

$$6.(a) \quad \frac{\partial y_L}{\partial b_L} = \frac{\partial}{\partial b_L} (A_L y_{L-1} + b_L) = 1$$

$$\frac{\partial y_L}{\partial y_{L-1}} = \frac{\partial}{\partial y_{L-1}} (A_L y_{L-1} + b_L) = A_L$$

For $l=1, \dots, L-1$,

$$\frac{\partial y_l}{\partial b_l} = \begin{pmatrix} \frac{\partial y_{l1}}{\partial b_{l1}} & \frac{\partial y_{l1}}{\partial b_{l2}} & \dots & \frac{\partial y_{l1}}{\partial b_{ln_l}} \\ \frac{\partial y_{l2}}{\partial b_{l1}} & \frac{\partial y_{l2}}{\partial b_{l2}} & \dots & \frac{\partial y_{l2}}{\partial b_{ln_l}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{ln_l}}{\partial b_{l1}} & \frac{\partial y_{ln_l}}{\partial b_{l2}} & \dots & \frac{\partial y_{ln_l}}{\partial b_{ln_l}} \end{pmatrix}$$

Define $\tilde{y}_l = A_l y_{l-1} + b_l$ (i.e. y_l pre-activation)

$$\begin{aligned} \frac{\partial \tilde{y}_{li}}{\partial b_{lj}} &= \frac{\partial}{\partial b_{lj}} ([A_l y_{l-1} + b_l]_i) \\ &= \frac{\partial}{\partial b_{lj}} ([A_l y_{l-1}]_i + b_{li}) \\ &= \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \end{aligned}$$

Hence, only diagonal entries of $\frac{\partial y_l}{\partial b_l}$ remain (i.e. $i=j$) are non-zero

$$\begin{aligned} \frac{\partial y_{li}}{\partial b_{lj}} &= \frac{\partial}{\partial b_{lj}} ([\sigma(\tilde{y}_l)]_i) \\ &= \frac{\partial \tilde{y}_{li}}{\partial b_{lj}} [\sigma'(\tilde{y}_l)]_i = \frac{\partial \tilde{y}_{li}}{\partial b_{lj}} \sigma'(\tilde{y}_{li}) \end{aligned}$$

by chain rule and elementwise operation

$$\text{with } \frac{\partial y_{li}}{\partial b_{li}} = \sigma'(\tilde{y}_{li}) = \sigma'([A_l y_{l-1} + b_l]_i)$$

$$= \begin{cases} 0 & \text{if } i \neq j \\ \sigma'(\tilde{y}_{li}) & \text{if } i = j \end{cases} \text{ by above}$$

$$\text{That is, } \frac{\partial y_l}{\partial b_l} = \text{diag}(\sigma'(A_l y_{l-1} + b_l)) \quad \square$$

For $l=2, \dots, L-1$,

For $\frac{\partial y_l}{\partial y_{l-1}}$, consider entries one row:

$$\begin{aligned} \frac{\partial \tilde{y}_{li}}{\partial y_{l-1j}} &= \frac{\partial}{\partial y_{l-1j}} ([A_l y_{l-1}]_i + b_{li}) \\ &= \frac{\partial}{\partial y_{l-1j}} \left(\sum_{k=1}^{n_{l-1}} (A_l)_{ik} (y_{l-1})_k + b_{li} \right) \\ &= (A_l)_{ij} \end{aligned}$$

$$i \left(\overbrace{\quad}^{A_l} \right) \left(\underbrace{\quad}_{y_{l-1}} \right)$$

$$\begin{aligned} \text{So } \frac{\partial y_{li}}{\partial y_{l-1j}} &= \frac{\partial}{\partial y_{l-1j}} (\sigma(\tilde{y}_{li})) = \frac{\partial \tilde{y}_{li}}{\partial y_{l-1j}} \sigma'(\tilde{y}_{li}) = (A_l)_{ij} \sigma'(\tilde{y}_{li}) \\ &= \sigma'(\tilde{y}_{li}) (A_l)_{ij} \\ &= [\sigma'(A_l y_{l-1} + b_l)]_i (A_l)_{ij} \end{aligned}$$

elements equivalent

$$\begin{aligned} [\text{diag}(\sigma'(A_l y_{l-1} + b_l)) A_l]_{ij} &= \sum_{k=1}^{n_l} [\text{diag}(\sigma'(A_l y_{l-1} + b_l))]_{ik} (A_l)_{kj} = [\sigma'(A_l y_{l-1} + b_l)]_i (A_l)_{ij} \\ &\quad \underbrace{\neq 0 \text{ iff } k=i} \\ \Rightarrow \frac{\partial y_l}{\partial y_{l-1}} &= \text{diag}(\sigma'(A_l y_{l-1} + b_l)) A_l \end{aligned}$$

$$i \left(\overbrace{\quad}^{A_l} \right) \left(\underbrace{\quad}_{y_{l-1}} \right)$$

as given \square

6.(b) Consider again, $y_l = \sigma(A_l y_{l-1} + \underline{b}_l)$

$$\begin{aligned}
 &= \sum_{i=1}^{n_l} \left(\sigma([A_l y_{l-1} + \underline{b}_l]_i) \right) \cdot \underline{e}_i \\
 &= \sum_{i=1}^{n_l} \left(\sigma \left(\sum_{j=1}^{n_{l-1}} (A_l)_{ij} (y_{l-1})_j + (\underline{b}_l)_i \right) \right) \cdot \underline{e}_i \\
 \Rightarrow \frac{\partial y_l}{\partial (A_l)_{ij}} &= \sigma'([A_l y_{l-1} + \underline{b}_l]_i) \cdot (y_{l-1})_j \cdot \underline{e}_i
 \end{aligned}$$

$$\begin{aligned}
 \left(\frac{\partial y_l}{\partial A_l} \right)_{ij} &= \frac{\partial y_l}{\partial (A_l)_{ij}} = \frac{\partial y_l}{\partial y_l} \frac{\partial y_l}{\partial (A_l)_{ij}} \quad \text{by Chain Rule} \\
 &= \frac{\partial y_l}{\partial y_l} \sigma'([A_l y_{l-1} + \underline{b}_l]_i) (y_{l-1})_j \underline{e}_i \\
 &= \sigma'([A_l y_{l-1} + \underline{b}_l]_i) \left(\frac{\partial y_l}{\partial y_l} \right)_i^T (y_{l-1})_j
 \end{aligned}$$

$$\Rightarrow \frac{\partial y_l}{\partial A_l} = \underbrace{\text{diag}(\sigma'(A_l y_{l-1} + \underline{b}_l))}_{\in \mathbb{R}^{n_l \times n_l}} \underbrace{\left(\frac{\partial y_l}{\partial y_l} \right)^T}_{\in \mathbb{R}^{n_l \times 1}} \underbrace{(y_{l-1})^T}_{\in \mathbb{R}^{1 \times n_{l-1}}}$$

$(\sim)(j) \rightarrow (\)_j \quad , \rightarrow (\)$

$$\begin{aligned}
 \left[\left(\frac{\partial y_l}{\partial A_l} \right)_{ij} \right] &= \sum_{k=1}^{n_l} \left(\text{diag}(\sigma'(A_l y_{l-1} + \underline{b}_l))_{ik} \left(\frac{\partial y_l}{\partial y_l} \right)_k^T \right) \cdot (y_{l-1})_j \\
 &= [\sigma'(A_l y_{l-1} + \underline{b}_l)]_i \left(\frac{\partial y_l}{\partial y_l} \right)_i^T (y_{l-1})_j \quad \text{so matches above } \checkmark
 \end{aligned}$$