}{\partial f(H_1)}$$

$$\frac{\partial E_{total}}{\partial f(H_1)} = 0.055399 + (-0.0381) = 0.03635$$

$$\frac{\partial f(H_1)}{\partial H_1} = f(H_1) * (1 - f(H_1))$$

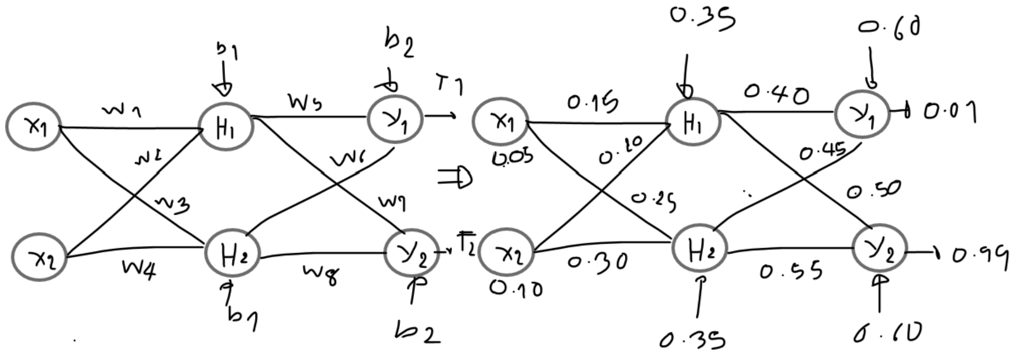
$$\frac{\partial f(H_1)}{\partial H_1} = 0.5932 * (1 - 0.5932) = 0.241300709$$

$$\frac{\partial H_1}{\partial w_1} = x_1 = 0.05$$

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial f(H_1)} * \frac{\partial f(H_1)}{\partial H_1} * \frac{\partial H_1}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial w_1} = 0.03635 * 0.241300 * 0.005 = 0.00438568$$

2. Backward to update weights



Hidden layer Update weights (w1,w2,w3,w4)

Update w1

$$W_1 = W_1 - n * \frac{\partial E_{total}}{\partial w_1}$$

$$W_1 = 0.15 - 0.5 * 0.00438$$

$$W_2 = 0.19956143$$

$$W_3 = 0.24975114$$

$$W_4 = 0.29950229$$