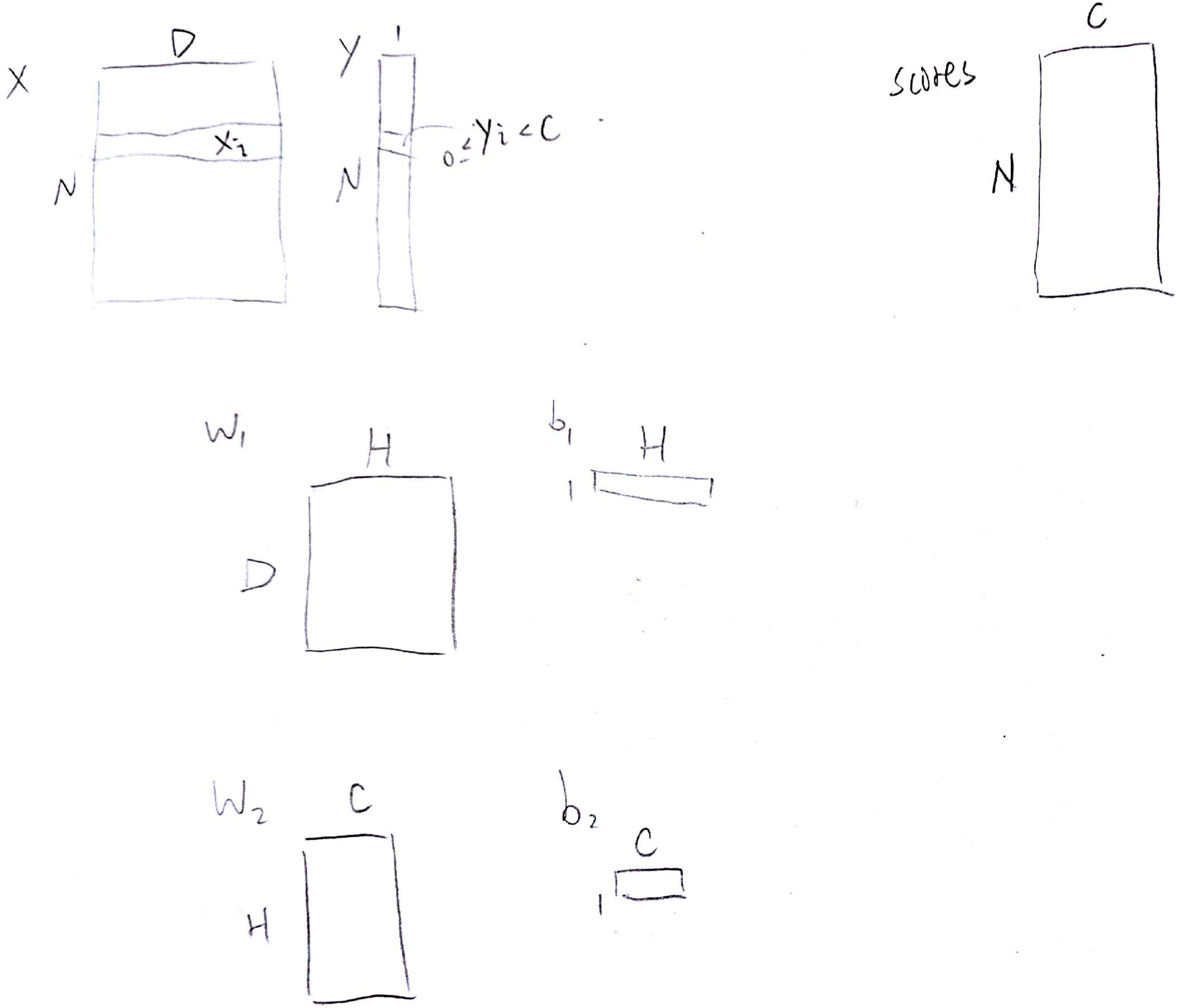
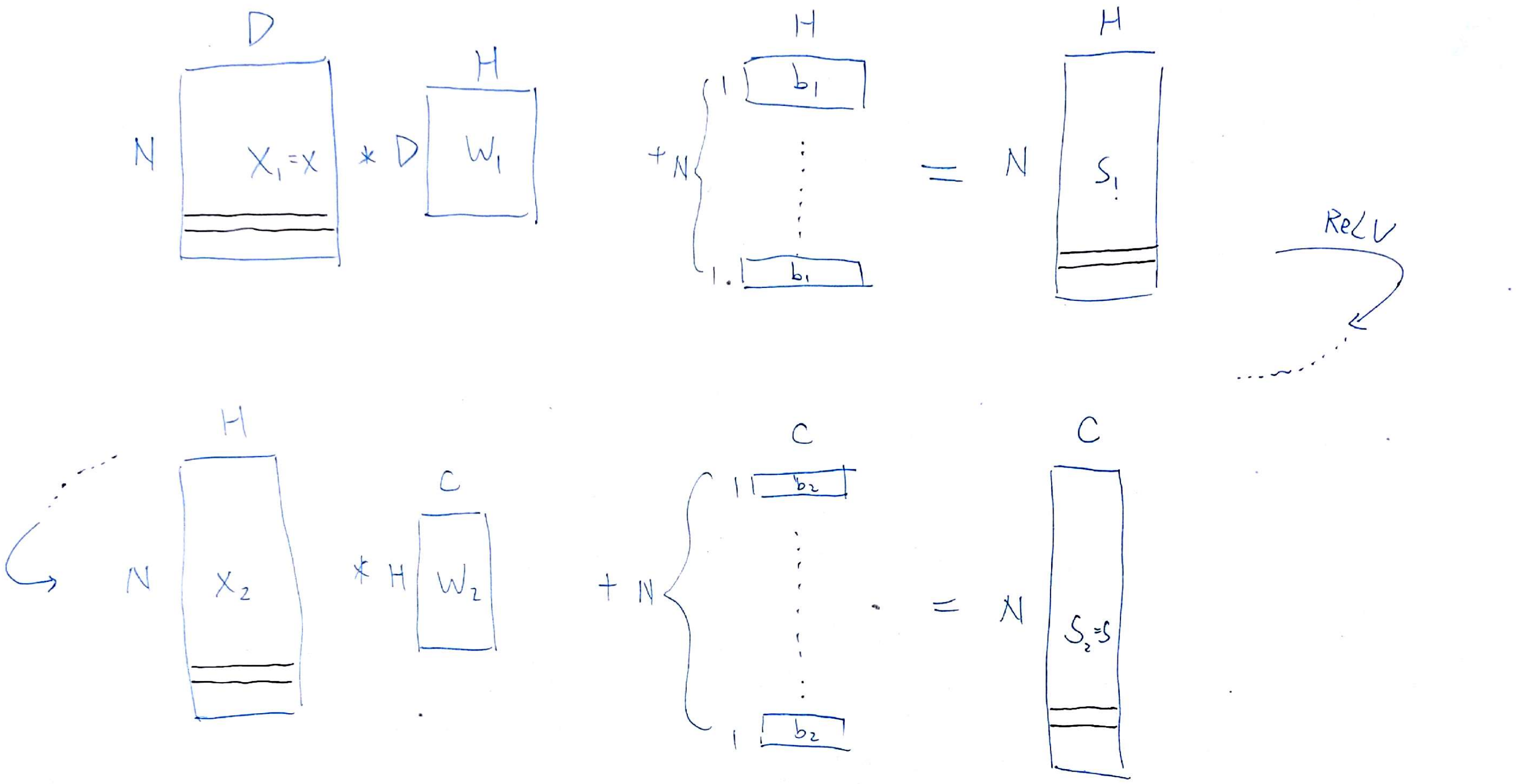


Scanned by CamScanner



Scanned by CamScanner

$$L_i = -\ln(p_{y_i}) \quad \text{where} \quad p_{y_i} = \frac{e^{s_{y_i}}}{\sum_j e^{s_j}}$$

$$s = W_2 \max(0, W_1 x_1 + b_1) + b_2$$

$$s_1 = W_1 x_1 + b_1$$

$$x_2 = s_1$$

$$s = s_2 = W_2 \; x_2 \; + b_2$$

Scanned by CamScanner

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

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

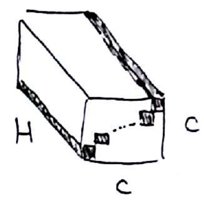
$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial S_2} \frac{\partial S_2}{\partial X_2} \frac{\partial X_2}{\partial S_1} \frac{\partial S_1}{\partial W_1}$$

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

Scanned by CamScanner

For one data point $X_1 i$,

$$\frac{\partial L}{\partial W_2} : {}_{H}\square^{C} = \frac{\partial L}{\partial S_2} : {}_{1}\square^{C} * \frac{\partial S_2}{\partial W_2} : C \times (H \times C) :$$



$L: {}_{1}\square^{1}$ $\quad$ $L: {}_{1}\square^{1}$ $\quad$ $S_i: {}_{1}\square^{C}$

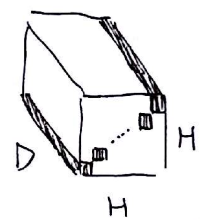
$W_2: {}_{H}\square^{C}$ $\quad$ $S_2: {}_{1}\square^{C}$ $\quad$ $W_2: {}_{H}\square^{C}$

$$\frac{\partial L}{\partial b_2} : {}_{1}\square^{C} = \frac{\partial L}{\partial S_2} : C * \frac{\partial S_2}{\partial b_2} : C \times C$$

$L: {}_{1}\square^{1}$

$b_2: {}_{1}\square^{C}$

$$\frac{\partial L}{\partial W_1} : D \times H = \frac{\partial L}{\partial S_2} : C * \frac{\partial S_2}{\partial X_2} : C \times H * \frac{\partial X_2}{\partial S_1} : H \times H * \frac{\partial S_1}{\partial W_1} : H \times (D \times H) :$$



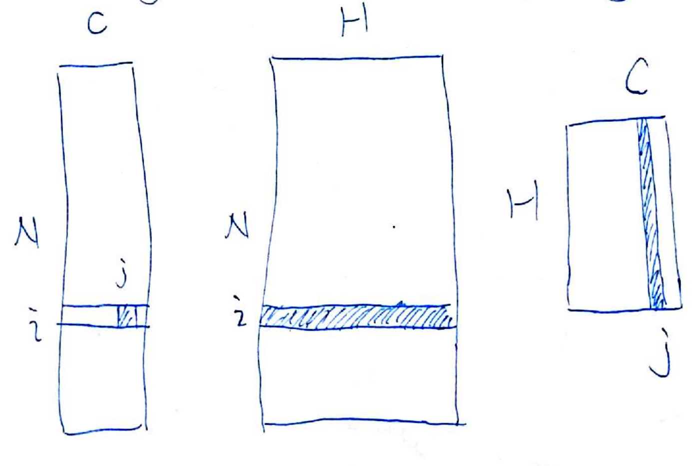
$$\frac{\partial L}{\partial b_1} : H = \frac{\partial L}{\partial S_2} : C * \frac{\partial S_2}{\partial X_2} : C \times H * \frac{\partial X_2}{\partial S_1} : H \times H * \frac{\partial S_1}{\partial b_1} : H \times H$$

Scanned by CamScanner

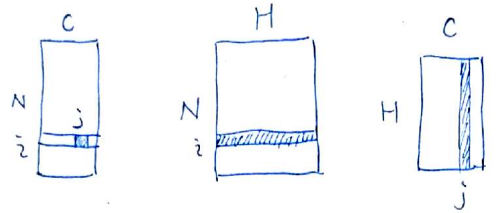
According to CS231 hw1 Q3,

$$\begin{cases} \dfrac{\partial L_i}{\partial S_{y_i}} = p_{y_i} - 1 \\[2ex] \dfrac{\partial L_i}{\partial S_j} = p_j \end{cases}$$

$$\frac{\partial S_{2j}}{\partial W_{2j}} = X_{2i} \text{ since } S_{2j} = X_{2i} * W_{2j} + b_{2j}.$$

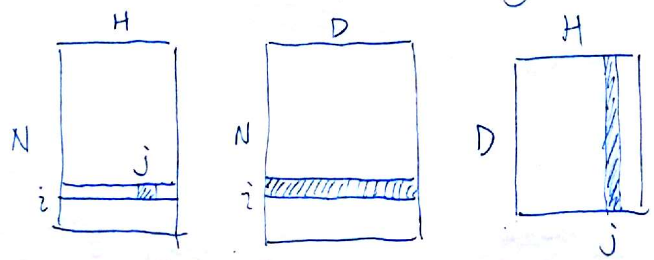


$$\frac{\partial S_{2j}}{\partial X_{2i}} = W_{2j} \quad \text{since} \quad S_{2j} = X_{2i} * W_{2j} + b_{2j}$$



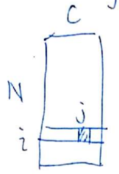
$$\frac{\partial X_{2i,h}}{\partial S_{1i,h}} = \begin{cases} 1 & \text{if } S_1 > 0 \\ 0 & \text{if } S \leq 0 \end{cases} \quad \text{since} \quad X_{2i,h} = \max(0, S_{1i,h})$$

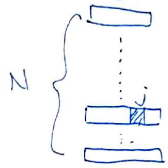
$$\frac{\partial S_{1j}}{\partial W_{1j}} = X_{1i} \quad \text{since} \quad S_{1j} = X_{1i} * W_{1j} + b_{1j}.$$



Scanned by CamScanner

$$\frac{\partial S_{2_j}}{\partial b_{2_j}} = 1 \quad \text{since} \quad S_{2_j} = X_{2_i} W_{2_j} + b_{2_j}$$





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

Scanned by CamScanner