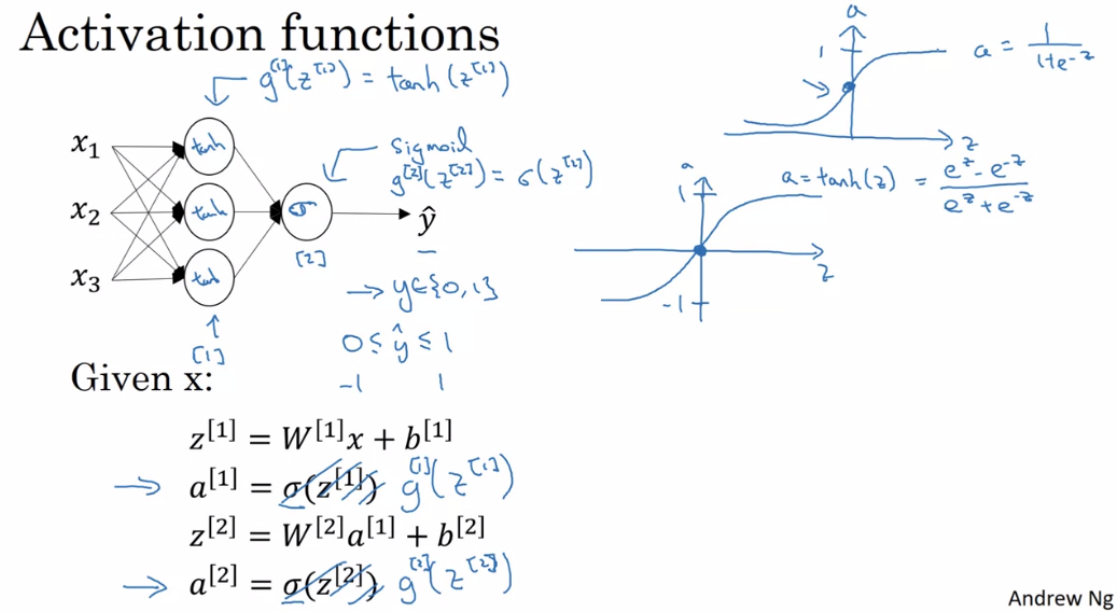
Week 3-2 Activation Functions

笔记本:	DL 1 - NN and DL		
创建时间:	2021/1/8 11:25	**更新时间:**	2021/1/8 11:49

sigmoid

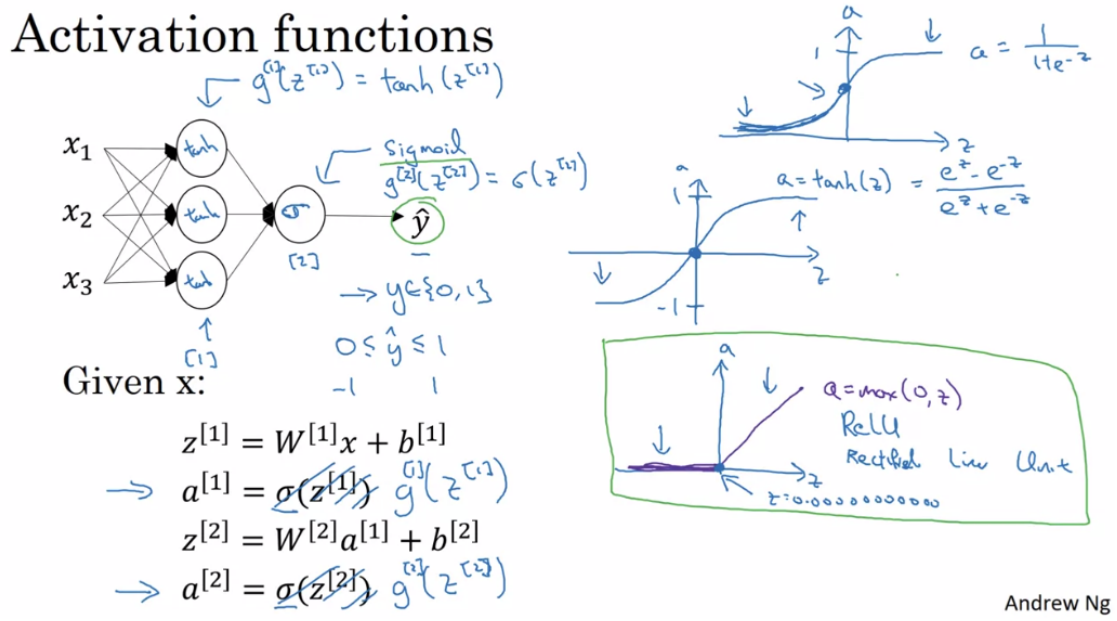
tanh -> almost always better than sigmoid (centering around 0)

but one exception: we should use sigmoid for output ({0,1})

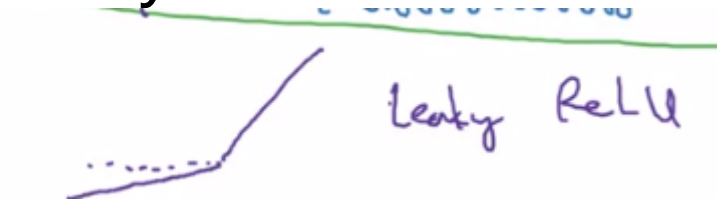


ReLU

(Rectified Linear Unit)

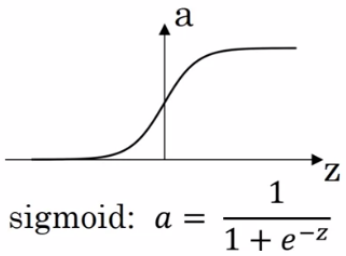


Leaky ReLU

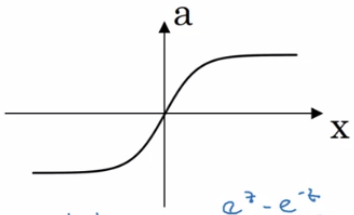


Adv: for a lot of the space of Z, the derivative of the activation function, the slope of the activation function **is very different from 0**. And so in practice, using the ReLU activation function, your neural network will often learn much faster than when using the tanh or the sigmoid activation function.

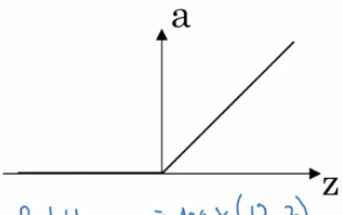
Pros and cons of activation functions



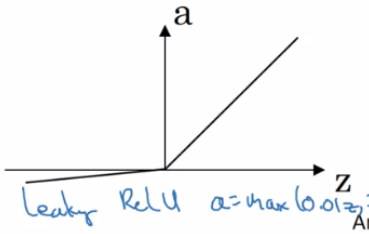
sigmoid: $a = \frac{1}{1+e^{-z}}$



tanh: $a = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$



ReLU $a = \max(0, z)$



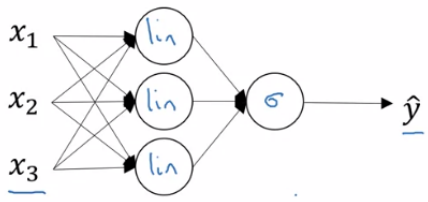
Leaky ReLU $a = \max(0.01z, z)$

Andrew Ng

Why do we need Activiation function?

If we use linear activation, the output would always be linear combination of the input

Activation function



Given x:

$$z^{[1]} = W^{[1]}x + b^{[1]}$$
$$a^{[1]} = g^{[1]}(z^{[1]})$$
$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$
$$a^{[2]} = g^{[2]}(z^{[2]})$$

$g(z) = z$ "linear activation function"

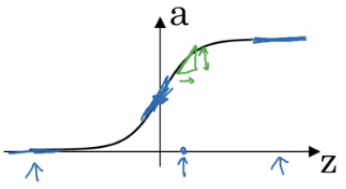
$$a^{[1]} = z^{[1]} = W^{[1]}x + b^{[1]}$$
$$a^{[2]} = z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$
$$a^{[2]} = W^{[2]}\left(W^{[1]}x + b^{[1]}\right) + b^{[2]}$$
$$= \underbrace{\left(W^{[2]}W^{[1]}\right)}_{W'}x + \underbrace{\left(W^{[2]}b^{[1]} + b^{[2]}\right)}_{b'}$$
$$= W'x + b'$$

Andrew Ng

---

derivatives

Sigmoid activation function



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

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

$$g'(z) = \frac{d}{dz}g(z) = \text{slope of } g(z) \text{ at } z$$
$$= \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right)$$
$$= g(z)\left(1 - g(z)\right)$$
$$= a(1-a)$$

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

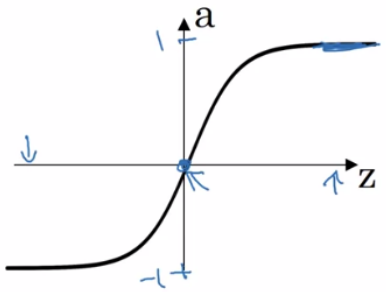
$z = 10$. $g(z) \approx 1$
$\frac{d}{dz}g(z) \approx 1\,(1-1) \approx 0$

$z = -10$ $g(z) \approx 0$
$\frac{d}{dz}g(z) \approx 0 \cdot (1-0) \approx 0$

$z = 0$ $g(z) = \frac{1}{2}$
$\frac{d}{dz}g(z) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$

Andrew Ng

Tanh activation function



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

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

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

$= 1 - (\tanh(z))^2$ ←

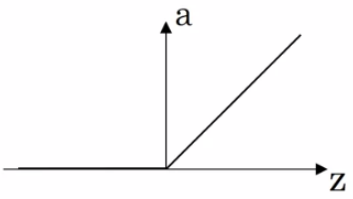
$a = g(z)$, $g'(z) = 1 - a^2$

$z = 10$ $\tanh(z) \approx 1$
$g'(z) \approx 0$

$z = -10$ $\tanh(z) \approx -1$
$g'(z) \approx 0$

$z = 0$ $\tanh(z) = 0$
$g'(z) = 1$

ReLU and Leaky ReLU



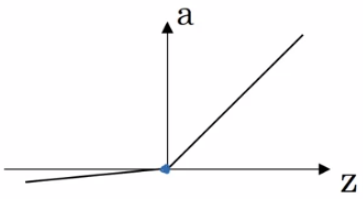
ReLU

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

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

~~undef if $z = 0$~~

$z = 0.0000\ldots0$



Leaky ReLU

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

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

Andrew Ng