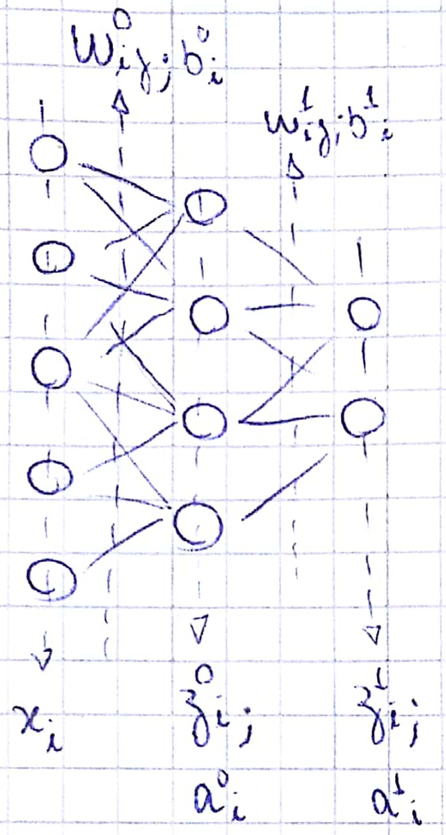

$L = n\_layers = 3$

$l = 0, \dots, L-2 = range(L-1)$

$$z^l_i = \sum_j w^l_{ij} a^{l-1}_j + b^l_i$$

$$a^l_i = \sigma(z^l_i)$$

if $l == 0$:

$$z^0_i = \sum_j w^0_{ij} x_j + b^0_i$$

$$\delta^l_i := \frac{\partial C}{\partial z^l_i} = \sum_j \frac{\partial C}{\partial z^{l+1}_j} \frac{\partial z^{l+1}_j}{\partial z^l_i} = \sum_j \delta^{l+1}_j \frac{\partial (W^{l+1}\sigma(z^l) + b^{l+1})}{\partial z^l_i}$$

$$\Rightarrow \delta^l_i = \sum_j \delta^{l+1}_j \sum_k w^{l+1}_{jk} \frac{\partial \sigma(z^l_k)}{\partial z^l_i} = \sum_j \delta^{l+1}_j w^{l+1}_{ji} \sigma'(z^l_i)$$

$$\Rightarrow \boxed{\delta^l = (W^{l+1})^T \delta^{l+1} \odot \sigma'(z^l)}$$

$$\delta^L_i = \frac{\partial C}{\partial z^L_i} = \frac{\partial C}{\partial a^L_i} \frac{da^L_i}{dz^L_i} = \boxed{\nabla_a C \odot \sigma'(z^L)}$$

$$\frac{\partial C}{\partial w^l_{ij}} = \frac{\partial C}{\partial z^l_i} \cdot \frac{\partial z^l_i}{\partial w^l_{ij}} = \delta^l_i a^{l-1}_j \Rightarrow \boxed{\frac{\partial C}{\partial W^l} = \delta^l \otimes a^{l-1}}$$