Simple Backpropagation Example with One Neuron and 3 Layers

Network Structure

- **Input Layer**: 1 input neuron.
- **Hidden Layer**: 1 hidden neuron with ReLU activation.
- **Output Layer**: 1 output neuron.

Parameters

$$\text{Input: } X = 2$$

$$\text{Weights:} \quad W_1 \text{ for input-to-hidden layer} = 0.5, \quad W_2 \text{ for hidden-to-output layer} = 0.8$$

$$\text{Target output: } Y = 1.5$$

$$\text{Learning rate: } \eta = 0.1$$

Step 1: Forward Pass

First, we calculate the output through the network.

Hidden Layer Calculation (with ReLU activation)

$$z_1 = X \times W_1 = 2 \times 0.5 = 1$$

Since we are using **ReLU**, we apply the activation function:

$$a_1 = \text{ReLU}(z_1) = \max(0, 1) = 1$$

Output Layer Calculation

$$z_2 = a_1 \times W_2 = 1 \times 0.8 = 0.8$$

(Since there is no activation function in the output layer, $a_2 = z_2$).

Thus, the predicted output is:

$$\hat{Y} = 0.8$$

Step 2: Compute the Loss

The loss function is the **mean squared error (MSE)**:

$$L = \frac{1}{2}(Y - \hat{Y})^2 = \frac{1}{2}(1.5 - 0.8)^2 = \frac{1}{2}(0.7)^2 = 0.245$$

Step 3: Backpropagation

Now we compute the gradients for each parameter using the chain rule.

1. Gradient of the Loss with Respect to the Output Layer (Weight W_2)

The gradient of the loss with respect to the output neuron W_2 is:

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial \hat{Y}} \times \frac{\partial \hat{Y}}{\partial W_2}$$

First, compute the derivative of the loss with respect to the output:

$$\frac{\partial L}{\partial \hat{Y}} = \hat{Y} - Y = 0.8 - 1.5 = -0.7$$

Now, compute the derivative of the output with respect to W_2 (which is just a_1):

$$\frac{\partial \hat{Y}}{\partial W_2} = a_1 = 1$$

Thus, the gradient for W_2 is:

$$\frac{\partial L}{\partial W_2} = (-0.7) \times 1 = -0.7$$

2. Gradient of the Loss with Respect to the Hidden Layer (Weight W_1)

Next, we compute the gradient for W_1. We use the chain rule again:

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial \hat{Y}} \times \frac{\partial \hat{Y}}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial W_1}$$

First, we know from earlier:

$$\frac{\partial L}{\partial \hat{Y}} = -0.7$$

Now, compute the derivative of $\hat{Y}$ with respect to a_1:

$$\frac{\partial \hat{Y}}{\partial a_1} = W_2 = 0.8$$

Next, compute the derivative of a_1 with respect to z_1 (the input to the ReLU function). Since $a_1 = \text{ReLU}(z_1)$, the derivative is:

$$\frac{\partial a_1}{\partial z_1} = 1 \quad \text{(since } z_1 = 1\text{, and ReLU is linear for positive inputs)}$$

Finally, compute the derivative of z_1 with respect to W_1:

$$\frac{\partial z_1}{\partial W_1} = X = 2$$

Thus, the gradient for W_1 is:

$$\frac{\partial L}{\partial W_1} = (-0.7) \times 0.8 \times 1 \times 2 = -1.12$$

Step 4: Update the Weights

Now, we update the weights using the gradients computed above and the learning rate $\eta = 0.1$.

Update for W_2

$$W_2^{\text{new}} = W_2 - \eta \times \frac{\partial L}{\partial W_2} = 0.8 - 0.1 \times (-0.7) = 0.8 + 0.07 = 0.87$$

Update for W_1

$$W_1^{\text{new}} = W_1 - \eta \times \frac{\partial L}{\partial W_1} = 0.5 - 0.1 \times (-1.12) = 0.5 + 0.112 = 0.612$$

Step 5: Summary of the Updates

- **Final updated weights**:

$$W_1^{\text{new}} = 0.612$$

$$W_2^{\text{new}} = 0.87$$