chapter-2 Backpropogation

Error

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}$$

It relates how a particular neuron (j) in layer l contribute to C (cost fn)

An eqn for the error in o/p layer, δ^L

$$\delta_j^L = \frac{\partial C}{\partial a_j^L}\sigma'(z_j^L) = \frac{\partial C}{\partial a_j^L}\frac{\partial a_j^L}{\partial z_j^L}$$

$$a = \frac{1}{1+e^{-z}} \rightarrow \sigma(z)$$

$$\boxed{\sigma'(z) = \sigma(z)(1-\sigma(z))}$$

$$\boxed{\frac{\partial C}{\partial a_j^L} = a_j^L - y_j} \quad (\because C = \frac{1}{2}\sum_j (y_j - a_j^L)^2)$$

Matrix form $\quad \delta^L = \nabla_a C \odot \sigma'(z^L)$

$$\delta^L = (a^L - y) \odot \sigma'(z^L)$$

$w_{jk}^l \Rightarrow$ weight from k^{th} neuron in $(l-1)^{th}$ layer to the j^{th} neuron in l^{th} layer.

$$\boxed{a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right)}$$

Eqn for the error in δ^L in terms of error in the next layer δ^{L+1}

$$\delta^l = \left((w^{l+1})^T \delta^{L+1}\right) \odot \sigma'(z^L)$$

how? $\delta_j^l = \sum_{j=1}^{n_{l+1}} \frac{\partial C}{\partial z_j^{l+1}} \frac{\partial z_j^{l+1}}{\partial z_k^l}$

but
$$\left.\begin{aligned} \frac{\partial C}{\partial z_j^{l+1}} &= \delta_j^{l+1}; \\ z_j^{l+1} &= \sum_k w_{jk}^{l+1} a_k^l + b_j^{l+1} \\ a_k^l &= \sigma(z_k^l) \end{aligned}\right\}$$

Thus, $\frac{\partial z_j^{l+1}}{\partial z_k^l} = w_{jk}^{l+1} \sigma'(z_k^l)$

so, $\delta_j^l = \left(\sum_j w_{jk}^{l+1} \delta_j^{l+1}\right) \sigma'(z_k^l)$.

Error, $\delta_j^l = \dfrac{\partial C}{\partial z_j^l}$

Each layer computes $z^l = w^l a^{l-1} + b^l$

· Ask :- Does z_j^l touch C directly ?

There is a chain of dependency.

ie, $\underbrace{z_j^l \to a_j^l}_{①} \to \{z_k^{l+1}\} \forall k \to$ Downstream layer neurons $\to C$

②

We shall work for each links ① & ②, then combine.

① $z_j^l \to a_j^l$

$$\frac{\partial a_j^l}{\partial z_j^l} = \sigma'(z_j^l) \qquad \left[\because a_j^l = \sigma(z_j^l)\right]$$

② $a_j^l \to z_k^{l+1}$ $\qquad \left[z_k^{l+1} = \sum_i w_{ki}^{l+1} a_i^l + b_k^{l+1}\right]$

$$\frac{\partial z_k^{l+1}}{\partial a_j^l} = w_{kj}^{l+1}$$

(other terms are independent of a_j^l)

Let's combine ① & ②

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

$\hookrightarrow$ (I)

Now, we know $\delta_j^l = \frac{\partial C}{\partial z_j^l}$

in case of we using layer $(l+1)$ neurons, the cost C depends on all of these neurons. $\{z_k^{l+1}\}$ $\forall$ k

so, Total change

$$\frac{\partial C}{\partial z_j^l} = \sum_k \frac{\partial C}{\partial z_k^{l+1}} \cdot \frac{\partial z_k^{l+1}}{\partial z_j^l}$$

After substituting known values, we get

$$\frac{\partial C}{\partial z_j^l} = \sum_k \delta_k^{l+1} \cdot w_{kj}^{l+1} \sigma'(z_j^l)$$

$\sigma'(z_j^l)$ is independent of k (summation index), so,

$$\frac{\partial C}{\partial z_j^l} = \sigma'(z_j^l) \sum_k \delta_k^{l+1} w_{kj}^{l+1}$$

$$\Rightarrow \quad \delta_j^l = \left(\sum_k w_{kj}^{l+1} \delta_k^{l+1}\right) \sigma'(z_j^l)$$

The rule for matrix multiplication forced us to take W^T

$$\Rightarrow \quad \delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l).$$