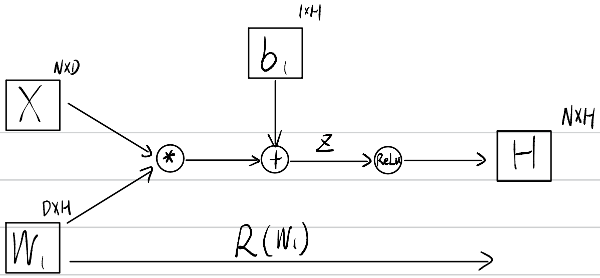
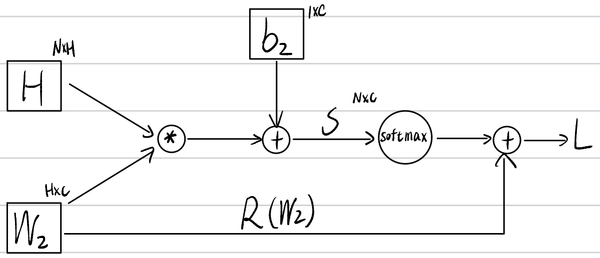
1\.

1st :

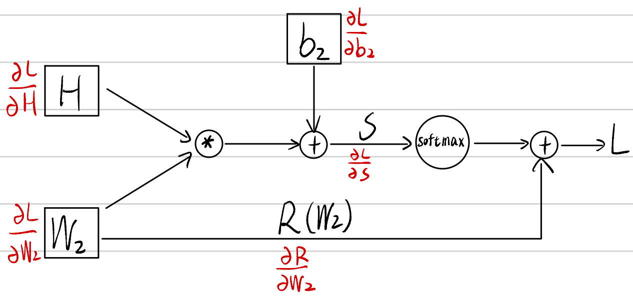


2nd :



2\.

2nd :



$$L = \frac{1}{N}\sum_{i=1}^{N} L_i + \lambda \cdot R(W)\ ;$$

$$R(W) = \left(\sum W_1^2 + \sum W_2^2\right)$$

$$\frac{\partial R}{\partial W_2} = 2 \times W_2$$

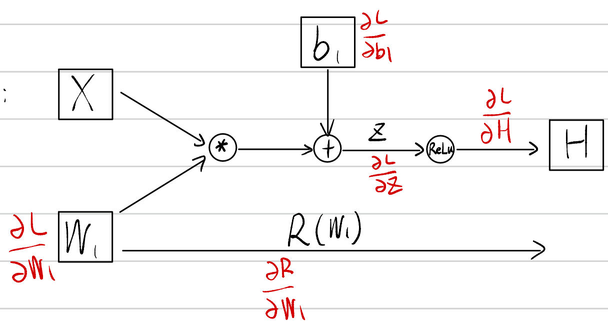
$$\frac{\partial L}{\partial S} = \begin{cases} \frac{1}{N}(P_{i,j} - 1), & j = y_i \\ \frac{1}{N}(P_{i,j}), & j \neq y_i \end{cases} \in R^{N\times C}$$

$$\frac{\partial L}{\partial b_2} = 1_N^T \frac{\partial L}{\partial S} \in R^{1\times C}$$

$$\frac{\partial L}{\partial H} = \frac{\partial L}{\partial S} W_2^T \in R^{N\times H}$$

$$\frac{\partial L}{\partial W_2} = H^T \frac{\partial L}{\partial S} + 2\cdot\lambda\cdot W_2 \in R^{H\times C}$$

1st :



$$\frac{\partial L}{\partial Z} = \begin{cases} \left[\frac{\partial L}{\partial H}\right]_{i,j}, & Z_{i,j} > 0 \\ 0, & \text{otherwise} \end{cases} \in R^{N\times H}$$

$$\frac{\partial L}{\partial b_1} = 1_N^T \cdot \frac{\partial L}{\partial Z} \in R^{1\times H}$$

$$\frac{\partial R}{\partial W_1} = 2W_1$$

$$\frac{\partial L}{\partial W_1} = X^T \frac{\partial L}{\partial Z} + 2\lambda\cdot W_1 \in R^{D\times H}$$

3.

$$g_t = \nabla L_t$$

$$m_t = \beta \cdot m_{t-1} + (1-\beta) g_t$$

$$m_0 = 0$$

$$m_1 = \beta \cdot m_0 + (1-\beta) \cdot g_1$$

$$= (1-\beta) \cdot g_1 \qquad * \ \hat{m}_1 = \frac{m_1}{1-\beta}$$

$$m_2 = \beta \cdot m_1 + (1-\beta) g_2$$

$$= \beta(1-\beta) g_1 + (1-\beta) g_2$$

$$m_3 = \beta \cdot m_2 + (1-\beta) g_3$$

$$= \beta^2 (1-\beta) g_1 + \beta (1-\beta) g_2 + (1-\beta) g_3$$

$$\vdots$$

$$m_t = (1-\beta) \sum_{i=1}^{t} \beta^{t-i} \cdot g_i$$

$$E[m_t] = (1-\beta) \sum_{i=1}^{t} \beta^{t-i} \cdot E[g_i]$$

Assume $E[g_i] = E[g]$

$$E[m_t] = (1-\beta) \sum_{i=1}^{t} \beta^{t-i} \cdot E[g]$$

$$= (1-\beta) \cdot \frac{1-\beta^t}{1-\beta} E[g]$$

$$= (1-\beta^t) \cdot E[g]$$

$$\hat{m}_t = \frac{m_t}{1-\beta^t}$$