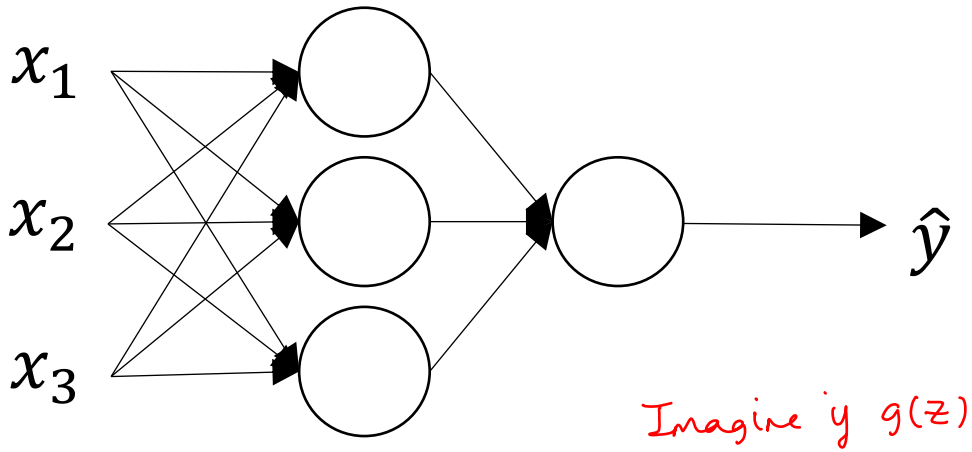
deeplearning.ai

One hidden layer Neural Network

Why do you need non-linear activation functions?

Activation function



Imagine if $g(z) = z$
ie, linear Activation func
or Identity " "

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

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

$a^{[2]} = z^{[2]} = (w^{[2]} \cdot w^{[1]}) \cdot X + w^{[2]} \cdot b^{[1]} + b^{[2]}$

$\Rightarrow a^{[2]} = z^{[2]} = w'x + b'$

$\Rightarrow$ linear

$\Rightarrow$ it doesnt matter how many layers there are, the NN will just be as good as a linear regression

Given x:

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

Say $a^{[1]} = g^{[1]}(z^{[1]}) = z^{[1]}$, then lets see what happens when we use a linear Activation func

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

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