1.

$$Y = XW \Rightarrow \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1H} \\ Y_{21} & \ddots & & \vdots \\ \vdots & & & \\ Y_{N1} & \cdots & & Y_{NH} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1D} \\ X_{21} & \ddots & & \vdots \\ \vdots & & & \\ X_{N1} & \cdots & & X_{ND} \end{bmatrix} \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1H} \\ W_{21} & \ddots & & \vdots \\ \vdots & & & \\ W_{D1} & \cdots & & W_{DH} \end{bmatrix}$$

$$\frac{\partial L}{\partial Y} = \nabla L_y = \begin{bmatrix} \frac{\partial L}{\partial Y_{11}} & \frac{\partial L}{\partial Y_{12}} & \cdots & \frac{\partial L}{\partial Y_{1H}} \\ \frac{\partial L}{\partial Y_{21}} & \ddots & & \vdots \\ \vdots & & & \\ \frac{\partial L}{\partial Y_{N1}} & \cdots & & \frac{\partial L}{\partial Y_{NH}} \end{bmatrix}$$

$$\frac{\partial L}{\partial X_{i,j}} = \frac{\partial Y}{\partial X_{i,j}} \cdot \frac{\partial L}{\partial Y}, \quad \frac{\partial Y}{\partial X_{i,j}} = \begin{bmatrix} \frac{\partial Y_{11}}{\partial X_{i,j}} & \frac{\partial Y_{12}}{\partial X_{i,j}} & \cdots & \frac{\partial Y_{1H}}{\partial X_{i,j}} \\ \frac{\partial Y_{21}}{\partial X_{i,j}} & \ddots & & \vdots \\ \vdots & & & \\ \frac{\partial Y_{N1}}{\partial X_{i,j}} & \cdots & & \frac{\partial Y_{NH}}{\partial X_{i,j}} \end{bmatrix}$$

EX:

$$\frac{\partial Y_{11}}{\partial X_{11}} = \frac{\partial (X_{11}W_{11} + X_{12}W_{21} + \cdots + X_{1D}W_{D1})}{\partial X_{11}} = W_{11}$$

$$\frac{\partial Y_{12}}{\partial X_{11}} = \frac{\partial (X_{11}W_{12} + X_{12}W_{22} + \cdots + X_{1D}W_{D2})}{\partial X_{11}} = W_{12} \qquad \Rightarrow \frac{\partial Y_{MN}}{\partial X_{ij}} = \begin{cases} W_{ij}, & \text{if } i = m \\ 0, & \text{if } i \neq m \end{cases}$$

$$\frac{\partial Y_{21}}{\partial X_{11}} = \frac{\partial (X_{21}W_{11} + X_{22}W_{21} + \cdots + X_{2D}W_{D1})}{\partial X_{11}} = 0$$

$$\Rightarrow \frac{\partial L}{\partial X_{11}} = \frac{\partial Y}{\partial X_{11}} \cdot \frac{\partial L}{\partial Y} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1H} \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 0 \end{bmatrix} \cdot \frac{\partial L}{\partial Y} = W_{11}\frac{\partial L}{\partial Y_{11}} + W_{12}\frac{\partial L}{\partial Y_{12}} + \cdots + W_{1H}\frac{\partial L}{\partial Y_{1H}}$$

$$\Rightarrow \frac{\partial L}{\partial X_{ij}} = W_{j1}\frac{\partial L}{\partial Y_{i1}} + W_{j2}\frac{\partial L}{\partial Y_{i2}} + \cdots + W_{jH}\frac{\partial L}{\partial Y_{iH}} = \left[\nabla L_y W^T\right]_{ij} \quad ※$$

2.

(a)

Scalar : from problem 2 in HW3

$$s \in R^{1\times c} \Rightarrow \frac{\partial L}{\partial s} \in R^{1\times c} \text{ and } \frac{\partial L}{\partial s_j} = \begin{cases} P_j - 1, & j = y \\ P_j, & j \neq y \end{cases}$$

Matrix : $$s \in R^{N\times C} \Rightarrow \frac{\partial L}{\partial s} \in R^{N\times C} \text{ and } \frac{\partial L}{\partial s_{ij}} = \begin{cases} P_{ij} - 1, & j = y_i \\ P_{ij}, & j \neq y_i \end{cases}$$ ※

(b)

Scalar : $$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial W_2} = \frac{\partial L}{\partial s} H$$

Matrix : $$W_2 \in R^{H\times C} \Rightarrow \overset{H\times C}{\frac{\partial L}{\partial W_2}} = \overset{N\times C}{\frac{\partial L}{\partial s}}\overset{X}{\frac{\partial s}{\partial W_2}} \text{ or } \overset{H\times C}{\frac{\partial L}{\partial W_2}} = \overset{H\times N}{\frac{\partial s}{\partial W_2}}\overset{N\times C}{\frac{\partial L}{\partial s}}$$

$$\because H \in R^{N\times H} \Rightarrow \frac{\partial L}{\partial W_2} = H^T \frac{\partial L}{\partial s}$$ ※

(c)

Scalar : $$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial b_2} = \frac{\partial L}{\partial s}\cdot 1$$

Matrix : $$b_2 \in R^{1\times C} \Rightarrow \overset{1\times C}{\frac{\partial L}{\partial b_2}} = \overset{N\times C}{\frac{\partial L}{\partial s}}\overset{X}{\frac{\partial s}{\partial b_2}} \text{ or } \overset{1\times C}{\frac{\partial L}{\partial b_2}} = \overset{1\times N}{\frac{\partial s}{\partial b_2}}\overset{N\times C}{\frac{\partial L}{\partial s}}$$

$$\because 1_N \in R^{N\times 1} \Rightarrow \frac{\partial L}{\partial b_2} = 1_N^T \frac{\partial L}{\partial s}$$ ※

(d)

Scalar : $$\frac{\partial L}{\partial H} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial H} = \frac{\partial L}{\partial s} W_2$$

Matrix : $$H \in R^{N\times H} \Rightarrow \overset{N\times H}{\frac{\partial L}{\partial H}} = \overset{N\times C}{\frac{\partial L}{\partial s}}\overset{C\times H}{\frac{\partial s}{\partial H}} \text{ or } \overset{N\times H}{\frac{\partial L}{\partial H}} = \overset{X}{\frac{\partial s}{\partial H}}\overset{N\times C}{\frac{\partial L}{\partial s}}$$

$$\because W_2 \in R^{H\times C} \Rightarrow \frac{\partial L}{\partial H} = \frac{\partial L}{\partial s} W_2^T$$ ※

(e)

Scalar : $$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial H}\frac{\partial H}{\partial z}$$

$$H = ReLu(z) \Rightarrow H = \begin{cases} z & , z > 0 \\ 0 & , \text{otherwise} \end{cases} \Rightarrow \frac{\partial H}{\partial z} = \begin{cases} 1 & , z > 0 \\ 0 & , \text{otherwise} \end{cases}$$

$$\Rightarrow \frac{\partial L}{\partial z} = \begin{cases} \frac{\partial L}{\partial H} & , z > 0 \\ 0 & , \text{otherwise} \end{cases}$$

Matrix : $$z \in R^{N\times H} \text{ and } H_{ij} = \begin{cases} z_{ij} & , z_{ij} > 0 \\ 0 & , \text{otherwise} \end{cases} \Rightarrow \frac{\partial H_{ij}}{\partial z_{ij}} = \begin{cases} 1 & , z_{ij} > 0 \\ 0 & , \text{otherwise} \end{cases}$$

$$\Rightarrow \frac{\partial L}{\partial z} = R^{N\times H} \text{ and } \frac{\partial L}{\partial z_{ij}} = \begin{cases} \frac{\partial L}{\partial H_{ij}} & , z_{ij} > 0 \\ 0 & , \text{otherwise} \end{cases}$$

(f)

Scalar : $$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial W_1} = \frac{\partial L}{\partial z} X$$

Matrix : $$W_1 \in R^{D\times H} \Rightarrow \overset{D\times H}{\frac{\partial L}{\partial W_1}} = \overset{N\times H}{\frac{\partial L}{\partial z}} \overset{X}{\frac{\partial z}{\partial W_1}} \text{ or } \overset{D\times H}{\frac{\partial L}{\partial W_1}} = \overset{D\times N}{\frac{\partial z}{\partial W_1}} \overset{N\times H}{\frac{\partial L}{\partial z}}$$

$$\because X \in R^{N\times D} \Rightarrow \frac{\partial L}{\partial W_1} = X^T \frac{\partial L}{\partial z}$$

(g)

Scalar : $$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial b_1} = \frac{\partial L}{\partial z} 1$$

Matrix : $$b_1 \in R^{1\times H} \Rightarrow \overset{1\times H}{\frac{\partial L}{\partial b_1}} = \overset{N\times H}{\frac{\partial L}{\partial z}} \overset{X}{\frac{\partial z}{\partial b_1}} \text{ or } \overset{1\times H}{\frac{\partial L}{\partial b_1}} = \overset{1\times N}{\frac{\partial z}{\partial b_1}} \overset{N\times H}{\frac{\partial L}{\partial z}}$$

$$\because 1_N \in R^{N\times 1} \Rightarrow \frac{\partial L}{\partial b_1} = 1_N^T \frac{\partial L}{\partial z}$$