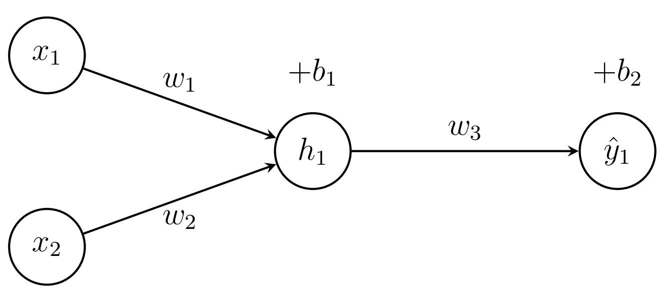
Neural Network Forward Pass and Backpropagation

Ethan Jed Carbonell

Given Information

- Inputs: $x_1 = 1.0,\ x_2 = 1.0$
- Hidden Layer Weights: $w_1 = 0.2,\ w_2 = 0.4$
- Hidden Layer Bias: $b_1 = 0.1$
- Output Layer Weight: $w_3 = 0.3$
- Output Layer Bias: $b_2 = 0.2$
- Target Output: $y_1 = 0$
- Activation Function: ReLU
- Loss Function: Mean Squared Error (MSE) with $n = 2$
- Learning Rate: $\alpha = 0.8$

Network Architecture



a) Forward Pass and Loss Function

Step 1: Hidden Layer Output

$$z_1 = w_1x_1 + w_2x_2 + b_1 = 0.2(1.0) + 0.4(1.0) + 0.1 = 0.7$$
$$h_1 = \text{ReLU}(z_1) = \max(0, 0.7) = 0.7$$

Step 2: Output Layer

$$z_2 = w_3 h_1 + b_2 = 0.3(0.7) + 0.2 = 0.41$$
$$\hat{y}_1 = z_2 = 0.41$$

Step 3: Loss Function

$$L = \frac{1}{2}(y_1 - \hat{y}_1)^2 = \frac{1}{2}(0 - 0.41)^2 = 0.08405$$

$$h_1 = 0.7, \quad \hat{y}_1 = 0.41, \quad L = 0.08405$$

b) Chain Rule Derivatives

Output Layer Weight w_3

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_3} = 0.41 \cdot 1 \cdot 0.7 = 0.287$$

Hidden Layer Weight w_1

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}_1} \cdot \frac{\partial \hat{y}_1}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$
$$= 0.41 \cdot 1 \cdot 0.3 \cdot 1 \cdot 1 = 0.123$$

Hidden Layer Weight w_2

$$\frac{\partial L}{\partial w_2} = 0.41 \cdot 1 \cdot 0.3 \cdot 1 \cdot 1 = 0.123$$

Summary of Derivatives

$$\frac{\partial L}{\partial w_1} = 0.123, \quad \frac{\partial L}{\partial w_2} = 0.123, \quad \frac{\partial L}{\partial w_3} = 0.287$$

c) Weight Updates

$$w_{\text{new}} = w_{\text{old}} - \alpha \cdot \frac{\partial L}{\partial w}$$

$$w_1^{\text{new}} = 0.2 - 0.8(0.123) = 0.1016$$
$$w_2^{\text{new}} = 0.4 - 0.8(0.123) = 0.3016$$
$$w_3^{\text{new}} = 0.3 - 0.8(0.287) = 0.0704$$

d) Second Forward Pass and Loss Comparison

Step 1: Hidden Layer Output

$$z_1^{\text{new}} = 0.1016(1) + 0.3016(1) + 0.1 = 0.5032$$
$$h_1^{\text{new}} = \text{ReLU}(0.5032) = 0.5032$$

Step 2: Output Layer

$$z_2^{\text{new}} = 0.0704(0.5032) + 0.2 = 0.2354$$
$$\hat{y}_1^{\text{new}} = 0.2354$$

Step 3: New Loss

$$L^{\text{new}} = \frac{1}{2}(0 - 0.2354)^2 = 0.0277$$

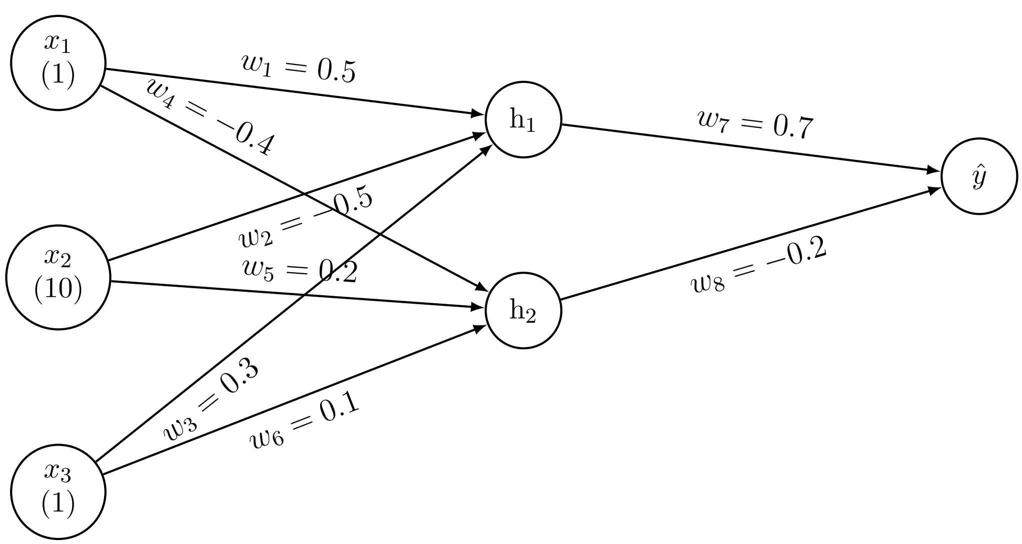
Comparison

$$L_{\text{initial}} = 0.08405, \quad L_{\text{new}} = 0.0277$$
$$\Delta L = 0.08405 - 0.0277 = 0.05635$$
$$\% \text{ Improvement} = \frac{0.05635}{0.08405} \times 100\% \approx 67.04\%$$

Conclusion

The loss decreased significantly after one gradient descent step, confirming that the weight updates moved the network closer to the optimal solution.

Network Architecture Diagram



Input Layer

$$x_1 = 1, \quad x_2 = 10, \quad x_3 = 1$$

Hidden Layer

$$h_1 \leftarrow w_1 x_1 + w_2 x_2 + w_3 x_3 + b_1$$
$$h_2 \leftarrow w_4 x_1 + w_5 x_2 + w_6 x_3 + b_2$$

Weights

$$\begin{aligned}
&w_1 = 0.5, \quad w_2 = -0.5, \quad w_3 = 0.3 && \text{(inputs to } h_1\text{)} \\
&w_4 = -0.4, \quad w_5 = 0.2, \quad w_6 = 0.1 && \text{(inputs to } h_2\text{)} \\
&w_7 = 0.7 && (h_1 \to \text{output}) \\
&w_8 = -0.2 && (h_2 \to \text{output})
\end{aligned}$$

Biases

$$b_1 = 0, \quad b_2 = 0, \quad b_{\text{out}} = 0$$

—

Forward Pass Equations

Hidden Layer Pre-activations

$$z_1 = w_1 x_1 + w_2 x_2 + w_3 x_3 + b_1 = 0.5(1) + (-0.5)(10) + 0.3(1) = -4.2$$
$$z_2 = w_4 x_1 + w_5 x_2 + w_6 x_3 + b_2 = (-0.4)(1) + 0.2(10) + 0.1(1) = 1.7$$

Hidden Activations

$$h_1 = f(z_1) = f(-4.2), \quad h_2 = f(z_2) = f(1.7)$$

Output Layer

$$z_{\text{out}} = w_7 h_1 + w_8 h_2 + b_{\text{out}}$$
$$\hat{y} = g(z_{\text{out}})$$

—

Chain Rule Derivatives

1. Derivative w.r.t. w_7

Path: $L \to \hat{y} \to z_{\text{out}} \to w_7$

$$\frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot h_1$$

—

2. Derivative w.r.t. w_2

Path: $L \to \hat{y} \to z_{\text{out}} \to h_1 \to z_1 \to w_2$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot w_7 \cdot f'(z_1) \cdot x_2$$

Substituting known values:

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot 0.7 \cdot f'(-4.2) \cdot 10$$

—

3. Derivative w.r.t. w_6

Path: $L \to \hat{y} \to z_{\text{out}} \to h_2 \to z_2 \to w_6$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot w_8 \cdot f'(z_2) \cdot x_3$$

Substituting known values:

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot (-0.2) \cdot f'(1.7) \cdot 1$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot (-0.2) \cdot f'(1.7)$$

—

Summary

$$\frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot h_1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot 0.7 \cdot f'(-4.2) \cdot 10$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial \hat{y}} \cdot g'(z_{\text{out}}) \cdot (-0.2) \cdot f'(1.7)$$

Where:

$$f'(z) = \text{derivative of hidden activation at } z, \quad g'(z) = \text{derivative of output activation at } z$$