[FMAN45] - Assignment3

Filip Kalkan (fi1231ka-s)

May 2022

1 Exercise 1

Firstly, we simplify the equation

$$\frac{dL}{dx_i} = \sum_{l=1}^{n} \frac{dL}{dy_l} \frac{dy_l}{dx_i} \tag{1}$$

as follows.

$$\frac{dL}{dx_i} = \sum_{l=1}^{n} \frac{dL}{dy_l} \frac{d}{dx_i} \left(\sum_{j=1}^{m} W_{lj} x_j + b_j \right) = \sum_{l=1}^{n} \frac{dL}{dy_l} W_{li} \tag{2}$$

Translating the result to apply to the entire vector $\mathbf{x}$ results in

$$\frac{L}{d\mathbf{x}} = \begin{bmatrix} W_{1,1} & W_{2,1} & \dots & W_{m,1} \\ W_{1,2} & W_{2,2} & \dots & W_{m,2} \\ \vdots & \ddots & & \vdots \\ W_{1,n} & W_{2,n} & \dots & W_{m,n} \end{bmatrix} \begin{bmatrix} \frac{dL}{dy_1} \\ \frac{dL}{dy_2} \\ \vdots \\ \frac{dL}{dy_m} \end{bmatrix} = \mathbf{W}^T \frac{dL}{d\mathbf{y}} \tag{3}$$

Moving on to the equation

$$\frac{dL}{dW_{ij}} = \sum_{l=1}^{n} \frac{dL}{dy_l} \frac{dy_l}{dW_{ij}} \tag{4}$$

we can apply similar computations.

$$\frac{dL}{dW_{ij}} = \sum_{l=1}^{n} \frac{dL}{dy_l} \frac{d}{dW_{ij}} \left(\sum_{k=1}^{m} W_{lk} x_k + b_l \right) = \tag{5}$$

$$= \sum_{l=1}^{n} \frac{dL}{dy_l} \left(\sum_{k=1}^{m} x_k \frac{d}{dW_{ij}} W_{lk} + b_l \right) \tag{6}$$

Now, as $\frac{d}{dW_{ij}} W_{lk} = 1$ if $i = l$ and $j = k$ and 0 otherwise, we can simplify the equation to

$$\frac{dL}{dW_{ij}} = \frac{dL}{dy_i} x_j \tag{7}$$

Translating the result to apply to $\mathbf{W}$ results in

$$\frac{dL}{d\mathbf{W}} = \begin{bmatrix} \frac{dL}{dy_1} \\ \frac{dL}{dy_2} \\ \vdots \\ \frac{dL}{dy_m} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} = \frac{dL}{d\mathbf{y}}\mathbf{x}^T \tag{8}$$

Finally we investigate

$$\frac{dL}{db_i} = \sum_{l=1}^{n} \frac{dL}{dy_l}\frac{dy_l}{db_i}. \tag{9}$$

This can be expanded as

$$\frac{dL}{db_i} = \sum_{l=1}^{n} \frac{dL}{dy_l}\frac{d}{db_i}\left(\sum_{j=1}^{m} W_{lj}x_j + b_j\right) \tag{10}$$

Once again $\frac{db_j}{db_i} = 1$ if $i = j$ and 0 otherwise. Thus,

$$\frac{dL}{db_i} = \frac{dL}{dy_i} \tag{11}$$

which, when concidering $\mathbf{b}$ translates to

$$\frac{dL}{d\mathbf{b}} = \frac{dL}{d\mathbf{y}} \tag{12}$$

2 Exercise 2

In order to derive the expressions, we make use of the results from Exercise 1.

$$\frac{dL}{d\mathbf{X}} = \begin{bmatrix} \mathbf{W}^T \frac{dL}{d\mathbf{y}^(1)} & \mathbf{W}^T \frac{dL}{d\mathbf{y}^(2)} & \dots & \mathbf{W}^T \frac{dL}{d\mathbf{y}^(N)} \end{bmatrix} = \tag{13}$$

$$= \mathbf{W}^T \begin{bmatrix} \frac{dL}{d\mathbf{y}^(1)} & \frac{dL}{d\mathbf{y}^(2)} & \dots & \frac{dL}{d\mathbf{y}^(N)} \end{bmatrix} = \mathbf{W}^T \frac{dL}{d\mathbf{Y}} \tag{14}$$

Furthermore, we can rewrite $\mathbf{Y}$ as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{W}\mathbf{x}^(1) + \mathbf{b} & \mathbf{W}\mathbf{x}^(2) + \mathbf{b} & \dots & \mathbf{W}\mathbf{x}^(N) + \mathbf{b} \end{bmatrix} = \mathbf{W}\mathbf{X} + \mathbf{b} \tag{15}$$

where $\mathbf{b}$ is added to each column of $\mathbf{WX}$.

Using equation (7), we find that