$$\frac{\partial L}{\partial w_{ij}^3} = \underbrace{\frac{\partial L}{\partial z_i^4}}_{\delta_i^4} \cdot \underbrace{\frac{\partial z_i^4}{\partial w_{ij}^3}}_{a_j^3}$$

$$\frac{\partial L}{\partial w_{ij}^2} = \underbrace{\frac{\partial L}{\partial z_i^3}}_{\delta_i^3} \cdot \underbrace{\frac{\partial z_i^3}{\partial w_{ij}^2}}_{a_j^2}$$

$$\frac{\partial L}{\partial w_{ij}^1} = \underbrace{\frac{\partial L}{\partial z_i^2}}_{\delta_i^2} \cdot \underbrace{\frac{\partial z_i^2}{\partial w_{ij}^1}}_{x_j}$$

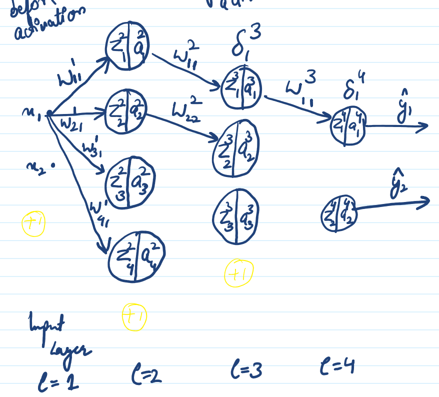
Calculate $\delta_i^4$ first and using that solve for $\delta_i^3$, $\delta_i^2$, $\delta_i^1$ in a recursive manner.

"Delta Learning Rule"

$w_{ij}^l, b_i^l$ $\quad z_i^l / a_i^l$ $\quad \delta_i^l \rightarrow$ error

before activation

after activation




$$L(\theta) = \frac{1}{2}\sum_{i=1}^{2}(\hat{y}_i - y_i)^2$$

$\theta \rightarrow w, b$ $\qquad y_i \rightarrow$ GT value

$$z_i^2 = \sum_{j=1}^{2} w_{ij}^1 x_j + b_i^1, \quad i = 1,2,3,4$$

$$a_i^2 = f(z_i^2)$$

$$z_i^3 = \sum_{j=1}^{4} w_{ij}^2 a_j^2 + b_i^2, \quad i = 1,2,3$$

$$a_i^3 = f(z_i^3)$$

$$z_i^4 = \sum w_{ij}^3 a_j^3 + b_i^3, \quad i = 1,2$$

$$a_i^4 = f(z_i^4)$$

$$\frac{\partial L}{\partial w_{ij}^3} = \underbrace{\frac{\partial L}{\partial z_i^4}}_{\delta_i^4} \cdot \underbrace{\frac{\partial z_i^4}{\partial w_{ij}^3}}_{a_j^3}$$

$$\frac{\partial L}{\partial w_{ij}^2} = \underbrace{\frac{\partial L}{\partial z_i^3}}_{\delta_i^3} \cdot \underbrace{\frac{\partial z_i^3}{\partial w_{ij}^2}}_{a_j^2}$$

$$\frac{\partial L}{\partial w_{ij}^1} = \underbrace{\frac{\partial L}{\partial z_i^2}}_{\delta_i^2} \cdot \underbrace{\frac{\partial z_i^2}{\partial w_{ij}^1}}_{x_j}$$

$$\because \frac{\partial z_i^4}{\partial w_{ij}^3} = a_j^3$$

$$\frac{\partial z_j^4}{\partial a_i^3} = w_{ji}^3$$

$$\delta_i^4 = \frac{\partial L}{\partial z_i^4} = (\hat{y}_i - y_i)\frac{\partial \hat{y}_i}{\partial z_i^4} \rightarrow f'(z_i^4)$$

$$\delta_i^3 = \frac{\partial L}{\partial z_i^3} = \frac{\partial L}{\partial a_i^3} \cdot \frac{\partial a_i^3}{\partial z_i^3} \quad f'(z_i^3)$$

$$\frac{\partial L}{\partial a_i^3} = \sum_{j=1}^{2} \underbrace{\frac{\partial L}{\partial z_j^4}}_{\delta_j^4} \cdot \underbrace{\frac{\partial z_j^4}{\partial a_i^3}}_{w_{ji}^3}$$

$$\Rightarrow \delta_i^3 = \left(\sum_{j=1}^{2} \delta_j^4 \cdot w_{ji}^3\right) \cdot f'(z_i^3)$$

wrt $z_i^4$

$$\frac{\partial L}{\partial z_i^4} = \frac{1}{2}\frac{\partial}{\partial z_i^4}\left(\sum_{i=1}^{2}(\hat{y}_i - y_i)^2\right)$$

$$= \frac{\cancel{1}\times\cancel{2}}{\cancel{2}}\frac{\partial}{\partial z_i^4}\left(\sum_{i=1}^{2}(\hat{y}_i - y_i)\right)$$

$$= \frac{\partial}{\partial z_i^4}\left((\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\right)$$

$$= (\hat{y}_1 - y_1)\left(\frac{\partial}{\partial z_i^4}(\hat{y}_1 - y_1)\right) + (\hat{y}_2 - y_2)\left(\frac{\partial}{\partial z_i^4}(\hat{y}_2 - y_2)\right)$$

$$= (\hat{y}_1 - y_1)\left(\frac{\partial \hat{y}_1}{\partial z_i^4}\right) + (\hat{y}_2 - y_2)\left(\frac{\partial \hat{y}_2}{\partial z_i^4}\right)$$

$$\Rightarrow \delta_i^3 = \left(\sum_{j=1}^{2} \delta_j^4 \cdot w_{ji}^3\right) \cdot f'(z_i^3)$$

$$= (\hat{y}_1 - y_1)\left(\frac{\partial \hat{y}_1}{\partial z_i^4}\right) + (\hat{y}_2 - y_2)\left(\frac{\partial \hat{y}_2}{\partial z_i^4}\right)$$

$$= (\hat{y}_i - y_i)\frac{\partial \hat{y}_i}{\partial z_i^4}$$

$$\Rightarrow \delta_i^2 = \left(\sum \delta_j^3 \cdot w_{ji}^2\right) \cdot f'(z_i^2)$$

In general $\delta_i^l = \sum_{j=1}^{s_{l+1}}$

$s_{l+1} \rightarrow$ # of neurons in $(l+1)^{th}$ layer

$$\delta_i^l = \left(\sum_{j=1}^{s_{l+1}} \delta_j^{l+1} \cdot w_{ji}^l\right) \cdot f'(z_i^l)$$