Partial Derivative Calculations - Assignment 2

Vijaya Sreyas Badde
AI24BTECH11003

AI1013
Programming for AI

1 Error Function With Respect to $W^{(2)}_{k,1}$

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

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

The loss function is

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

Taking its partial differentiation w.r.t. $\hat{y}_i$, we get:

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = -\frac{1}{N}\left(Y_i - \hat{y}_i\right), \quad \text{since no other term depends on } y_i.$$

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

Since y_i is computed using the sigmoid function:

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

The required derivative becomes:

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

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

From the network definition,

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

Taking the partial derivative, we get

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

Final Expression

Substituting everything back, we get

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

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

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

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

From the mean squared error (MSE) loss function:

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

Taking the derivative:

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

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

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

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

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

From the definition:

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

Differentiating w.r.t. $Z_{i,l}$:

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

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:

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

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

From the definition:

$$H = XW^{(1)}$$

So:

$$H_{i,l} = \sum_{k=1}^{3} X_{i,k} W_{k,l}^{(1)}$$

Taking the partial derivative:

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

Step 6: Final Expression

Substituting all the components:

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