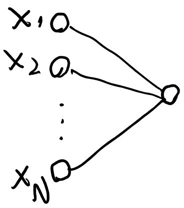
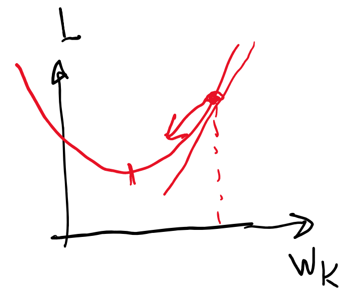
Backprop (lecture)

Tuesday, April 28, 2020 13:50



$$y = \sum_j^N w_j x_j = \underline{w} \cdot \underline{x}$$



$$L = \sum_\mu^P \left(y^\mu - \sum_j^N w_j x_j^\mu \right)^2$$

$$\Delta w_K = - \alpha \frac{\partial L}{\partial w_K}$$

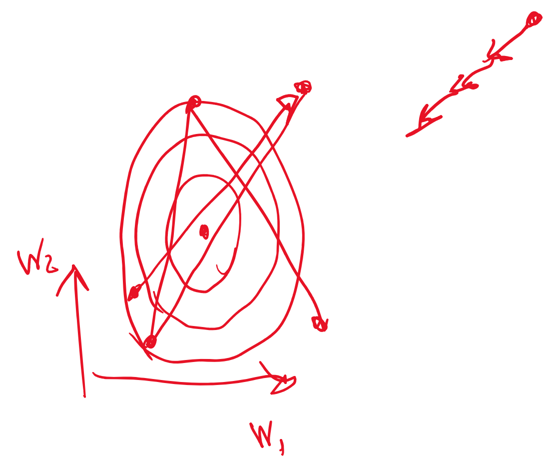
$$\Delta w_k = -2\alpha \sum_\mu^P \underbrace{\left(y^\mu - \sum_j^N w_j x_j^\mu \right)}_{\rightsquigarrow\, \delta^\mu} \cdot \left(- x_k^\mu\right) = \underbrace{2\alpha}_{\alpha} \sum_\mu^P \delta^\mu x_k^\mu$$

SGD: $\mu \rightarrow \underline{\Delta w_K = \alpha \, \delta^\mu x_k^\mu}$ LTP/LTD

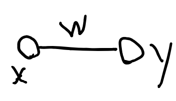
$$y = f\left(\sum_j^N w_j x_j\right)$$

$$\Delta w_K = \alpha \sum_\mu^P \delta^\mu \cdot f'\left(\sum_j^N w_j x_j^\mu\right) \cdot x_K^\mu$$

$$\delta^\mu \rightarrow \delta^\mu f'\left(\sum_j^N w_j x_j^\mu\right)$$



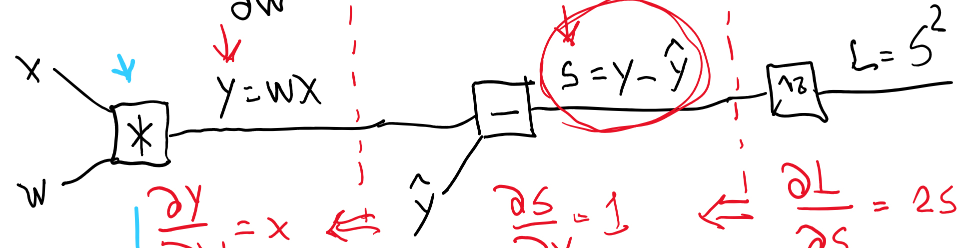
$$y = wx$$



$$L = (y - \hat{y})^2 = (wx - \hat{y})^2$$

$$\Delta w = - \alpha \frac{\partial L}{\partial w}$$

$$\Delta W = -\alpha \frac{\partial L}{\partial W}$$



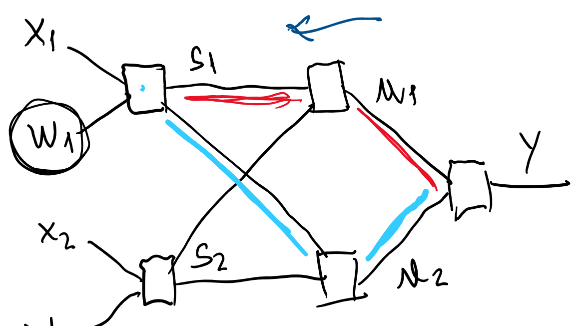
$$\frac{\partial y}{\partial w} = x \Leftarrow \qquad \frac{\partial s}{\partial y} = 1 \Leftarrow \qquad \frac{\partial L}{\partial s} = 2s$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial y}\frac{\partial y}{\partial w}$$

FW: $x = 1$, $w = 2$, $\hat{y} = 0$ $\Rightarrow$ $y = 2$ $\Rightarrow$ $s = 2$ $\Rightarrow$ $L = 4$

BW: $\Delta W = 1 \cdot 4 \cdot 1 \cdot \alpha \leftarrow \frac{\partial s}{\partial y} = 1 \cdot 4 \leftarrow \frac{\partial L}{\partial s} = 4$

$w \rightarrow w + \Delta w$



$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial n_1}\frac{\partial n_1}{\partial s_1}\frac{\partial s_1}{\partial w_1} + \frac{\partial L}{\partial y}\frac{\partial y}{\partial n_2}\frac{\partial n_2}{\partial s_1}\frac{\partial s_1}{\partial w_1}$$

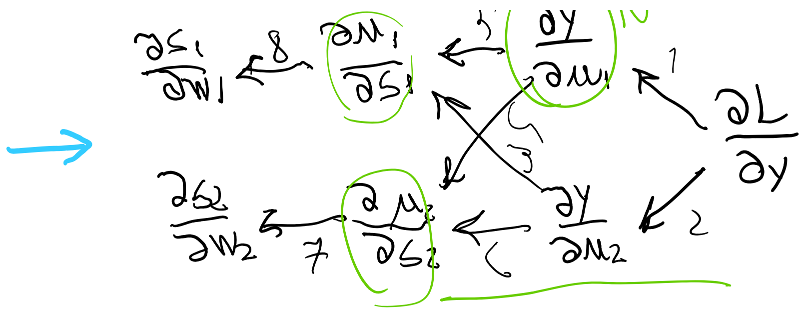
16

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial n_1}\frac{\partial n_1}{\partial s_2}\frac{\partial s_2}{\partial w_2} + \frac{\partial L}{\partial y}\frac{\partial y}{\partial n_2}\frac{\partial n_2}{\partial s_2}\frac{\partial s_2}{\partial w_2}$$

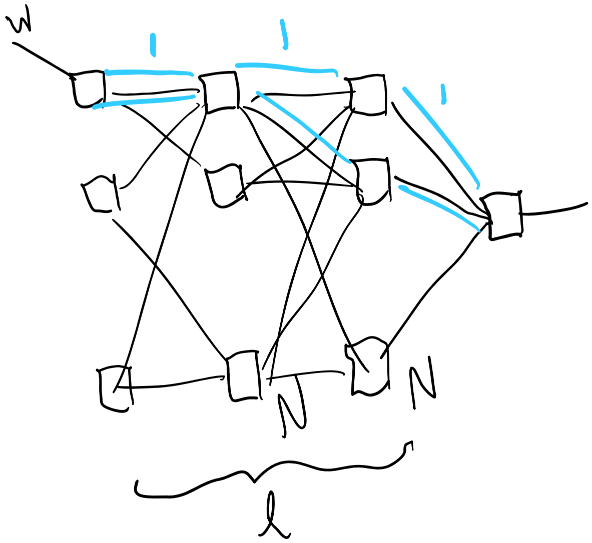
4 4 4 4

$$\frac{\partial s_1}{\partial w_1} \xleftarrow{8} \frac{\partial n_1}{\partial s_1} \xleftarrow{5} \frac{\partial y}{\partial n_1} \xleftarrow{1} \partial L$$

N N



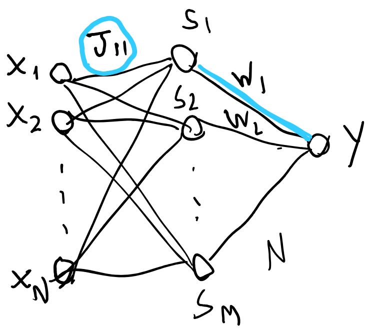
8



$$\sim N^{\ell}$$

$$N \cdot N \cdot (\ell - 1) = N^2 \ell \ll N^{\ell}$$

$$\gg 1$$



$N \times M$

$$y = f\Big(\overbrace{\sum_{j=1}^{M} w_j s_j}^{h}\Big)$$

$$f\Big(\underbrace{\sum_{k=1}^{N} J_{jk} x_k}_{h_j}\Big)$$

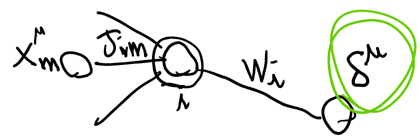
$$L(\{w, J\}) = \sum_{\mu=1}^{p} \Big(y^{\mu} - f\Big(\sum_{j=1}^{M} w_j s_j\Big)\Big)^2$$

$$\Delta w_i = -\alpha \frac{\partial L}{\partial w_i} = +2\alpha \sum_{\mu=1}^{p} \underbrace{\big(y^{\mu} - f(h^{\mu})\big)\big(+f'(h^{\mu})\big)}_{\delta^{\mu}} s_i^{\mu} = 2\alpha \sum_{\mu=1} \delta^{\mu} s_i^{\mu}$$

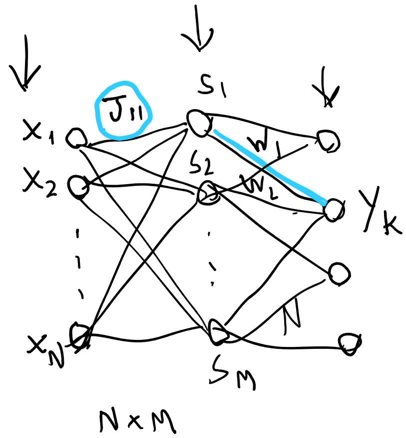
$$\Delta J_{im} = -\alpha \frac{\partial L}{\partial J_{im}} = -\alpha \sum_{\mu=1}^{p} \frac{\partial L}{\partial s_i} \frac{\partial s_i}{\partial J_{im}} \rightarrow f'(h_i^{\mu})\, x_m^{\mu}$$

$$- \delta^{\mu} w_i$$

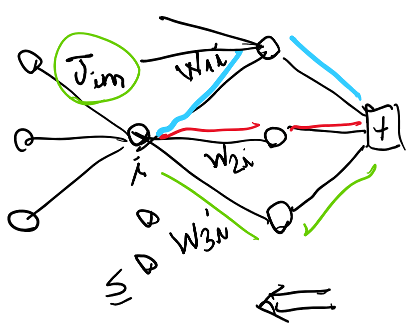
$$\Delta J_{im} = 2\alpha \sum_{\mu=1}^{p} \underbrace{\delta^{\mu} w_i f'(h_i^{\mu})}_{\overset{DEF}{=} \delta_i^{\mu}} \cdot x_m^{\mu}$$



$$\delta_i^{\mu} = f'(h_i^{\mu}) w_i \delta^{\mu}$$



$$L = \sum_{k=1}^{N_y} \sum_{\mu=1}^{p} \left[y_k^{\mu} - f\left(\sum_{j=1}^{M} w_{kj} S_j\right)\right]$$



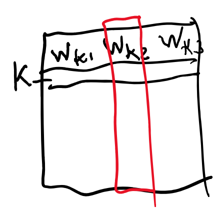
$$\delta_{y_k}^{\mu} = (y_k^{\mu} - f(\dots)) f'(h_{y_k}^{\mu})$$

$$\Delta J_{im} = \alpha \sum_{\mu=1}^{p} \delta_i^{\mu} x_m^{\mu}$$

$$f'(h_i^{\mu}) \sum_{k=1}^{N_y} w_{ki} \delta_{y_k}^{\mu}$$

$$W \rightarrow W^T$$

$$\underline{y} = W \underline{S}$$



$$y_k = f\left(\sum_{j=1}^{M} w_{kj} S_j\right)$$

$$\sum_{k=1}^{N_y} w_{ki} \delta_{y_k}^{\mu} = W^T \cdot \underline{\delta_y}$$