

④

Define  $\delta_j^{(l)} = \frac{\partial c}{\partial z_j^{(l)}}$

We will later relate  $\delta_j^{(l)}$  back to the desired gradients.

$c$  is <sup>written as</sup> a function of  $a^{(L)}$  :  $c = c(a^{(L)})$   
 $\uparrow$  last layer

Let's get  $\delta_j^{(L)}$  first:

$$\delta_j^{(L)} = \frac{\partial c}{\partial z_j^{(L)}} = \frac{\partial c}{\partial a_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}}$$

$$\delta_j^{(L)} = \frac{\partial c}{\partial a_j^{(L)}} g_L'(z_j^{(L)}) \quad (1)$$

For layer  $l < L$ :

$$\delta_j^{(l)} = \frac{\partial c}{\partial z_j^{(l)}} = \sum_{k=1}^{n_{l+1}} \underbrace{\frac{\partial c}{\partial z_k^{(l+1)}}}_{= \delta_k^{(l+1)}} \frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}}$$

$$\text{where } \frac{\partial z_k^{(l+1)}}{\partial z_j^{(l)}} = \frac{\partial}{\partial z_j^{(l)}} \left[ \sum_{i=1}^{n_{l+1}} a_i^{(l)} W_{ik}^{(l+1)} + b_k^{(l+1)} \right]$$

$$= \sum_i g_l'(z_i^{(l)}) \frac{\partial z_i^{(l)}}{\partial z_j^{(l)}} W_{ik}^{(l+1)}$$

$$= g_l'(z_j^{(l)}) W_{jk}^{(l+1)}$$

$$\Rightarrow \delta_j^{(l)} = \sum_k \delta_k^{(l+1)} W_{kj}^{(l+1)T} g_l'(z_j^{(l)}) \quad (2)$$

(5)

Now let's get the gradients:

$$\begin{aligned}\frac{\partial C}{\partial W_{ij}^{(l)}} &= \sum_{k=1}^{n_l} \frac{\partial C}{\partial z_k^{(l)}} \frac{\partial z_k^{(l)}}{\partial W_{ij}^{(l)}} \\ &= \sum_k \delta_k^{(l)} \frac{\partial}{\partial W_{ij}^{(l)}} \sum_m \left( a_m^{(l-1)} W_{mk}^{(l)} + b_k^{(l)} \right) \\ \boxed{\frac{\partial C}{\partial W_{ij}^{(l)}}} &= a_i^{(l-1)} \delta_j^{(l)} \quad (3)\end{aligned}$$

$$\begin{aligned}\frac{\partial C}{\partial b_j^{(l)}} &= \sum_k \frac{\partial C}{\partial z_k^{(l)}} \frac{\partial z_k^{(l)}}{\partial b_j^{(l)}} \\ \boxed{\frac{\partial C}{\partial b_j^{(l)}}} &= \delta_j^{(l)} \quad (4)\end{aligned}$$

Summary: training algorithm (one "epoch")

- $\vec{a}^{(0)} = \vec{x}$
- # forward pass:  
for  $l = 1, \dots, L$ :  
 $\vec{a}^{(l)} = g_l(\vec{a}^{(l-1)} W^{(l)} + \vec{b}^{(l)})$
- # backpropagation:  
get  $\vec{\delta}^{(L)}$  from (1)  
for  $l = L-1, \dots, 1$ :  
get  $\vec{\delta}^{(l)}$  from (2)  
get  $\partial C / \partial W_{ij}^{(l)}$  and  $\partial C / \partial b_j^{(l)}$  from (3) & (4)
- update  $\left. \begin{aligned} W_{ij}^{(l)} &= W_{ij}^{(l)} - \eta \frac{\partial C}{\partial W_{ij}^{(l)}} \\ b_j^{(l)} &= b_j^{(l)} - \eta \frac{\partial C}{\partial b_j^{(l)}} \end{aligned} \right\} \begin{array}{l} \text{or use an} \\ \text{alternative} \\ \text{to GD} \end{array}$