Back propagation

Note: $z_j^\ell = \sum_k w_{jk}^\ell a_k^{\ell-1} + b^\ell$, $a_k^{\ell-1} = \sigma(z_k^{\ell-1})$

$$\boxed{\frac{\partial L}{\partial w_{jk}^\ell} = \frac{\partial L}{\partial z_j^\ell} \cdot \frac{\partial z_j^\ell}{\partial w_{jk}^\ell} = d_j^\ell \cdot a_k^{\ell-1}}$$

$$\boxed{\frac{\partial L}{\partial b_j^\ell} = \frac{\partial L}{z_j^\ell} \cdot 1 = d_j^\ell}$$

☆ Label: $d_j^\ell = \frac{\partial L}{\partial z_j^\ell}$, $\vec{d}^\ell = \begin{bmatrix} d_1^\ell \\ d_2^\ell \\ \vdots \\ d_n^\ell \end{bmatrix}$

n= number of nodes in layer ℓ.

Now we need to compute d_j^ℓ.

↳ We'll do this recursively.

Final Layer:

L is a function of the nodes $a_1^F, \ldots, a_n^F$

$$a_j^F = \sigma(z_j^F)$$

$$d_j^F = \frac{\partial L}{\partial z_j^F} = \sum_k \frac{\partial L}{\partial a_k^F} \frac{\partial a_k^F}{\partial z_j^F} = \frac{\partial L}{\partial a_j^F} \frac{\partial a_j^F}{\partial z_j^F}$$

$$= \underbrace{\frac{\partial L}{\partial a_j^F}}_{\text{depends on choice of loss fxn}} \cdot \sigma'(z_j^F)$$

$$\Rightarrow \boxed{\vec{d}^F = \nabla_a L \odot \sigma'(\vec{z}^F)}$$

Intermediate Layers

$$d_j^{\ell} = \frac{\partial L}{\partial z_j^{\ell}}$$

Where do z_j^{ℓ}'s appear?

$\rightarrow$ in a^{ℓ}'s, which appear in $z^{\ell+1}$'s

$$d_j^{\ell} = \frac{\partial L}{\partial z_j^{\ell}} = \sum_k \frac{\partial L}{\partial z_k^{\ell+1}} \frac{\partial z_k^{\ell+1}}{\partial z_j^{\ell}}$$

$$= \sum_k d_k^{\ell+1} \frac{\partial z_k^{\ell+1}}{\partial z_j^{\ell}}$$

$$z_k^{\ell+1} = \sum_j w_{kj}^{\ell+1} a_j^{\ell} + b_k^{\ell+1}$$

$$= \sum_k w_{kj}^{\ell+1} \sigma(z_j^{\ell}) + b_k^{\ell+1}$$

$$\frac{\partial z_k^{\ell+1}}{\partial z_j^{\ell}} = W_{kj}^{\ell+1} \sigma'(z_j^{\ell})$$

$$\Rightarrow d_j^{\ell} = \sum_k w_{kj}^{\ell+1} d_k^{\ell+1} \sigma'(z_j^{\ell})$$

$$\Rightarrow \boxed{\vec{d}^{\ell} = \left((W^{\ell+1})^T \vec{d}^{\ell+1}\right) \odot \sigma'(\vec{Z}^{\ell})}$$