Partial Derivative Calculations

Pushkar Gudla
AI24BTECH11012

1 Loss Function With Respect to $W_{k,1}^{(2)}$

We aim to compute the gradient of the loss function with respect to $W_{k,1}^{(2)}$. The gradient is given by:

$$\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}} = \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \times \frac{\partial \hat{y}_i}{\partial O_i} \times \frac{\partial O_i}{\partial W_{k,1}^{(2)}}, \quad k \in \{1, 2, 3, 4\}. \tag{1}$$

Step 1: Compute $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$

The loss function used is the Mean Squared Error (MSE):

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^{N} (Y_i - \hat{y}_i)^2. \tag{2}$$

Taking the derivative with respect to $\hat{y}_i$:

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = -\frac{1}{N}(Y_i - \hat{y}_i). \tag{3}$$

Step 2: Compute $\frac{\partial \hat{y}_i}{\partial O_i}$

Since the predicted output $\hat{y}_i$ is computed using the sigmoid function:

$$\hat{y}_i = \sigma(O_i) = \frac{1}{1 + e^{-O_i}}, \tag{4}$$

its derivative with respect to O_i is:

$$\frac{\partial \hat{y}_i}{\partial O_i} = \hat{y}_i(1 - \hat{y}_i). \tag{5}$$

Step 3: Compute $\frac{\partial O_i}{\partial W_{k,1}^{(2)}}$

The output O_i is a linear combination of the hidden layer activations:

$$O_i = \sum_{k=1}^{4} Z_{i,k} W_{k,1}^{(2)}. \tag{6}$$

Taking its derivative with respect to $W_{k,1}^{(2)}$:

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

Final Expression

Substituting all the computed derivatives into the gradient equation, we obtain:

$$\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}} = \sum_{i=1}^{N} \left(-\frac{1}{N}(Y_i - \hat{y}_i) \times \hat{y}_i(1 - \hat{y}_i) \times Z_{i,k} \right).$$

This expression provides the gradient of the loss function with respect to $W_{k,1}^{(2)}$, which is used for updating the weights in backpropagation.

2 Loss Function With Respect to $W_{k,1}^{(1)}$

$$\frac{\partial \mathcal{L}}{\partial W^{(1)}k,l} = \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial \hat{y}i} \times \frac{\partial \hat{y}i}{\partial O_i} \times \frac{\partial O_i}{\partial Zi,l} \times \frac{\partial Zi,l}{\partial H_{i,l}} \times \frac{\partial H_{i,l}}{\partial W_{k,l}^{(1)}}, \quad k,l \in 1,2,3,4 \tag{8}$$

Step 1: Compute $\frac{\partial \mathcal{L}}{\partial \hat{y}i}$

From the mean squared error (MSE) loss function:

$$\mathcal{L} = \frac{1}{2N} \sum i = 1^N (Y_i - \hat{y}_i)^2 \tag{9}$$

Taking the derivative:

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = -\frac{1}{N}(Y_i - \hat{y}_i) \tag{10}$$

Step 2: Compute $\frac{\partial \hat{y}_i}{\partial O_i}$

Since $\hat{y}_i = \sigma(O_i)$ (sigmoid activation):

$$\frac{\partial \hat{y}_i}{\partial O_i} = \hat{y}_i(1 - \hat{y}_i) \tag{11}$$

Step 3: Compute $\frac{\partial O_i}{\partial Z_{i,l}}$

From the definition:

$$O_i = \sum_{k=1}^{4} Z_{i,k} W^{(2)} k, 1 \tag{12}$$

Differentiating w.r.t. Zi, l:

$$\frac{\partial O_i}{\partial Z_{i,l}} = W_{l,1}^{(2)} \tag{13}$$

Step 4: Compute $\frac{\partial Z_{i,l}}{\partial H_{i,l}}$

Since $Z_{i,l} = \sigma(H_{i,l})$, applying the derivative of the sigmoid function:

$$\frac{\partial Z_{i,l}}{\partial H_{i,l}} = Z_{i,l}(1 - Z_{i,l}) \tag{14}$$

Step 5: Compute $\frac{\partial H_{i,l}}{\partial W^{(1)} k,l}$

From the definition:

$$H = XW^{(1)} \tag{15}$$

So:

$$Hi, l = \sum_{k=1}^{3} X_{i,k} W^{(1)} k, l \tag{16}$$

Taking the partial derivative:

$$\frac{\partial Hi, l}{\partial W^{(1)} k, l} = Xi, k \tag{17}$$

Step 6: Final Expression

Substituting all the components:

$$\frac{\partial \mathcal{L}}{\partial W^{(1)} k, l} = \sum_{i=1}^{N} \left(-\frac{1}{N} (Y_i - \hat{y}i) \times \hat{y}i(1 - \hat{y}i) \times W^{(2)} l, 1 \times Zi, l(1 - Zi, l) \times X_{i,k} \right) \tag{18}$$