Lec 17

1. Recall: back propagation Algo

↳ Input : $(\underline{x}, y)$, $\Omega = \{ w^{(1)}, w^{(2)} \cdots w^{(L)} \}$

Output : $\dfrac{\partial e(\Omega)}{\partial w_{i,j}^{(\ell)}}$, $\ell = 1, 2, \cdots L$

↳ Steps:

① Run forward propagation to compute

$$\{ s^{(\ell)}, x^{(\ell)} \} , \ell = 1, 2, \cdots L$$

$$e(\Omega) = g(\underline{x}^{(L)}, y)$$

② $\underline{\delta}^{(L)} = \underline{\delta}^{(\ell)} = [\hat{w}^{(\ell+1)} \underline{\delta}^{(\ell+1)}] \otimes \theta'(\underline{s}^{(\ell)})$

③ For $\ell = L-1$ to 1 do:

$$\delta^{(\ell)} = [\hat{w}^{(\ell+1)} \underline{\delta}^{(\ell+1)}] \otimes \theta'(\underline{s}^{(\ell)})$$

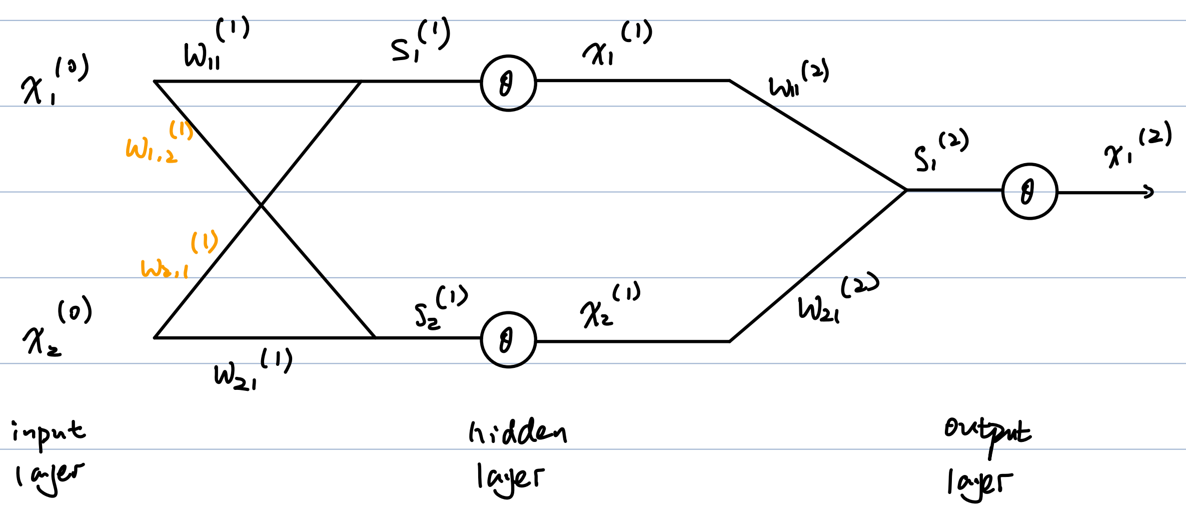
$$\frac{\partial e(\Omega)}{\partial w^{(\ell)}} = \underline{x}^{(\ell-1)} (\delta^{(\ell)})^T$$

End

↳ Complexity : $O(Q)$

// one forward propagation to compute all $\dfrac{\partial e}{\partial w^{(\ell)}}$

e.g.



$$e(\Omega) = (x_1^{(2)} - y)^2 \qquad L = 2$$

<u>Layer 2:</u> $$\delta_1^{(2)} = \frac{\partial e}{\partial x_1^{(2)}} \cdot \frac{\partial x_1^{(2)}}{\partial s_1^{(2)}} = 2(x_1^{(2)} - y)\, \theta'(s_1^{(2)})$$

$$\frac{\partial e}{\partial w_{11}^{(2)}} = x_1^{(1)}\, \delta_1^{(2)}, \qquad \frac{\partial e}{\partial w_{21}^{(2)}} = x_2^{(1)}\, \delta_1^{(2)}$$

<u>Layer 1:</u> $$\delta_1^{(1)} = \theta'(s_1^{(1)})\, \delta_1^{(2)}\, w_{11}^{(2)}$$

$$\delta_2^{(1)} = \theta'(s_2^{(1)})\, \delta_1^{(2)}\, w_{21}^{(2)}$$

$$\frac{\partial e}{\partial w_{ij}^{(1)}} = x_i^{(0)} \cdot \delta_j^{(1)}, \qquad i, j \in \{1, 2\}$$

Done.

↳ <u>Note:</u>

① Can do more:

$$\frac{\partial e}{\partial x_1^{(0)}} \triangleq \delta_1^{(0)} = \frac{\partial e}{\partial s_1^{(0)}}$$

$$= \frac{\partial e}{\partial s_1^{(1)}} \cdot \frac{\partial s_1^{(1)}}{\partial x_1^{(0)}} + \frac{\partial e}{\partial s_2^{(0)}} \cdot \frac{\partial s_2^{(1)}}{\partial x_1^{(0)}}$$

$$= \delta_1^{(1)}\, w_{11}^{(1)} + \delta_2^{(1)}\, w_{12}^{(1)}$$

$$\frac{\partial e}{\partial x_2^{(0)}} = \cdots$$

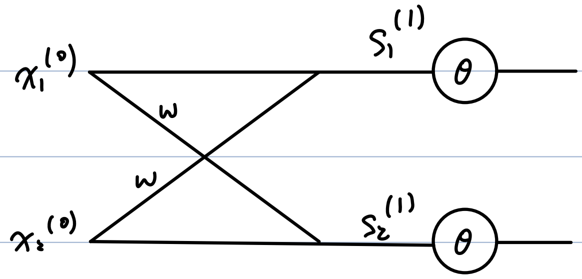
This is the sensitivity of $e(\Omega)$ to the input

→ Too sensetive to change in input. chaos

→ 〃 〃 〃 to noise in training. overfit

② what if we force

$w_{12}^{(1)} = w_{21}^{(1)} \triangleq w$?



$$\frac{\partial e}{\partial w} = \frac{\partial e}{\partial s_1^{(1)}} \cdot \frac{\partial s_1^{(1)}}{\partial w} + \frac{\partial e}{\partial s_2^{(1)}} \cdot \frac{\partial s_2^{(1)}}{\partial w}$$

$$= \delta_1^{(1)} x_2^{(0)} + \delta_2^{(1)} x_1^{(0)}$$

General: CNN