

Gradient Descent & Back prop

n :- Number of neurons

W :- layer weight matrix

b :- layer bias vector

a :- layer activation vector

z :- vector of weighted inputs from the ~~previous~~ previous layer before activation

σ :- layer activation function

$$\hookrightarrow \sigma(x) = \frac{1}{1 + e^{-x}} \quad ; \text{ sigmoid } \downarrow$$

square error

C :- cost function ; ~~se~~ $\frac{1}{2} \sum (y - \hat{y})^2$
 \uparrow
 $\hat{y} = h(x)$

* Backprop goal: Find the partial derivatives ~~max~~ of weights & biases w.r.t prediction cost

* with partial derivatives

$$\hookrightarrow w_{k+1}^l = w_k^l - \alpha \frac{\partial C}{\partial w_k^l}$$

$$\hookrightarrow b_{k+1}^l = b_k^l - \alpha \frac{\partial C}{\partial b_k^l}$$

* Forward pass

$$a^{(0)} = x$$

$$z^1 = w^1 a^0 + b^1$$

$$\frac{\partial C}{\partial w^1}, \frac{\partial C}{\partial b^1}$$

$$\frac{\partial C}{\partial w^1} = \frac{\partial C}{\partial a^3} \frac{\partial a^3}{\partial z^3} \frac{\partial z^3}{\partial z^2} \frac{\partial z^2}{\partial a^2} \frac{\partial a^2}{\partial z^1} \frac{\partial z^1}{\partial w^1}$$

$$a^1 = \sigma(z^1)$$

$$z^2 = w^2 a^1 + b^2$$

$$\frac{\partial C}{\partial w^2}, \frac{\partial C}{\partial b^2}$$

$$a^2 = \sigma(z^2)$$

$$z^3 = w^3 a^2 + b^3$$

$$\frac{\partial C}{\partial w^3}, \frac{\partial C}{\partial b^3}$$

$$a^3 = \sigma(z^3) = \hat{y}$$

$$C = \frac{1}{2} (y - \hat{y})^2$$

Square error per sample

$$C = \frac{1}{2} (y - \sigma(z(w^3, b^3, \sigma(z(w^2, b^2, \sigma(z(w^1, b^1, x))))))^2$$

* Because of the function composition property, the chain rule can be used to find the weight, bias partial derivatives

* Partial derivatives required for back prop of gradients

$$\begin{array}{ll}
 1 \quad \frac{\partial C}{\partial z^L} & z^L \rightarrow 0 \rightarrow C \\
 & \quad \quad \quad \sigma(z^L) \quad \downarrow y \\
 2 \quad \frac{\partial C}{\partial z^l} & z^l \rightarrow 0 \dots \dots 0 \rightarrow C \\
 & \quad \quad \quad \sigma(z^l) \quad \sigma(z^l) \\
 3 \quad \frac{\partial z^l}{\partial w^l} & w^{l, l-1} \rightarrow z^l \\
 4 \quad \frac{\partial z^l}{\partial b^l} & b^l \rightarrow z^l
 \end{array}$$

$$1. \frac{\partial C}{\partial z^L} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L} = \frac{\partial C}{\partial a^L} \circ \sigma'(z^L)$$

derivative of activation

$$\hookrightarrow \frac{\partial C}{\partial a^L} = -(y - \hat{y})$$

$$\hookrightarrow \frac{\partial a^L}{\partial z^L}; \quad a^L = \sigma(z^L) = \frac{1}{1 + e^{-z^L}}$$

$$\frac{\partial a^L}{\partial z^L} = \frac{\partial \sigma(z^L)}{\partial z^L} = \frac{\partial}{\partial z^L} \left(\frac{1}{1 + e^{-z^L}} \right)$$

* using Reciprocal rule

$$\hookrightarrow \frac{\frac{d}{dx} \left(\frac{1}{f(x)} \right)}{\frac{d}{dx} f(x)} = - \frac{\frac{df(x)}{dx}}{f(x)^2} \bigg| \frac{\partial a^L}{\partial z^L} = - \frac{\frac{\partial (1 + e^{-z^L})}{\partial z^L}}{(1 + e^{-z^L})^2}$$

$$= - \frac{1}{(1 + e^{-z^L})^2} \frac{\partial (1 + e^{-z^L})}{\partial z^L}$$

$\frac{d e^x}{dx} = e^x$

* Exponent rule

$$\begin{aligned} \frac{\partial a^L}{\partial z^L} &= - \frac{1}{(1 + e^{-z^L})^2} \left(\frac{\partial -z^L}{\partial z^L} \right) \\ &= - \frac{e^{-z^L}}{(1 + e^{-z^L})^2} \end{aligned}$$

$$\begin{aligned}
 \hookrightarrow \frac{\partial a^L}{\partial z^L} &= \frac{e^{-z^L} + 1 - 1}{(e^{-z^L} + 1)^2} \\
 &= \frac{1}{(1 + e^{-z^L})} \left[\frac{e^{-z^L} + 1}{e^{-z^L} + 1} - \frac{1}{e^{-z^L} + 1} \right] \\
 &= \sigma(z^L) [1 - \sigma(z^L)]
 \end{aligned}$$

$$\begin{aligned}
 \text{So } \frac{\partial C}{\partial z^L} &= -(y - \hat{y}) \sigma(z^L) [1 - \sigma(z^L)] \\
 &= -(y - a^L) a^L [1 - a^L]
 \end{aligned}$$

$$2. \frac{\partial C}{\partial z^l} = \frac{\partial C}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial a^l} \frac{\partial a^l}{\partial z^l}$$

y

$$z^{l+1} = w^{l+1} a^l + b^{l+1}$$

$$\frac{\partial z^{l+1}}{\partial a^l} = w^{l+1}$$

$$\frac{\partial a^l}{\partial z^l} = \sigma'(z^l) = \frac{\partial \sigma(z^l)}{\partial z^l}$$

$$\hookrightarrow \text{so } \frac{\partial C}{\partial z^l} = \left[(w^{l+1})^T \frac{\partial C}{\partial z^{l+1}} \right] \odot \sigma'(z^l)$$

$$3/4. \frac{\partial C}{\partial z^l} = \frac{\partial C}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial a^l} \frac{\partial a^l}{\partial z^l} \left(\frac{\partial z^l}{\partial w^l} / \frac{\partial z^l}{\partial b^l} \right)$$

$$\hookrightarrow \frac{\partial z^l}{\partial w^l} = w^l a^{l-1} + b^l$$

$$= a^{l-1}$$

$$\hookrightarrow \frac{\partial z^l}{\partial b^l} = 1$$