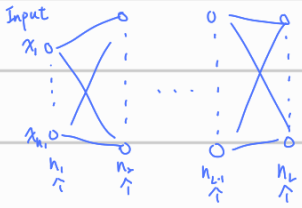

$a^{[1]} = \sigma(z^{[1]})$

$z^{[1]} = W^{[1]} x + b^{[1]}$

$\mathbb{R}^{n_1}$ $M_{n_1 \times 1}$ $\mathbb{R}^{n_1}$ $M_{n_1 \times 1}$

$\Rightarrow z^{[l]} = W^{[l-1]} a^{[l-1]} + b^{[l]}$

$a^{[l]} = \sigma(z^{[l]})$ ——— ⊛

$H(x) = a^{[L]}$, $H: \mathbb{R}^{n_1} \to \mathbb{R}^{n_L}$

Define $\delta_j^{[l]} = \frac{\partial H}{\partial z_j^{[l]}}$, $l = 2, \dots, L$, $\delta_j^{[L]} = \frac{\partial H}{\partial z_j^{[L]}} = \sigma'(z_j^{[L]})$

$H(x) = \sigma(z^{[L]})$, $z^{[L]} \in \mathbb{R}$

$\delta_j^{[L-1]} = \frac{\partial H}{\partial \delta^{[L]}} \cdot \frac{\partial \delta^{[L]}}{\partial z_j^{[L-1]}}$

* HW 9/1

$n_L = 1$

$H: \mathbb{R}^{n_1} \to \mathbb{R}^1$ Calculate $\nabla H = \begin{bmatrix} \frac{\partial H}{\partial x_1} \\ \vdots \\ \frac{\partial H}{\partial x_{n_1}} \end{bmatrix}$

<ans> $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$ ⊛, $a^{[l]} = \sigma(z^{[l]})$, $H(x) = a^{[l]}$, calculate $\frac{\partial H(x)}{\partial x}$

Let (MSE) $C = \frac{1}{2}\|y - H(x)\|_2^2$, $\delta_j^{[l]} = \frac{\partial C}{\partial z_j^{[l]}}$ $1 \le j \le n_l$, $2 \le l \le L$

$\frac{\partial a_j^{[L]}}{\partial z_j^{[L]}} = \sigma'(z_j^{[L]})$, $\frac{\partial C}{\partial a_j^{[L]}} = \frac{\partial}{\partial a_j^{[L]}} \frac{1}{2} \sum_{k=1}^{n_L} (y_k - a_k^{[L]})^2 = -(y_j - a_j^{[L]})$

using chain rule, $\delta_j^{[L]} = \frac{\partial C}{\partial z_j^{[L]}} = \frac{\partial C}{\partial a_j^{[L]}} \frac{\partial a_j^{[L]}}{\partial z_j^{[L]}} = (a_j^{[L]} - y_j)\sigma'(z_j^{[L]}) = \sigma'(z^{[L]}) \circ (a^L - y)$

$\delta_j^{[l]} = \frac{\partial C}{\partial z_j^{[l]}} = \sum_{k=1}^{n_{l+1}} \frac{\partial C}{\partial z_k^{[l+1]}} \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} = \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} \frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} = \sum_{k=1}^{n_{l+1}} \delta_k^{[l+1]} w_{kj}^{[l+1]} \sigma'(z_j^{[l]}) = \sigma'(z_j^{[l]}) \left[(W^{[l+1]})^T \delta^{[l+1]} \right]_j = \left[(W^{[l+1]})^T \delta^{[l+1]} \circ \sigma'(z_j^{[l]}) \right]$

By ⊛, $z_k^{[l+1]} = \sum_{s=1}^{n_l} w_{ks}^{[l+1]} \sigma(z_s^{[l]}) + b_k^{[l+1]}$. Hence, $\frac{\partial z_k^{[l+1]}}{\partial z_j^{[l]}} = w_{kj}^{[l+1]} \sigma'(z_j^{[l]})$

By ⊛ and ⊛, $z_j^{[l]} = (W^{[l]} \sigma(z^{[l-1]}))_j + b_j^{[l]}$ $\Rightarrow \frac{\partial z_j^{[l]}}{\partial b_j^{[l]}} = 1$ $\xrightarrow{\text{chain rule}}$ $\frac{\partial C}{\partial b_j^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} \frac{\partial z_j^{[l]}}{\partial b_j^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} = \delta_j^{[l]}$

$z_j^{[l]} = \sum_{k=1}^{n_{l-1}} w_{jk}^{[l]} a_k^{[l-1]} + b_j^{[l]}$ and $\frac{\partial z_j^{[l]}}{\partial w_{jk}^{[l]}} = a_k^{[l-1]}$, indep. of j, and $\frac{\partial z_s^{[l]}}{\partial w_{jk}^{[l]}} = 0$ for $s \neq j$

$\Rightarrow \frac{\partial C}{\partial w_{jk}^{[l]}} = \sum_{s=1}^{n_l} \frac{\partial C}{\partial z_s^{[l]}} \frac{\partial z_s^{[l]}}{\partial w_{jk}^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} \frac{\partial z_j^{[l]}}{\partial w_{jk}^{[l]}} = \frac{\partial C}{\partial z_j^{[l]}} a_k^{[l-1]} = \delta_j^{[l]} a_k^{[l-1]}$

$\frac{\partial z^{[l]}}{\partial a^{[l-1]}} = W^{[l]}$, $a^{[l]} = \sigma(z^{[l]})$ $\Rightarrow \frac{\partial a^{[l]}}{\partial z^{[l]}} = \mathrm{diag}(\sigma'(z^{[l]})) = D^{[l]}$

layer 1 $\frac{\partial a^{[1]}}{\partial x} = \frac{\partial a^{[1]}}{\partial z^{[1]}} \frac{\partial z^{[1]}}{\partial x} = D^{[1]} W^{[1]}$

layer 2 $\frac{\partial a^{[2]}}{\partial x} = \frac{\partial a^{[2]}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial a^{[1]}} \frac{\partial a^{[1]}}{\partial x} = D^{[2]} W^{[2]} (D^{[1]} W^{[1]})$

layer l. $\frac{\partial a^{[l]}}{\partial x} = \frac{\partial a^{[l]}}{\partial z^{[l]}} \frac{\partial z^{[l]}}{\partial a^{[l-1]}} \frac{\partial a^{[l-1]}}{\partial x} = D^{[l]} W^{[l]} \left(\frac{\partial a^{[l-1]}}{\partial x} \right)$

$\frac{\partial H}{\partial x} = \frac{\partial a^{[L]}}{\partial x} = \frac{\partial a^{[L]}}{\partial z^{[L]}} \frac{\partial z^{[L]}}{\partial a^{[L-1]}} \cdots \frac{\partial a^{[2]}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial a^{[1]}} \frac{\partial a^{[1]}}{\partial z^{[1]}} \frac{\partial z^{[1]}}{\partial x}$

$\Rightarrow \frac{\partial H}{\partial x} = (D^{[L]} W^{[L]})(D^{[L-1]} W^{[L-1]}) \cdots (D^{[2]} W^{[2]})(D^{[1]} W^{[1]})$

The derivation is based on the paper Higham 2019 _ Deeplearning Intro Applied Methematic.