4. $$\frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j, \quad \text{where } x = (x_1, \cdots, x_d)^T$$

- By chain rule,

$$\frac{\partial J}{\partial W_{ij}} = \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \cdot \frac{\partial y_r}{\partial W_{ij}}$$

- Since only the item with $r = i$ is non-zero,

$$\frac{\partial J}{\partial W_{ij}} = \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \cdot \frac{\partial y_r}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} \cdot \frac{\partial y_i}{\partial W_{ij}}$$

- $\frac{\partial y_i}{\partial W_{ij}} = x_j$ (as $y = Wx + b$)

$$\boxed{\frac{\partial J}{\partial W_{ij}} = \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \cdot \frac{\partial y_r}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} \cdot \frac{\partial y_i}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} \cdot x_j}$$

5. $\frac{\partial J}{\partial W} = \frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial y_i} x_j = \frac{\partial J}{\partial y} \otimes x$

---

6. $\boxed{\frac{\partial J}{\partial x}} = \frac{\partial J}{\partial x_i} = \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \cdot \frac{\partial y_r}{\partial x_i}$

$\left(\text{as } y = Wx + b, \quad \frac{\partial y_r}{\partial x_i} = W_{r,i} \right)$

$= \sum_{r=1}^{m} \frac{\partial J}{\partial y_r} \cdot W_{r,i} = (W_i)^T \cdot \frac{\partial J}{\partial y_r} = \boxed{W^T \left(\frac{\partial J}{\partial y} \right)}$

---

7. From $y = Wx + b$, $\frac{\partial y}{\partial b} = I$

$\therefore \boxed{\frac{\partial J}{\partial b}} = \frac{\partial J}{\partial y} \cdot \frac{\partial y}{\partial b} = \frac{\partial J}{\partial y} \cdot I = \boxed{\frac{\partial J}{\partial y}}$

8. Show $\frac{\partial J}{\partial A} = \frac{\partial J}{\partial S} \odot \sigma'(A)$, where $(A \odot B)_i = A_i B_i$

Based on definition of $\sigma'(A)$, $\frac{\partial S_i}{\partial A_i} = \sigma'(A_i)$

So, $\frac{\partial J}{\partial A_i} = \frac{\partial J}{\partial S_i} \cdot \frac{\partial S_i}{\partial A_i} = \frac{\partial J}{\partial S_i} \cdot \sigma'(A_i)$

$$\boxed{\therefore \frac{\partial J}{\partial A} = \frac{\partial J}{\partial S} \odot \sigma'(A)}$$