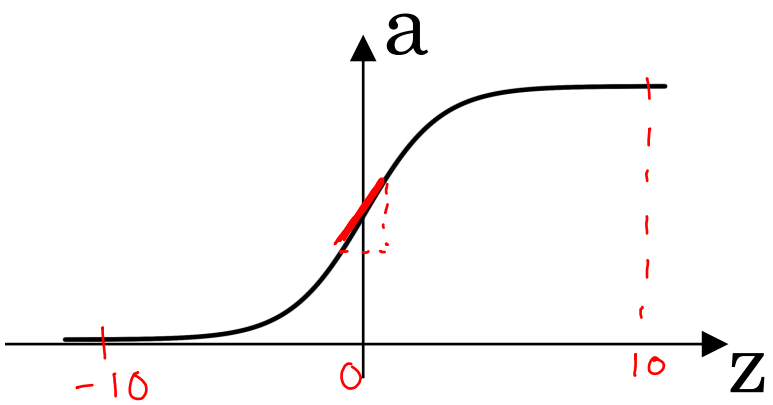
deeplearning.ai

One hidden layer Neural Network

Derivatives of activation functions

Sigmoid activation function

When you do back prop, $\frac{\partial J}{\partial a}$, you need to compute derivative of Activations "a", ie, $g(z)$



$$g(z) = \frac{1}{1+e^{-z}}$$

Say $z = 10$
$g(z) \approx 1$
$\frac{\partial g(z)}{\partial z} = 1 \cdot (1-1) \approx 0$
which is true as slope ≈ 0 at $z = 10$

$\frac{\partial J}{\partial a} \rightarrow \partial a = \frac{\partial g(z)}{\partial z}$ = slope of $g(z)$ or "a" at pt "z" ($g'(z)$)

$z = -10$, $g(z) = 0$
$\frac{\partial g(z)}{\partial z} = 0 \cdot (1-0) = 0$

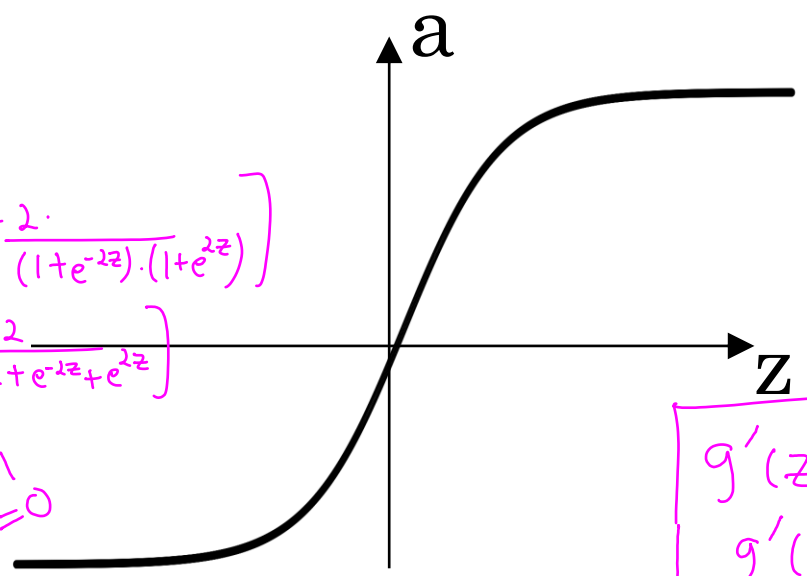
If $z = 0$, $g(z) = 1/2$
then $\frac{\partial g(z)}{\partial z} = \frac{1}{2}(1-1/2) = \frac{1}{4}$

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

$$g'(z) = a(1-a)$$

Andrew Ng

Tanh activation function



$$g(z) = \tanh(z)$$

$$g'(z) = 1 - (\tanh(z))^2$$

$$g'(z) = 1 - a^2$$

$1-(\tanh(z))^2$

$1-\left[\frac{1}{1+e^{-2z}} - \frac{1}{1+e^{2z}}\right]^2$

$1-\left[\left(\frac{1}{1+e^{-2z}}\right)^2 + \left(\frac{1}{1+e^{2z}}\right)^2 - 2\cdot\frac{}{(1+e^{-2z})\cdot(1+e^{2z})}\right]$

$1-\left[(1+e^{-2z})^{-2} + (1+e^{2z})^{-2} - \frac{2}{2+e^{-2z}+e^{2z}}\right]$

LHS

To prove LHS = RHS

put $z=0$, LHS = RHS = 1

put $z=100$, LHS = RHS = 0

Hence proved

$\frac{d}{dz} g(z)$ = slope of $g(z)$ at "z"

$= \frac{d}{dz}\left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right)$

$= \frac{1}{1+e^{-2z}} - \frac{1}{e^{2z}+1}$

$= \frac{1}{1+e^{-2z}} - \frac{1}{1+e^{2z}}$

$= \frac{d}{dz}\left(\frac{1}{1+e^{-2z}}\right) - \frac{d}{dz}\left(\frac{1}{1+e^{2z}}\right)$

$= \frac{d}{dz}(1+e^{-2z})^{-1} - \frac{d}{dz}(1+e^{2z})^{-1}$

(A) $(-1)(1+e^{-2z})^{-2}\cdot e^{-2z}\cdot(-2)$

$= 2(1+e^{-2z})^{-2}\cdot e^{-2z}$

(B) $(-1)(-1)\cdot(1+e^{2z})^{-2}\cdot e^{2z} * 2$

$= 2(1+e^{2z})^{-2}\cdot e^{2z}$

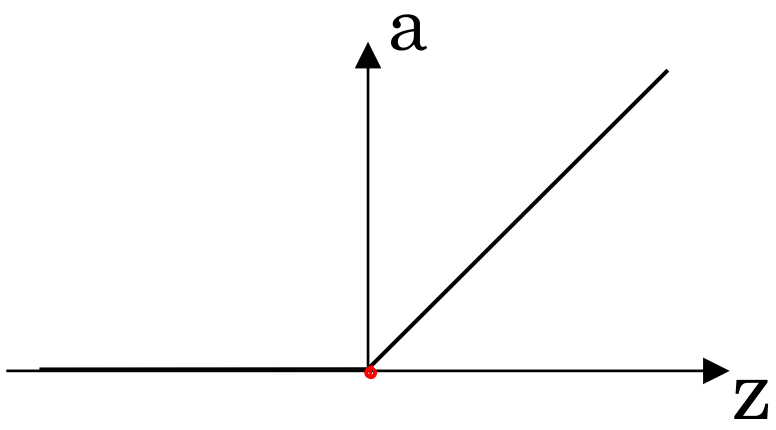
A + B

$= 2\left[(1+e^{-2z})^{-2}\cdot e^{-2z} + (1+e^{2z})^{-2}\cdot e^{2z}\right]$

RHS

Andrew Ng

ReLU and Leaky ReLU



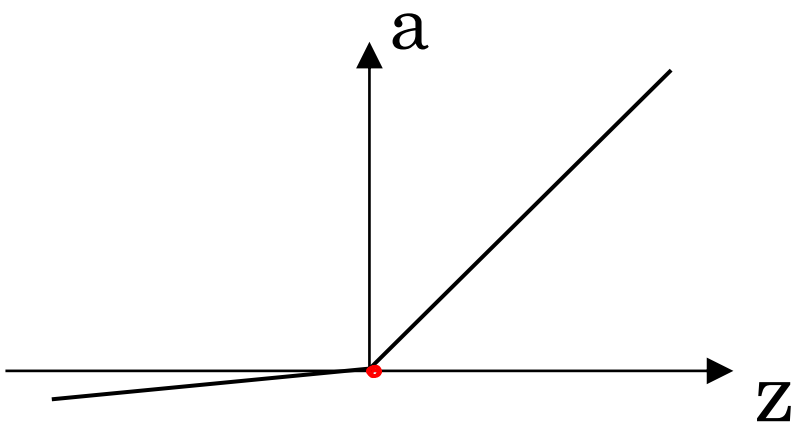
ReLU

$g(z) = \max(0, z)$

$g'(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z > 0 \end{cases}$

undefined at $z = 0$

↳ Highly unlikely to have hidden unit w/ value 0.0000…



Leaky ReLU

$g(z) = \max(0.01 \cdot z, z)$

$g'(z) = \begin{cases} 0.01 & \text{if } z < 0 \\ 1 & \text{if } z > 0 \end{cases}$

undefined at $z = 0$

whenever $z = 0$, then
By default make $g'(z) = 1$
(Its fine, no big deal)

Andrew Ng