CS 1678/2078 Homework 3 (Back prop)

Alaa Alghwiri

Question 1

1.

What is the partial derivative of the loss function with respect to the weight $W^3_{1,1}$ in the output layer?

$$\frac{\partial l(\theta)}{\partial W^3_{1,1}} = \frac{\partial h^3}{\partial W^3_{1,1}} \frac{\partial l(\theta)}{\partial h^3}$$

$$\frac{\partial h^3}{\partial W^3_{1,1}} = h^2_1$$

$$\frac{\partial l(\theta)}{\partial h^3} = \sigma$$

Accordingly:

$$\frac{\partial l(\theta)}{\partial W^3_{1,1}} = h^2_1 \times \sigma$$

2.

$$\frac{\partial l(\theta)}{\partial W^3_{1,2}} = h^2_2 \times \sigma$$

3.

$$\frac{\partial l(\theta)}{\partial W^3} = \begin{bmatrix} h^2_1 \times \sigma & h^2_2 \times \sigma \end{bmatrix}$$

4.

$$\frac{\partial l(\theta)}{\partial h^2_{1,1}} = \frac{\partial h^3}{\partial h^2_{1,1}} \frac{\partial l(\theta)}{\partial h^3}$$

$$\frac{\partial l(\theta)}{\partial h^2_{1,1}} = W^3_{1,1} \times \sigma$$

5.

$$\frac{\partial l(\theta)}{\partial h^2} = \begin{bmatrix} W^3_{1,1} \times \sigma & W^3_{1,2} \times \sigma \end{bmatrix}$$

6.

$$\frac{\partial \sigma(x)}{\partial x} =$$

$$\frac{d\sigma(x)}{dx} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

7.

$$\frac{\partial h^2_{1,1}}{\partial W^2_{1,j}} = \begin{bmatrix} \sigma(z^1_1) & \sigma(z^1_2) & \sigma(z^1_3) \end{bmatrix}$$

8.

$$\frac{\partial h^2_{1,1}}{\partial W^2_{2,j}} = \begin{bmatrix} \sigma(z^1_1) & \sigma(z^1_2) & \sigma(z^1_3) \end{bmatrix}$$

9.

$$\frac{\partial h^2_{1,1}}{\partial W^2} = \begin{bmatrix} \sigma(z^1_1) & \sigma(z^1_2) & \sigma(z^1_3) \\ \sigma(z^1_1) & \sigma(z^1_2) & \sigma(z^1_3) \end{bmatrix}$$

10.

$$\frac{\partial l(\theta)}{\partial W^2_{i,j}} = \frac{\partial l(\theta)}{\partial h^3_{i,j}} \times \frac{\partial h^3_{i,j}}{\partial h^2_{i,j}} \times \frac{\partial h^2_{i,j}}{\partial W^2_{i,j}}$$

$$\frac{\partial l(\theta)}{\partial W^2_{i,j}} = \delta W^3_{i,j} \sigma(z^1)$$

11.

$$\frac{\partial l(\theta)}{\partial W_2} = \begin{bmatrix} \delta W^3_{(1,1)}\sigma(z^1_1) & \delta W^3_{(1,1)}\sigma(z^1_2) & \delta W^3_{(1,1)}\sigma(z^1_3) \\ \delta W^3_{(1,2)}\sigma(z^1_1) & \delta W^3_{(1,2)}\sigma(z^1_2) & \delta W^3_{(1,2)}\sigma(z^1_3) \end{bmatrix}$$

12.

$$\frac{\partial h^2_1}{\partial h^1_{1,j}} = W^2_{(1,j)}$$

13.

$$\frac{\partial h_2}{\partial h_1} = \begin{bmatrix} W^2_{(1,1)} & W^2_{(1,2)} & W^2_{(1,3)} \end{bmatrix}$$

14.

$$\frac{\partial l(\theta)}{\partial h_1^{(1,j)}} = \sum_{i=1}^{n_2} \delta W^3_{(1,i)} W^2_{(i,j)}$$

15.

$$\frac{\partial l(\theta)}{\partial h_1} = \begin{bmatrix} \sum_{i=1}^{n_2} \delta W^3_{(1,i)} W^2_{(i,1)} & \sum_{i=1}^{n_2} \delta W^3_{(1,i)} W^2_{(i,2)} & \sum_{i=1}^{n_2} \delta W^3_{(1,i)} W^2_{(i,3)} \end{bmatrix}$$

16.

$$\frac{\partial l(\theta)}{\partial W^1_{(i,j)}} = \sum_{k=1}^{n_1} \left(\sum_{m=1}^{n_2} \delta W^3_{(1,m)} W^2_{(m,k)} \right)$$

17.

$$\frac{\partial l(\theta)}{\partial W_1} = \begin{bmatrix} \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,1)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,2)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,3)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,4)} \\ \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,1)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,2)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,3)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,4)} \\ \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x^{(1,1)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,2)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,3)} & \sum_{k=1}^{n_1} \frac{\partial l(\theta)}{\partial h^1_{(1,k)}} x_{(1,4)} \end{bmatrix}$$