One hidden layer
Neural Network

Activation functions

deeplearning.ai

Activation functions

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

x_1

x_2

x_3

tanh

tanh

tanh

σ

$\hat{y}$

[2]

[1]

Sigmoid

$g^{[2]}(z^{[2]}) = \sigma(z^{[1]})$

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

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

$-1 \qquad 1$

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

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

-1

$z > 0$

$a = \max(0, z)$
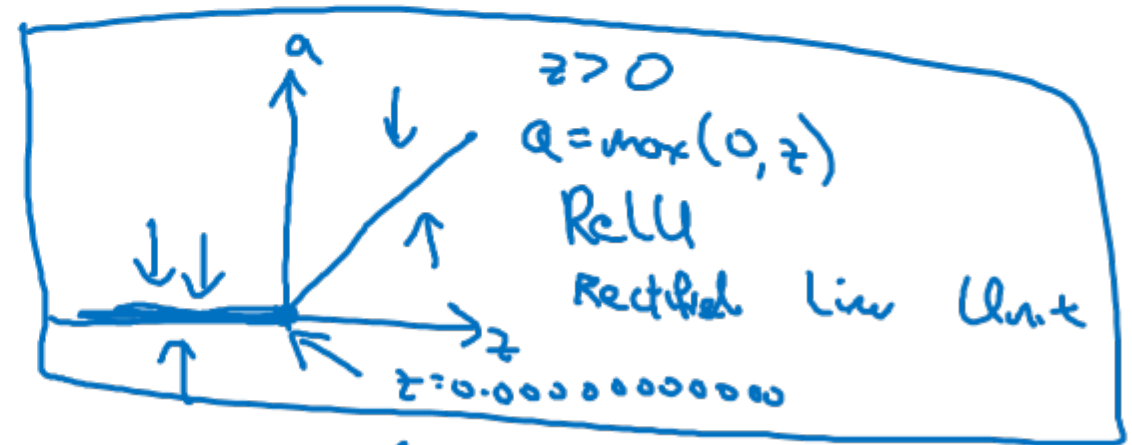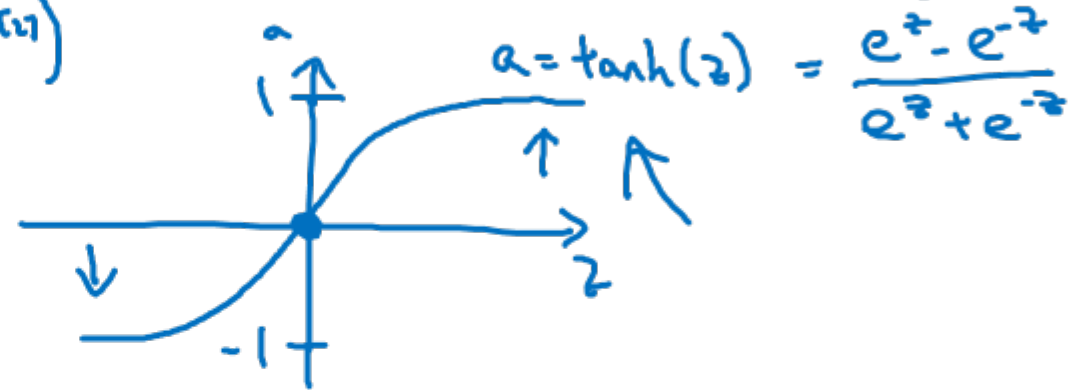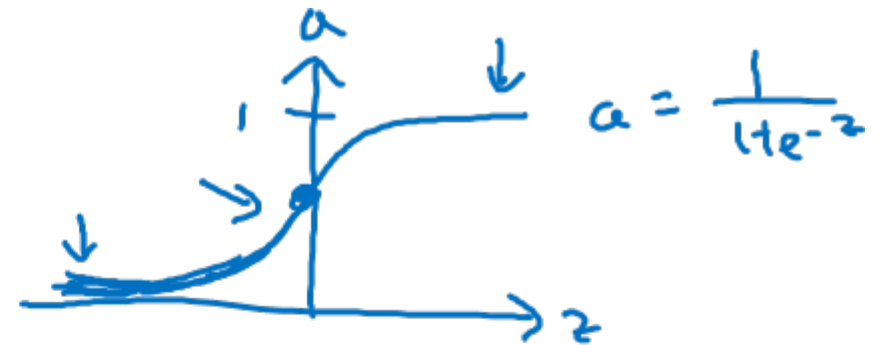
ReLU
Rectified Linear Unit

$z = 0.0000000000$
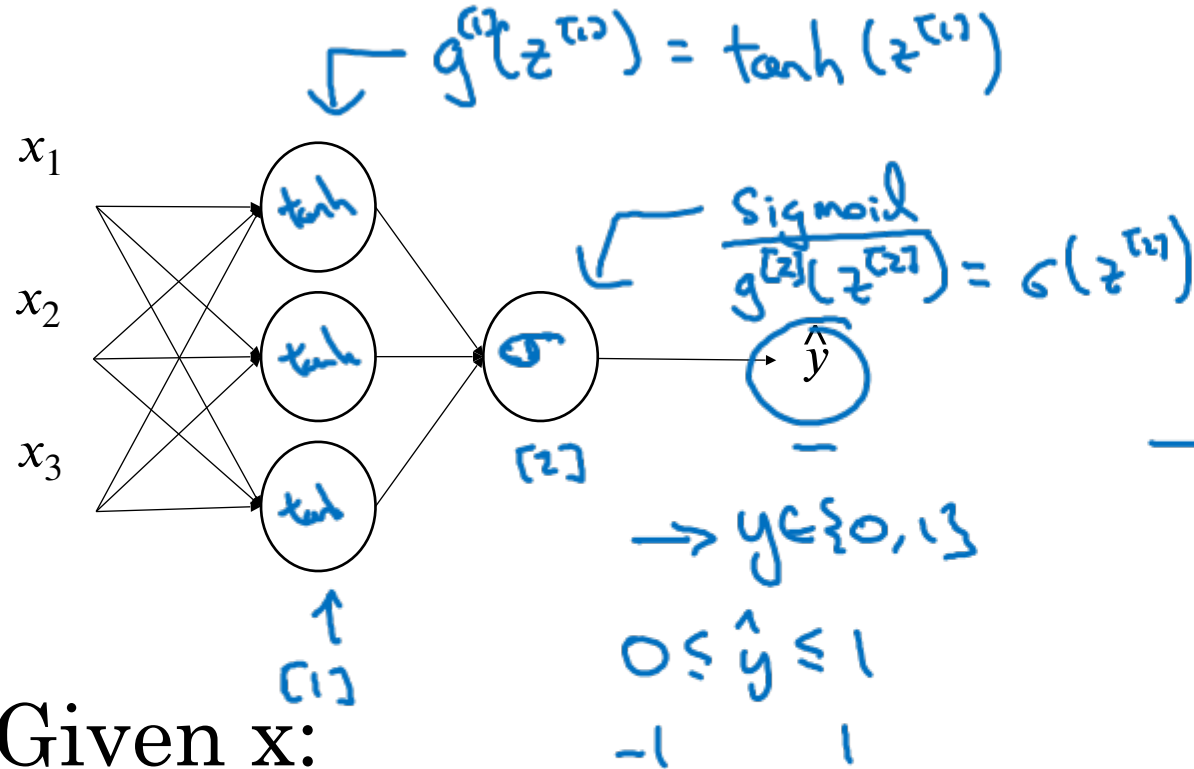
leaky ReLU

Given x:

$z^{[1]} = W^{[1]}x + b^{[1]}$
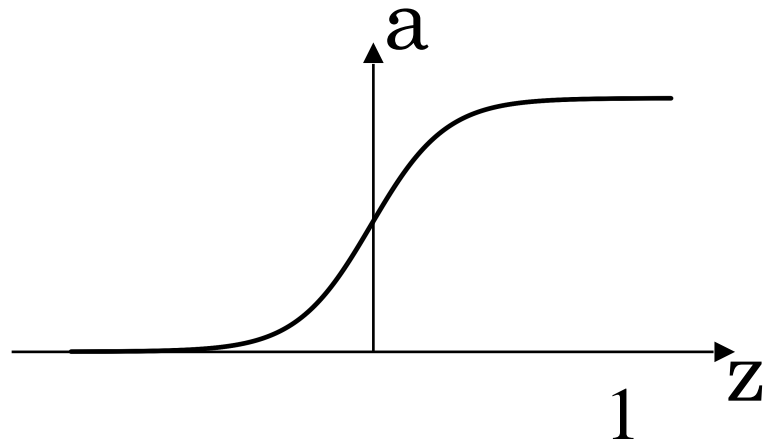
$\rightarrow a^{[1]} = \sigma(z^{[1]}) \quad g^{[1]}(z^{[1]})$

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

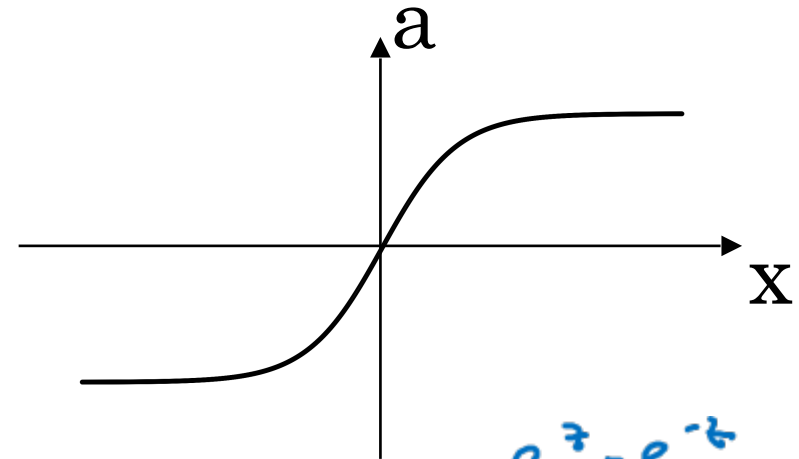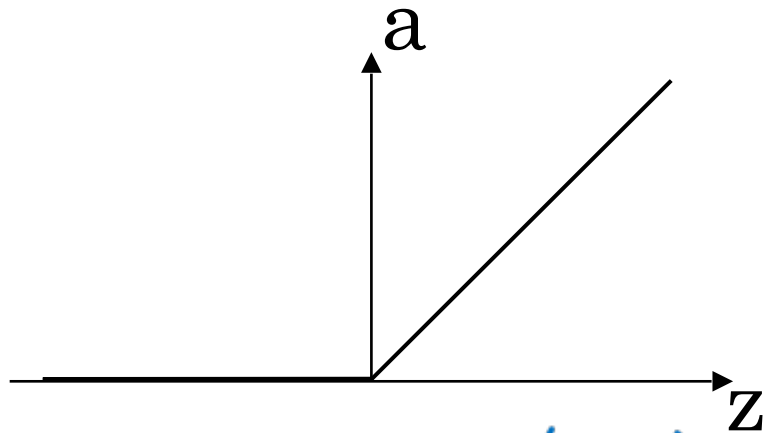$\rightarrow a^{[2]} = \sigma(z^{[2]}) \quad g^{[2]}(z^{[2]})$
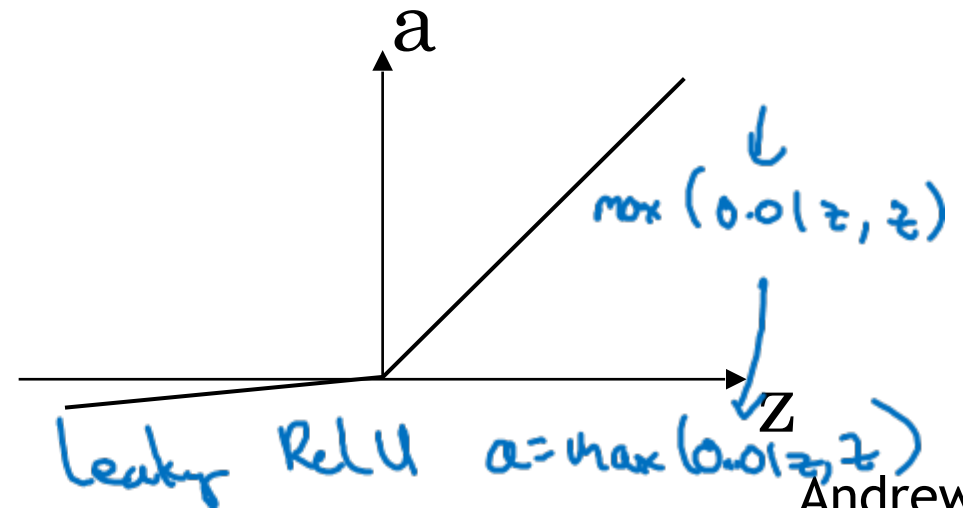
Andrew Ng

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

$\max(0.01z, z)$