Derivation of Gradients for Neural Network Training

Jakkula Adishesh Balaji
AI24BTECH11016

1 Introduction

In this document, we derive the gradient calculations for the weight matrices $W^{(1)}$ and $W^{(2)}$ in order to perform gradient descent.

2 Gradient of $W^{(2)}$

The weight matrix $W^{(2)}$ connects the hidden layer to the output layer. The loss function is defined as the mean squared error (MSE):

$$C = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{1}$$

where y_i is the true label and $\hat{y}_i$ is the predicted output.

The gradient with respect to $W^{(2)}$ is given as:

$$\frac{\partial C}{\partial W^{(2)}_{k,l}} = \sum_{i=1}^{N} \frac{\partial C}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial O_i} \cdot \frac{\partial O_i}{\partial W^{(2)}_{k,l}}. \tag{2}$$

2.1 Computing Partial Derivatives

We compute each term separately:

$$\frac{\partial C}{\partial \hat{y}_i} = -\frac{(y_i - \hat{y}_i)}{N}. \tag{3}$$

Since $\hat{y}_i = \sigma(O_i)$, we have:

$$\frac{\partial \hat{y}_i}{\partial O_i} = \sigma(O_i)(1 - \sigma(O_i))$$

Since O_i is given by $\sum_{k=1}^{4} Z_{i,k} W^{(2)}_{k,1}$, it follows that:

$$\frac{\partial O_i}{\partial W^{(2)}_{k,1}} = Z_{i,k}. \tag{4}$$

2.2 Final Gradient Expression

Thus, the gradient simplifies to:

$$\frac{\partial C}{\partial W^{(2)}_{k,l}} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)\cdot \sigma(O_i)(1 - \sigma(O_i)) \cdot Z_{i,k}, \quad k \in \{1,2,3,4\}. \tag{5}$$

3 Gradient of $W^{(1)}$

The weight matrix $W^{(1)}$ connects the input layer to the hidden layer. The gradient is given by:

$$\frac{\partial C}{\partial W_{k,l}^{(1)}} = \sum_{i=1}^{N} \frac{\partial C}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial O_i} \cdot \frac{\partial O_i}{\partial Z_{i,l}} \cdot \frac{\partial Z_{i,l}}{\partial H_{i,l}} \cdot \frac{\partial H_{i,l}}{\partial W_{k,l}^{(1)}}, \quad k,l \in \{1,2,3\}. \tag{6}$$

3.1 Computing Partial Derivatives

We already computed $\frac{\partial C}{\partial \hat{y}_i}$ and $\frac{\partial \hat{y}_i}{\partial O_i}$.

Since $O_i = \sum_{l=1}^{3} W_{l,1}^{(2)} Z_{i,l}$, we get:

$$\frac{\partial O_i}{\partial Z_{i,l}} = W_{l,1}^{(2)} \tag{7}$$

Since $Z_{i,l} = \sigma(H_{i,l})$, we use the sigmoid derivative:

$$\frac{\partial Z_{i,l}}{\partial H_{i,l}} = Z_{i,l}(1 - Z_{i,l}). \tag{8}$$

Finally, for the input-hidden weight matrix:

$$\frac{\partial H_{i,k}}{\partial W_{k,l}^{(1)}} = X_{i,k}. \tag{9}$$

3.2 Final Gradient Expression

Thus, the gradient simplifies to:

$$\frac{\partial C}{\partial W_{k,l}^{(1)}} = -\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot \sigma(O_i)(1 - \sigma(O_i)) \cdot W_{l,1}^{(2)} \cdot Z_{i,l}(1 - Z_{i,l}) \cdot X_{i,k}, \quad l \in \{1,2,3\}. \tag{10}$$