Backpropagation

<u>Note</u>: $z_j^l = \left(\sum_k w_{jk}^l \, a_k^{l-1} \right) + b_j^l$

$$a_k^{l-1} = \sigma\left(z_k^{l-1} \right)$$

To compute ∇L, we must compute:

$$\boxed{\frac{\partial L}{\partial w_{jk}^l} = \frac{\partial L}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial w_{jk}^l} = d_j^l \, a_j^{l-1}}$$

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

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

sometimes called "error"

Final Layer ($\ell = F$)

Consider L as a fxn of the nodes in the final layer $a_1^F, \ldots, a_n^F$.

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

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

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

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

$$\vec{d}^F = \begin{bmatrix} d_1^F \\ d_2^F \\ \vdots \\ d_n^F \end{bmatrix}$$

n nodes in the final layer

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

Intermediate Layers ($\ell = 2, 3, \ldots, F-1$)

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

Question: Where do z_j^ℓ's appear?

→ In a_j^ℓ's, which appear in $z_j^{\ell+1}$'s

sum over nodes in $\ell+1$ layer

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

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

sum over nodes in ℓ^{th} layer

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

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

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

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

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