deeplearning.ai

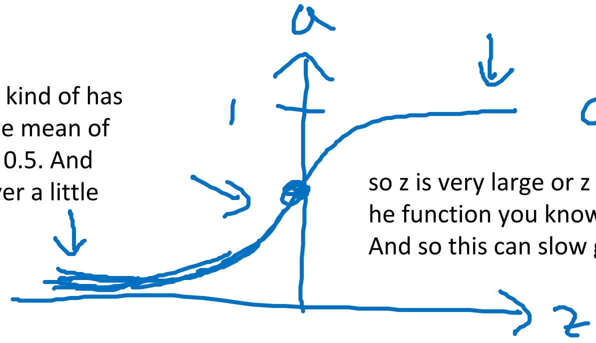
One hidden layer Neural Network

Activation functions

Activation functions

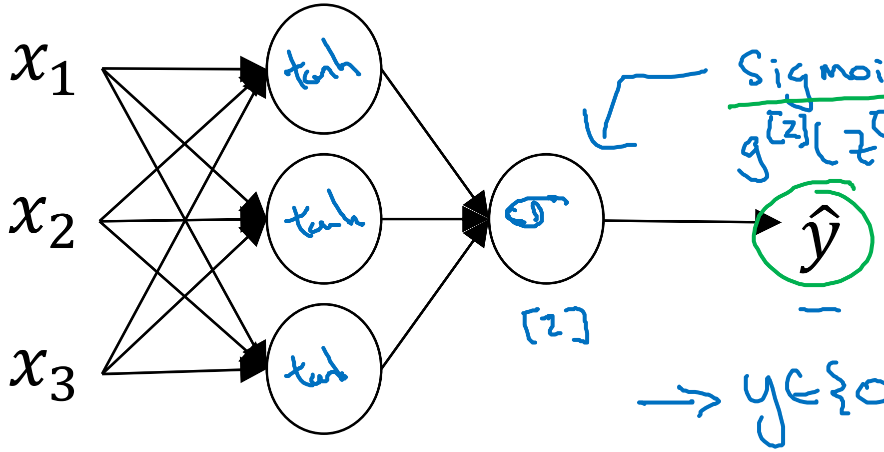
$g^{[1]}(z^{[1]}) = \tanh(z^{[1]})$

using a tan(h) instead of a sigmoid function kind of has the effect of centering your data, so that the mean of the data is close to zero rather than maybe 0.5. And this actually makes learning for the next layer a little bit easier.



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

so z is very large or z is very small, the slope of t he function you know ends up being close to 0. And so this can slow gradient.



Sigmoid

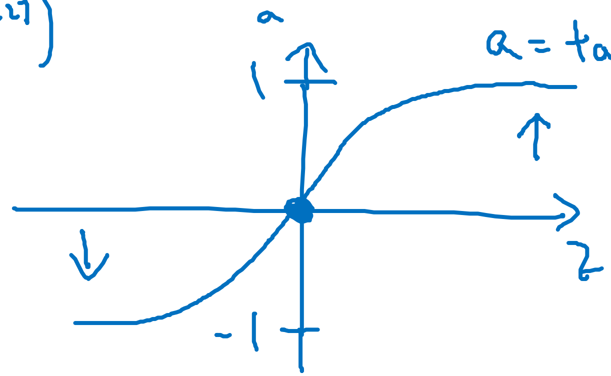
$g^{[2]}(z^{[2]}) = \sigma(z^{[2]})$

$\rightarrow y \in \{0, 1\}$

$0 \leq \hat{y} \leq 1$

$-1 \qquad 1$

the tanh function is almost always strictly superior the one exception is for the output layer



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

Either one should be fine although if you had to pick one. And advantage of both the ReLU and the leaky ReLU is that for a lot of space of z, the derivative of the activation function, the slope of the activation function is very different from zero, and so in practice using the ReLU activation function, your neural network usually learning faster than using the tanh or the sigmoid activation function. The reason is that there is less of this effect of the slope of the function going to zero which slow the learning.

Given x:

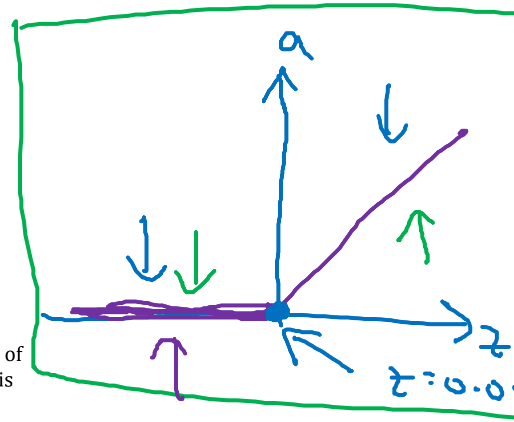
$$z^{[1]} = W^{[1]}x + b^{[1]}$$

$$\rightarrow a^{[1]} = \cancel{\sigma(z^{[1]})} \; g^{[1]}(z^{[1]})$$

$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$

$$\rightarrow a^{[2]} = \cancel{\sigma(z^{[2]})} \; g^{[2]}(z^{[2]})$$

for half of the range of z, the slope of ReLU is zero but in practice, enough of your hidden units will have z greater than zero. So learning is still be quite fast for most training example

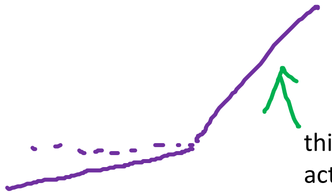


$z > 0$

$a = \max(0, z)$

ReLU

Rectified Linear Unit

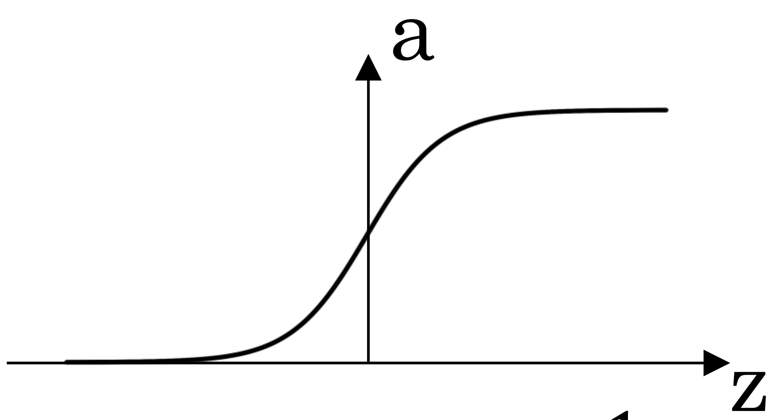


Leaky ReLU

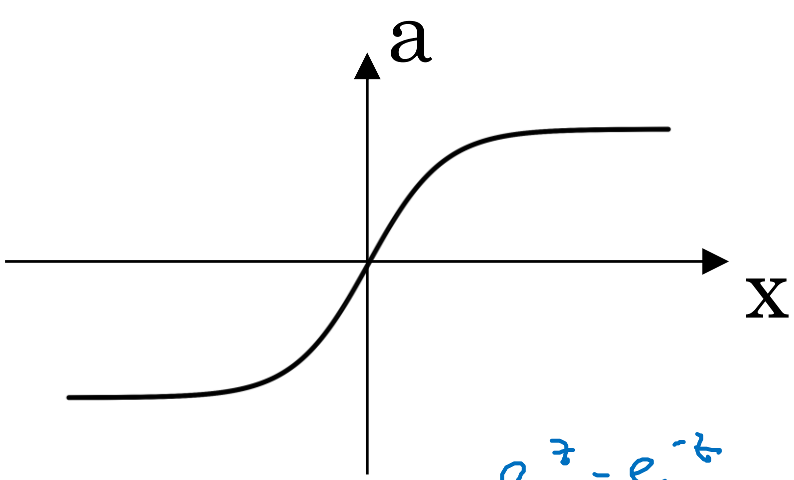
this usually works better than the ReLU activation function, although it's just not used as much in practice.

Andrew Ng

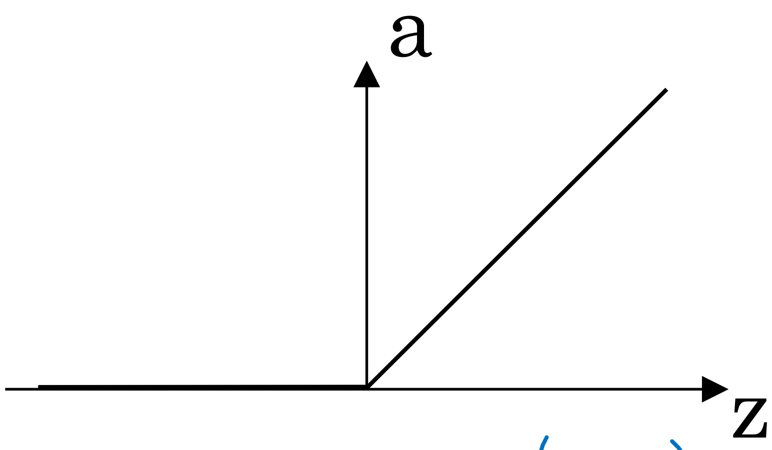
Pros and cons of activation functions



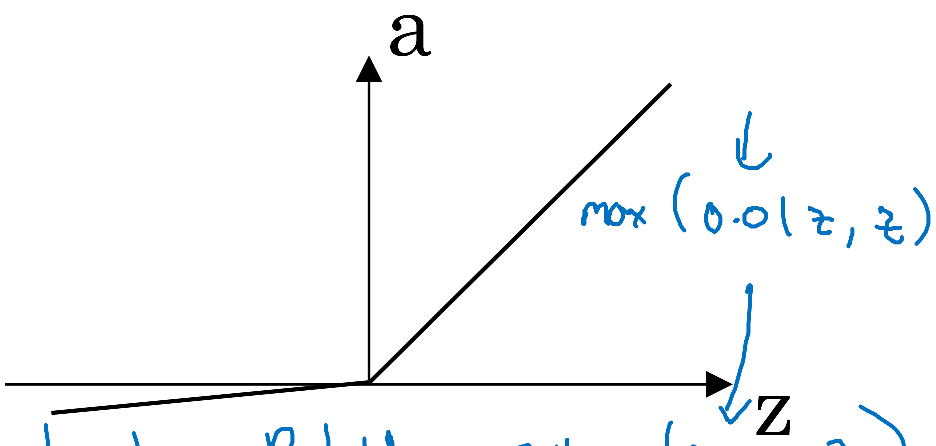
sigmoid: $a = \dfrac{1}{1 + e^{-z}}$



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



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



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

Andrew Ng