

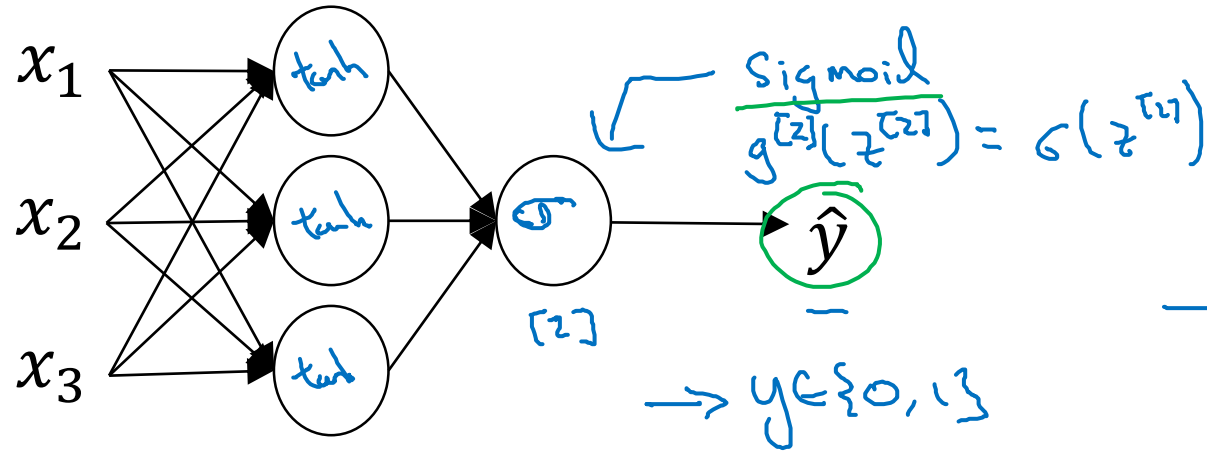


deeplearning.ai

One hidden layer Neural Network

Activation functions

Activation functions



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

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

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

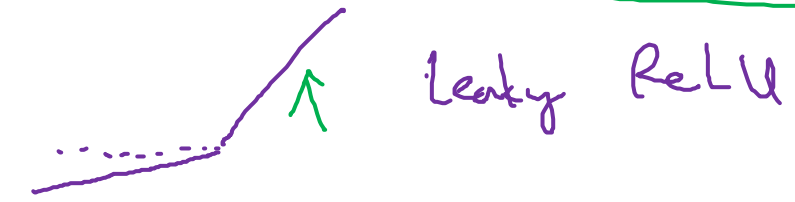
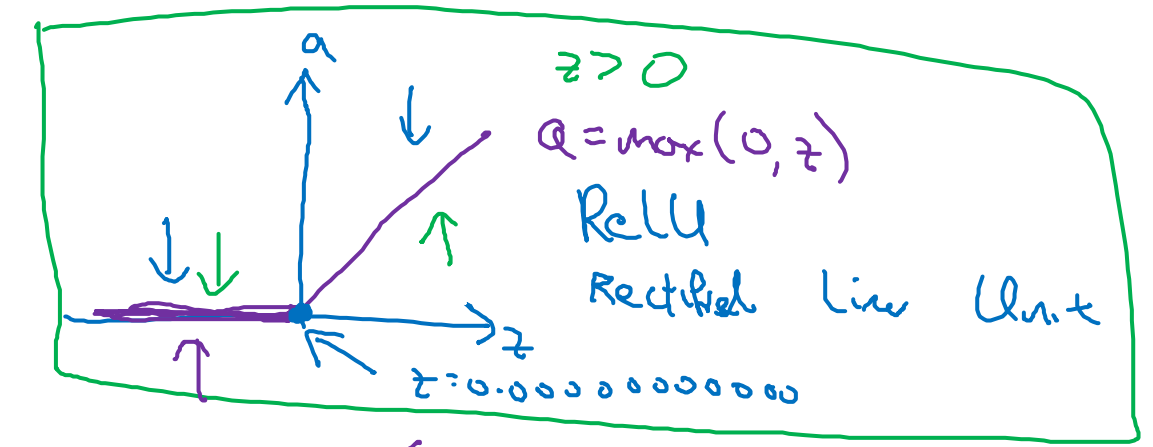
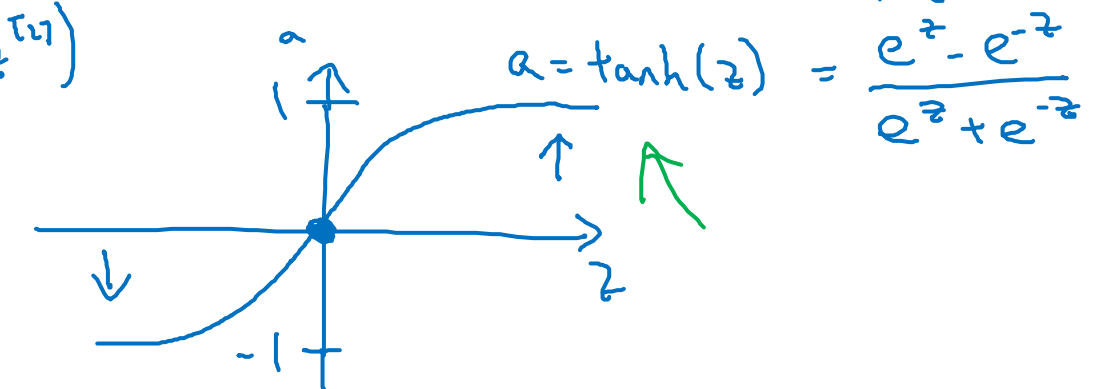
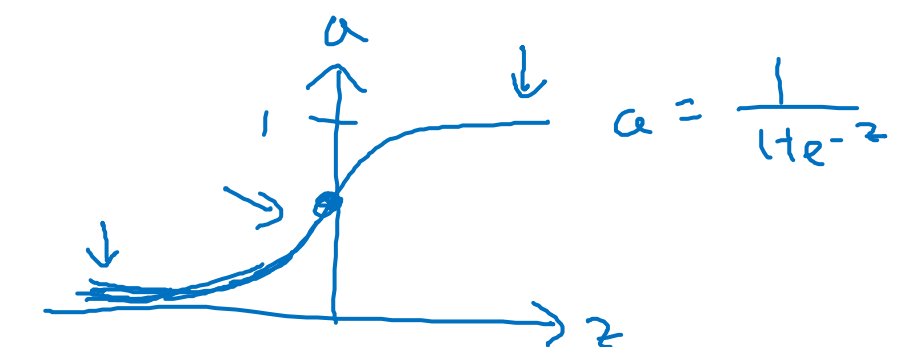
Given x :

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

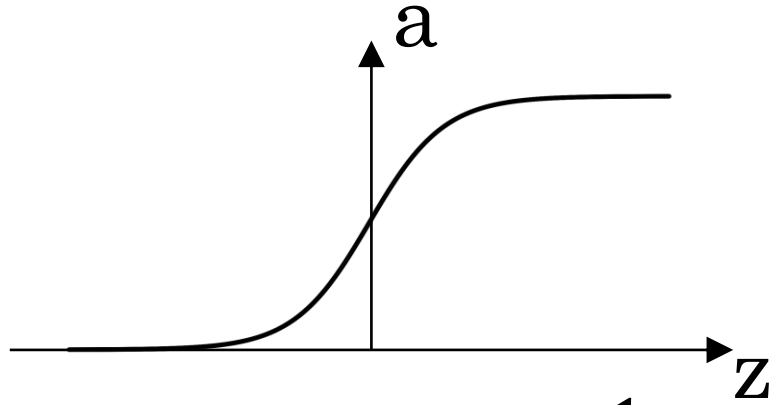
$\rightarrow a^{[1]} = \cancel{\sigma(z^{[1]})} g^{(1)}(z^{(1)})$

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

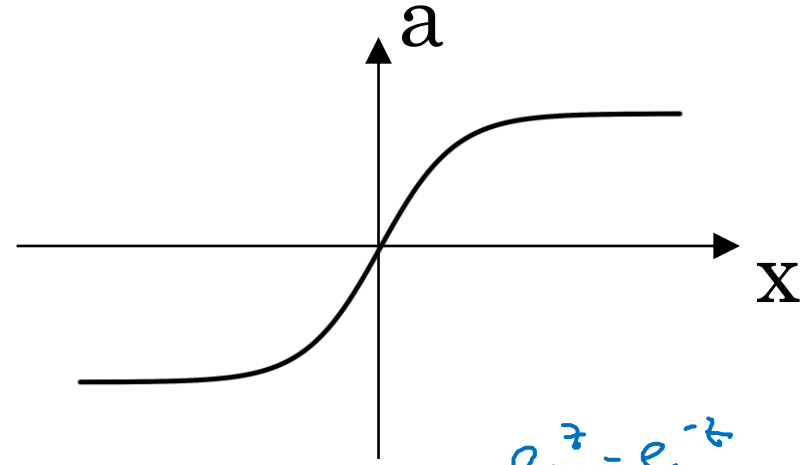
$\rightarrow a^{[2]} = \cancel{\sigma(z^{[2]})} g^{(2)}(z^{(2)})$



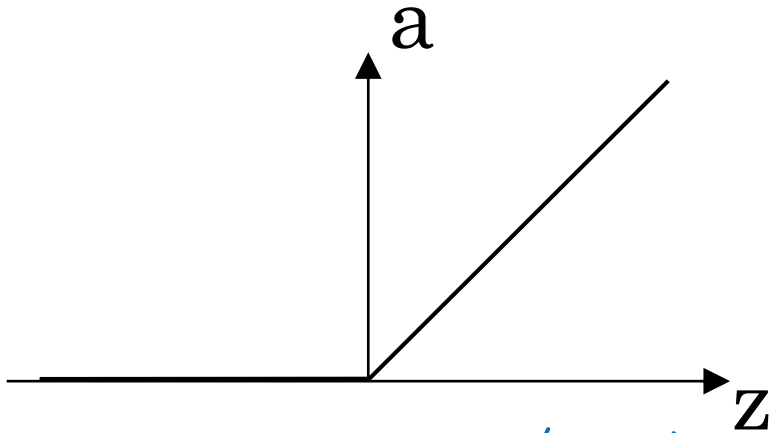
Pros and cons of activation functions



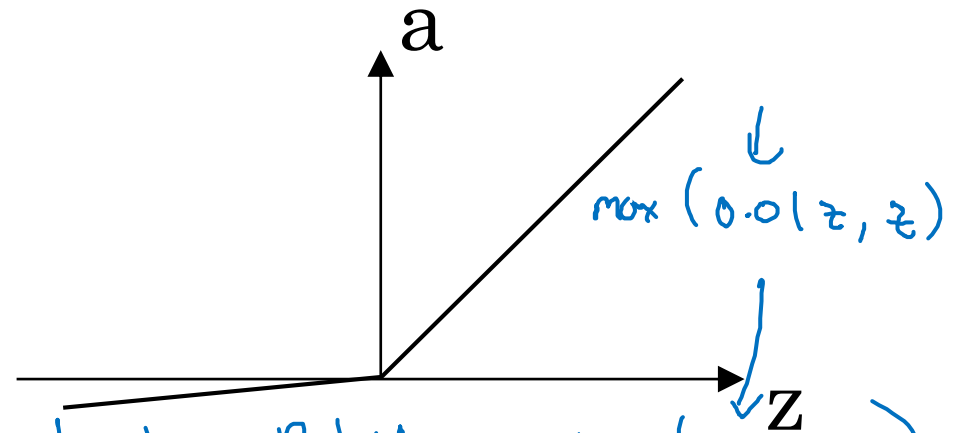
sigmoid: $a = \frac{1}{1 + e^{-z}}$



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



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



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