Calculation of partial derivatives with respect to $W^{(1)}$ and $W^{(2)}$

B. Sai Likith Reddy
AI24BTECH11004

March 21, 2025

We have the following:

1. Loss function is $L(Y, \hat{y}) := \frac{1}{2N} \sum_{i=1}^{N} (Y_i - \hat{y}_i)^2$,
2. $\hat{y} = \sigma(O)$
3. $O_i = \sum_{k=1}^{4} Z_{i,k} W_{k,1}^{(2)}$
4. $Z_{i,l} = \sigma(H_{i,l})$
5. $H_{i,l} = \sum_{k=1}^{3} X_{i,k} W_{k,l}^{(1)}$

Partial derivatives w.r.t $W^{(2)}$

We need to find

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

For that we do the following:

1. Since $L(Y, \hat{y}) := \frac{1}{2N} \sum_{i=1}^{N} (Y_i - \hat{y}_i)^2$, the partial derivative becomes
$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{-1}{N}(Y_i - \hat{y}_i)$$
2. Since $\hat{y} = \sigma(O)$ the partial derivative becomes
$$\frac{\partial \hat{y}_i}{\partial O_i} = \hat{y}_i(1 - \hat{y}_i)$$
3. Since $O_i = \sum_{k=1}^{4} Z_{i,k} W_{k,1}^{(2)}$ the partial derivative becomes
$$\frac{\partial O_i}{\partial W_{k,1}^{(2)}} = Z_{i,k}$$

Finally the simplified expression will be

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

Partial derivatives w.r.t $W^{(1)}$

We need to find

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

For that we do the following:

1. Since $L(Y,\hat{y}) := \frac{1}{2N}\sum_{i=1}^{N}(Y_i - \hat{y}_i)^2$, the partial derivative becomes

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

2. Since $\hat{y} = \sigma(O)$ the partial derivative becomes

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

3. Since $O_i = \sum_{k=1}^{4} Z_{i,k}W_{k,1}^{(2)}$ the partial derivative becomes

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

4. Since $Z_{i,l} = \sigma(H_{i,l})$ the partial derivative becomes

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

5. Since $H_{i,l} = \sum_{k=1}^{3} X_{i,k}W_{k,l}^{(1)}$ the partial derivative becomes

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

The Final simplified expression will be

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