Assignment 5

Fudi(Fred) Wang

November 2019

Part 1

$$\frac{\partial J}{\partial \boldsymbol{w}} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \boldsymbol{O}} \frac{\partial \boldsymbol{O}}{\partial M_1} \frac{\partial M_1}{\partial \boldsymbol{c}} \frac{\partial c}{\partial \boldsymbol{w}} = -\frac{1}{\hat{y}} \hat{y}(1-\hat{y}) f'(\boldsymbol{MU} + \boldsymbol{b_2}) \boldsymbol{U}^T \boldsymbol{I} f'(\boldsymbol{w} \cdot \boldsymbol{x_{i:i+h-1}} + b_1) \boldsymbol{x}^T_{\boldsymbol{i:i+h-1}}$$
$$= (\hat{y} - 1) f'(\boldsymbol{MU} + \boldsymbol{b_2}) \boldsymbol{U}^T \boldsymbol{I} f'(\boldsymbol{w} \cdot \boldsymbol{x_{i:i+h-1}} + b_1) \boldsymbol{x}^T_{\boldsymbol{i:i+h-1}}$$

The results varies by different activation functions. Say that we use ReLU as our f. Since

$$ReLU(x) = \begin{cases} 0, x \leq 0 \\ \\ x, x > 0 \end{cases}$$

So its derivative is zero when the input is less than or equal to zero and one when the input is greater than zero. Hence, when both $\boldsymbol{MU} + \boldsymbol{b_2}$ and $\boldsymbol{w} \cdot \boldsymbol{x_{i:i+h-1}} + b_1$ are greater than zero, we have

$$\frac{\partial J}{\partial \boldsymbol{w}} = (\hat{y} - 1) \boldsymbol{U}^T \boldsymbol{I} \boldsymbol{x}^T$$

Otherwise,

$$\frac{\partial J}{\partial \boldsymbol{w}} = 0$$