

$$N \begin{matrix} D \\ \boxed{X_1 = X} \\ \hline \hline \end{matrix} * D \begin{matrix} H \\ \boxed{W_1} \end{matrix}$$

$$+ N \left\{ \begin{matrix} H \\ \boxed{b_1} \\ \vdots \\ \boxed{b_1} \end{matrix} \right.$$

$$= N \begin{matrix} H \\ \boxed{S_1} \\ \hline \hline \end{matrix}$$

$ReLU$

$$N \begin{matrix} H \\ \boxed{X_2} \\ \hline \hline \end{matrix} * H \begin{matrix} C \\ \boxed{W_2} \end{matrix}$$

$$+ N \left\{ \begin{matrix} C \\ \boxed{b_2} \\ \vdots \\ \boxed{b_2} \end{matrix} \right.$$

$$= N \begin{matrix} C \\ \boxed{S_2 = S} \\ \hline \hline \end{matrix}$$

$$L_i = -\ln(p_{y_i})$$

where

$$p_{y_i} = \frac{e^{s_{y_i}}}{\sum_j e^{s_j}}$$

$$s = w_2 \max(0, w_1 x_1 + b_1) + b_2$$

$$s_1 = w_1 x_1 + b_1$$

$$x_2 = s_1$$

$$s = s_2 = w_2 x_2 + b_2$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial S_2} \frac{\partial S_2}{\partial W_2}$$

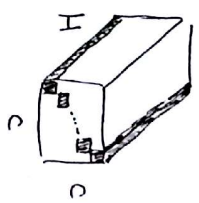
$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial S_2} \frac{\partial S_2}{\partial b_2}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial S_2} \frac{\partial S_2}{\partial X_2} \frac{\partial X_2}{\partial S_1} \frac{\partial S_1}{\partial W_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial S_2} \frac{\partial S_2}{\partial X_2} \frac{\partial X_2}{\partial S_1} \frac{\partial S_1}{\partial b_1}$$

For one data point X_{12} ,

$$\frac{\partial L}{\partial w_2} : \begin{matrix} C \\ H \end{matrix} = \frac{\partial L}{\partial s_2} : \begin{matrix} C \\ 1 \end{matrix} * \frac{\partial s_2}{\partial w_2} : C \times (H \times C) :$$



$L : \begin{matrix} 1 \\ 1 \end{matrix}$

$L : \begin{matrix} 1 \\ 1 \end{matrix}$

$s_2 : \begin{matrix} C \\ 1 \end{matrix}$

$w_2 : \begin{matrix} C \\ H \end{matrix}$

$s_2 : \begin{matrix} C \\ 1 \end{matrix}$

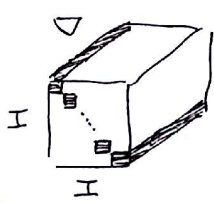
$w_2 : \begin{matrix} H \\ C \end{matrix}$

$$\frac{\partial L}{\partial b_1} : \begin{matrix} C \\ 1 \end{matrix} = \frac{\partial L}{\partial s_1} : C * \frac{\partial s_1}{\partial b_1} : C \times C$$

$L : \begin{matrix} 1 \\ 1 \end{matrix}$

$b_1 : \begin{matrix} C \\ 1 \end{matrix}$

$$\frac{\partial L}{\partial w_1} : D \times H = \frac{\partial L}{\partial s_2} : C * \frac{\partial s_2}{\partial x_2} : C \times H * \frac{\partial x_2}{\partial s_1} : H \times H * \frac{\partial s_1}{\partial w_1} : H \times (D \times H) :$$



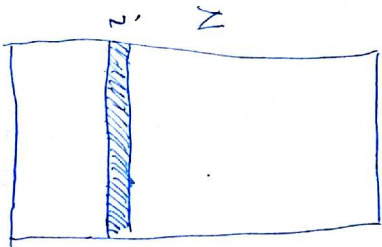
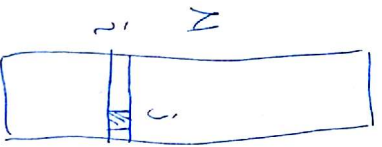
$$\frac{\partial L}{\partial b_1} : H = \frac{\partial L}{\partial s_2} : C * \frac{\partial s_2}{\partial x_2} : C \times H * \frac{\partial x_2}{\partial s_1} : H \times H * \frac{\partial s_1}{\partial b_1} : H \times H$$

According to CS231 hw1 Q3,

$$\left\{ \begin{array}{l} \frac{dL_i}{dS_i} = p_i - 1 \\ \frac{dL_i}{dS_i} = p_i \end{array} \right.$$

$$\frac{dS_j}{dW_j} = X_{2j} \text{ since}$$

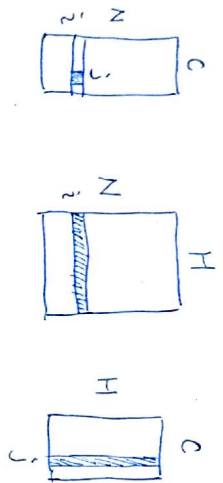
$$S_{2j} = X_{2j} * W_{2j} + b_{2j}$$



$$\frac{\partial S_{2j}}{\partial X_{2i}} = W_{2j}$$

since

$$S_{2j} = X_{2i} * W_{2j} + b_{2j}$$



$$\frac{\partial X_{2i,h}}{\partial S_{1i,h}} = \begin{cases} 1 \\ 0 \end{cases}$$

if

$$S_{1i,h} > 0$$

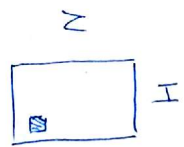
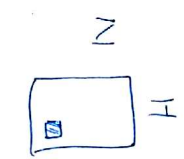
0

if

$$S_{1i,h} \leq 0$$

since

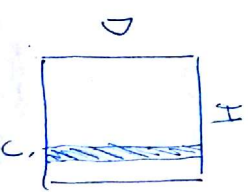
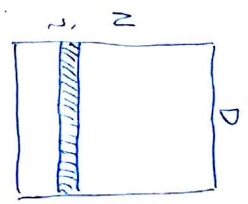
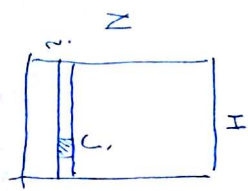
$$X_{2i,h} = \max(0, S_{1i,h})$$

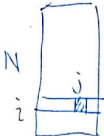
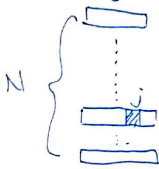


$$\frac{\partial S_{1j}}{\partial W_{1j}} = X_{1i}$$

since

$$S_{1j} = X_{1i} * W_{1j} + b_{1j}$$



$$\frac{\partial S_{2j}}{\partial b_{2j}} = 1 \quad \text{since} \quad S_{2j} = \sum_i X_{2i} W_{2j} + b_{2j}$$



$$\frac{\partial S_{1j}}{\partial b_{1j}} = 1 \quad \text{for similar reason}$$