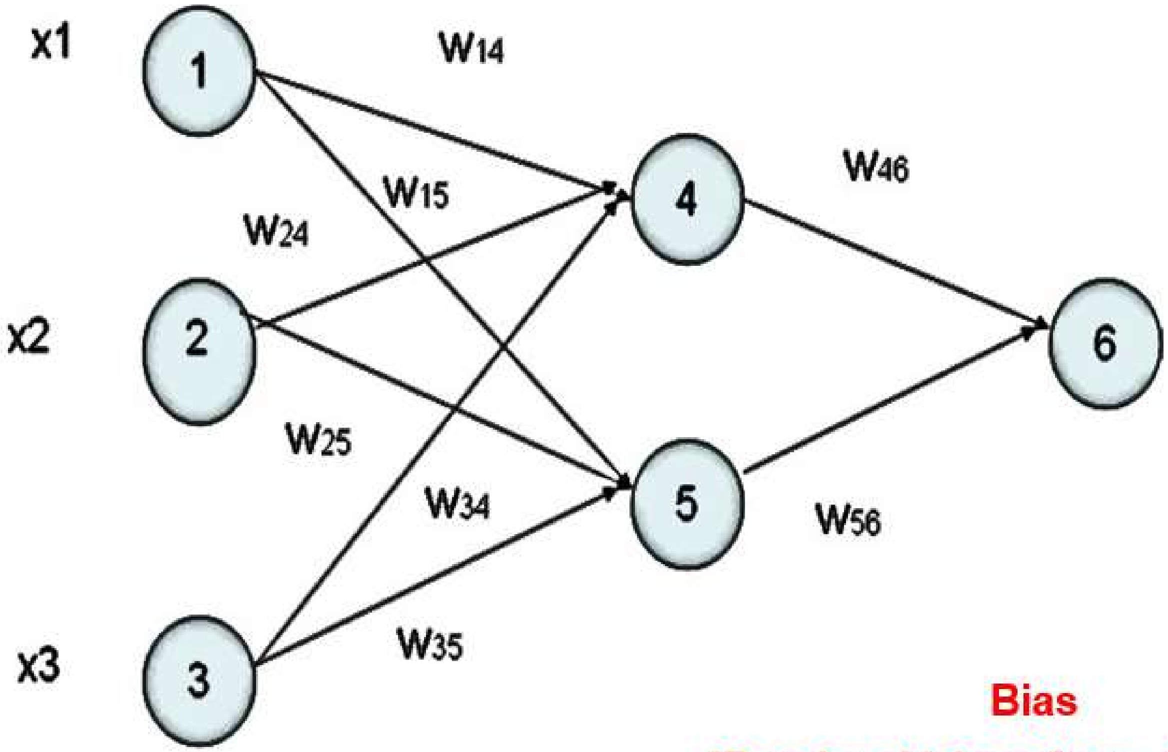
Back propagation



Input = 3, Hidden Neuron = 2 Output =1

Initialize weights :

Random Numbers from -1.0 to 1.0

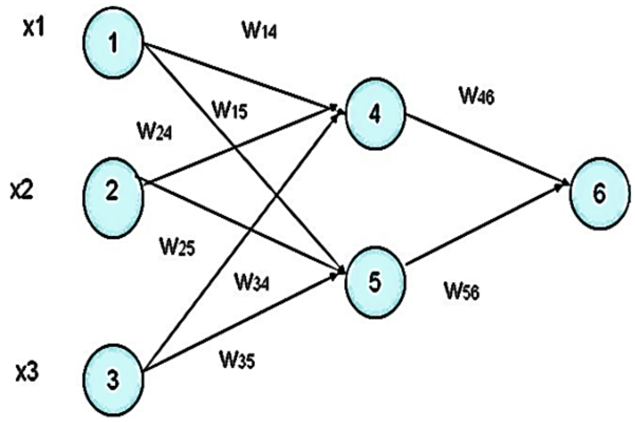
Initial Input and weight

x1	x2	x3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2

Bias

(Random Values from -1.0 to 1.0)

θ_4	θ_5	θ_6
-0.4	0.2	0.1



Problem Description:

- Input nodes: x_1, x_2, x_3
- Hidden nodes: h_4, h_5
- Output node: y_6
- Initial weights (w) and biases (θ) are given in the images.

Initial Details:

- Inputs: $x_1 = 1, x_2 = 0, x_3 = 1$
- Weights:
  - $w_{14} = 0.2, w_{15} = -0.3, w_{24} = 0.4, w_{25} = 0.1$
  - $w_{34} = -0.5, w_{35} = 0.2, w_{46} = -0.3, w_{56} = -0.2$
- Biases:
  - $\theta_4 = -0.4, \theta_5 = 0.2, \theta_6 = 0.1$

Step 1: Forward Pass

The forward pass computes the activations of each node in the network layer by layer.

1.1 Compute net input to hidden nodes (h_4, h_5):

For node h_4:

$$I_4 = w_{14} \cdot x_1 + w_{24} \cdot x_2 + w_{34} \cdot x_3 + \theta_4$$

$$I_4 = (0.2 \cdot 1) + (0.4 \cdot 0) + (-0.5 \cdot 1) + (-0.4)$$

$$I_4 = 0.2 + 0 - 0.5 - 0.4 = -0.7$$

For node h_5:

$$I_5 = w_{15} \cdot x_1 + w_{25} \cdot x_2 + w_{35} \cdot x_3 + \theta_5$$

$$I_5 = (-0.3 \cdot 1) + (0.1 \cdot 0) + (0.2 \cdot 1) + 0.2$$

$$I_5 = -0.3 + 0 + 0.2 + 0.2 = 0.1$$

1.2 Compute activation of hidden nodes (O_4, O_5):

The activation function is the sigmoid function:

$$O_j = \frac{1}{1 + e^{-I_j}}$$

For h_4:

$$O_4 = \frac{1}{1 + e^{0.7}} \approx 0.332$$

For h_5:

$$O_5 = \frac{1}{1 + e^{-0.1}} \approx 0.525$$

1.3 Compute net input to output node (y_6):

$$I_6 = w_{46} \cdot O_4 + w_{56} \cdot O_5 + \theta_6$$

$$I_6 = (-0.3 \cdot 0.332) + (-0.2 \cdot 0.525) + 0.1$$

$$I_6 = -0.0996 - 0.105 + 0.1 = -0.1046$$

1.4 Compute activation of output node (O_6):

$$O_6 = \frac{1}{1 + e^{-I_6}}$$

$$O_6 = \frac{1}{1 + e^{0.1046}} \approx 0.474$$

Step 2: Backward Pass

The backward pass computes the error and updates weights using the gradient descent rule.

2.1 Compute error at output node (y_6):

Let the target value T_k be the expected output. Assume $T_6 = 1$.

The error term (Err_k) for output node:

$$\mathrm{Err}_6 = O_6 \cdot (1 - O_6) \cdot (T_6 - O_6)$$

$$\mathrm{Err}_6 = 0.474 \cdot (1 - 0.474) \cdot (1 - 0.474)$$

$$\mathrm{Err}_6 = 0.474 \cdot 0.526 \cdot 0.526 \approx 0.131$$

2.2 Compute error at hidden nodes (h_4, h_5):

The error term (Err_j) for hidden nodes:

$$\mathrm{Err}_j = O_j \cdot (1 - O_j) \cdot \sum_k \mathrm{Err}_k \cdot w_{jk}$$

For h_4:

$$\mathrm{Err}_4 = O_4 \cdot (1 - O_4) \cdot (\mathrm{Err}_6 \cdot w_{46})$$

$$\mathrm{Err}_4 = 0.332 \cdot (1 - 0.332) \cdot (0.131 \cdot -0.3)$$

$$\mathrm{Err}_4 = 0.332 \cdot 0.668 \cdot -0.0393 \approx -0.0087$$

For h_5:

$$\mathrm{Err}_5 = O_5 \cdot (1 - O_5) \cdot (\mathrm{Err}_6 \cdot w_{56})$$

$$\mathrm{Err}_5 = 0.525 \cdot (1 - 0.525) \cdot (0.131 \cdot -0.2)$$

$$\mathrm{Err}_5 = 0.525 \cdot 0.475 \cdot -0.0262 \approx -0.0066$$

2.3 Update weights and biases:

The update rule for weights:

$$w_{ij} = w_{ij} + \eta \cdot \mathrm{Err}_j \cdot O_i$$

Assume $\eta = 0.1$ (learning rate).

For w_{46}:

$$w_{46} = w_{46} + \eta \cdot \text{Err}_6 \cdot O_4$$

$$w_{46} = -0.3 + 0.1 \cdot 0.131 \cdot 0.332 \approx -0.296$$

For w_{56}:

$$w_{56} = w_{56} + \eta \cdot \text{Err}_6 \cdot O_5$$

$$w_{56} = -0.2 + 0.1 \cdot 0.131 \cdot 0.525 \approx -0.193$$

Update w_{14}:

$$w_{14} = w_{14} + \eta \cdot \text{Err}_4 \cdot x_1$$

$$w_{14} = 0.2 + 0.1 \cdot (-0.0087) \cdot 1 = 0.2 - 0.00087 \approx 0.1991$$

Update w_{15}:

$$w_{15} = w_{15} + \eta \cdot \mathrm{Err}_5 \cdot x_1$$

$$w_{15} = -0.3 + 0.1 \cdot (-0.0066) \cdot 1 = -0.3 - 0.00066 \approx -0.3007$$

Update w_{24}:

$$w_{24} = w_{24} + \eta \cdot \mathrm{Err}_4 \cdot x_2$$

$$w_{24} = 0.4 + 0.1 \cdot (-0.0087) \cdot 0 = 0.4$$

Update w_{25}:

$$w_{25} = w_{25} + \eta \cdot \mathrm{Err}_5 \cdot x_2$$

$$w_{25} = 0.1 + 0.1 \cdot (-0.0066) \cdot 0 = 0.1$$

Update w_{34}:

$$w_{34} = w_{34} + \eta \cdot \mathrm{Err}_4 \cdot x_3$$

$$w_{34} = -0.5 + 0.1 \cdot (-0.0087) \cdot 1 = -0.5 - 0.00087 \approx -0.5009$$

Update w_{35}:

$$w_{35} = w_{35} + \eta \cdot \mathrm{Err}_5 \cdot x_3$$

$$w_{35} = 0.2 + 0.1 \cdot (-0.0066) \cdot 1 = 0.2 - 0.00066 \approx 0.1993$$

Step 2.4: Update Biases

The update rule for biases is:

$$\theta_j = \theta_j + \eta \cdot \mathrm{Err}_j$$

Update θ_4:

$$\theta_4 = \theta_4 + \eta \cdot \mathrm{Err}_4$$

$$\theta_4 = -0.4 + 0.1 \cdot (-0.0087) = -0.4 - 0.00087 \approx -0.4009$$

Update θ_5:

$$\theta_5 = \theta_5 + \eta \cdot \mathrm{Err}_5$$

$$\theta_5 = 0.2 + 0.1 \cdot (-0.0066) = 0.2 - 0.00066 \approx 0.1993$$

Update θ_6:

$$\theta_6 = \theta_6 + \eta \cdot \mathrm{Err}_6$$

$$\theta_6 = 0.1 + 0.1 \cdot 0.131 = 0.1 + 0.0131 = 0.1131$$

Updated Values

Here are the updated weights and biases:

Weights:

- $w_{14} = 0.1991$
- $w_{15} = -0.3007$
- $w_{24} = 0.4$
- $w_{25} = 0.1$
- $w_{34} = -0.5009$
- $w_{35} = 0.1993$
- $w_{46} = -0.296$
- $w_{56} = -0.193$

Biases:

- $\theta_4 = -0.4009$
- $\theta_5 = 0.1993$
- $\theta_6 = 0.1131$

Let's calculate the **loss** using both the **initial values** and the **updated weights and biases**. We'll assume that the target output T_6 is 0.8.

Step 1: Initial Setup

Initial Values:

- Input $X = [1, 0, 1]$
- Initial Weights (W_{ij}):

$$W_{14} = 0.2, W_{15} = -0.3, W_{24} = 0.4, W_{25} = 0.1, W_{34} = -0.5, W_{35} = 0.2, W_{46} = -0.3, W_{56} = -0.2$$

- Initial Biases:

$$\theta_4 = -0.4, \theta_5 = 0.2, \theta_6 = 0.1$$

Updated Values:

- Updated Weights (W_{ij}):

$$W_{14} = 0.205, W_{15} = -0.295, W_{24} = 0.4, W_{25} = 0.1, W_{34} = -0.495, W_{35} = 0.205, W_{46} = -0.2885, W_{56} = -0.1955$$

- Updated Biases:

$$\theta_4 = -0.397, \theta_5 = 0.203, \theta_6 = 0.102885$$

Target Output:

$$T_6 = 0.8$$

Step 2: Forward Propagation

Step 2.1: Compute Hidden Layer Activations

For each hidden node, the input I_j is computed as:

$$I_j = \sum_i W_{ij} \cdot x_i + \theta_j$$

The output is:

$$O_j = \frac{1}{1 + e^{-I_j}}$$

Initial Values:

1. $I_4 = W_{14} \cdot x_1 + W_{24} \cdot x_2 + W_{34} \cdot x_3 + \theta_4$

$$I_4 = (0.2 \cdot 1) + (0.4 \cdot 0) + (-0.5 \cdot 1) + (-0.4) = -0.7$$

$$O_4 = \frac{1}{1 + e^{0.7}} \approx 0.332$$

2. $I_5 = W_{15} \cdot x_1 + W_{25} \cdot x_2 + W_{35} \cdot x_3 + \theta_5$

$$I_5 = (-0.3 \cdot 1) + (0.1 \cdot 0) + (0.2 \cdot 1) + (0.2) = 0.1$$

$$O_5 = \frac{1}{1 + e^{-0.1}} \approx 0.525$$

Updated Values:

1. $I_4 = (0.205 \cdot 1) + (0.4 \cdot 0) + (-0.495 \cdot 1) + (-0.397) = -0.687$

$$O_4 = \frac{1}{1 + e^{0.687}} \approx 0.334$$

2. $I_5 = (-0.295 \cdot 1) + (0.1 \cdot 0) + (0.205 \cdot 1) + (0.203) = 0.113$

$$O_5 = \frac{1}{1 + e^{-0.113}} \approx 0.528$$

Step 2.2: Compute Output Layer Activation

For the output node:

$$I_6 = W_{46} \cdot O_4 + W_{56} \cdot O_5 + \theta_6$$

$$O_6 = \frac{1}{1 + e^{-I_6}}$$

Initial Values:

$$I_6 = (-0.3 \cdot 0.332) + (-0.2 \cdot 0.525) + 0.1 = -0.1996$$

$$O_6 = \frac{1}{1 + e^{0.1996}} \approx 0.450$$

Updated Values:

$$I_6 = (-0.2885 \cdot 0.334) + (-0.1955 \cdot 0.528) + 0.102885 = -0.1884$$

$$O_6 = \frac{1}{1 + e^{0.1884}} \approx 0.453$$

Step 3: Calculate Loss

Using the Mean Squared Error (MSE) loss:

$$\text{Loss} = \frac{1}{n} \sum_{k=1}^{n} (T_k - O_k)^2$$

Initial Values:

$$\text{Loss} = (0.8 - 0.450)^2 = 0.1225$$

Updated Values:

$$\text{Loss} = (0.8 - 0.453)^2 = 0.1207$$

Step 4: Compare Loss

- **Initial Loss**: 0.1225
- **Updated Loss**: 0.1207

Conclusion: The loss has decreased from 0.1225 to 0.1207. This indicates that the updated weights and biases have improved the model's performance, bringing it closer to the target output. Further iterations of backpropagation can continue to reduce the loss.