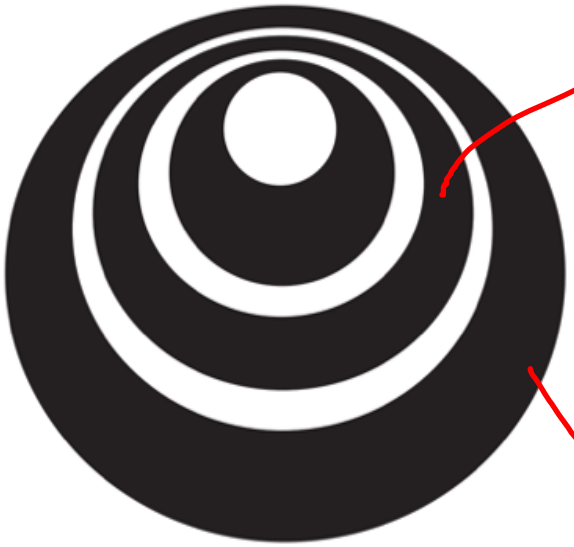
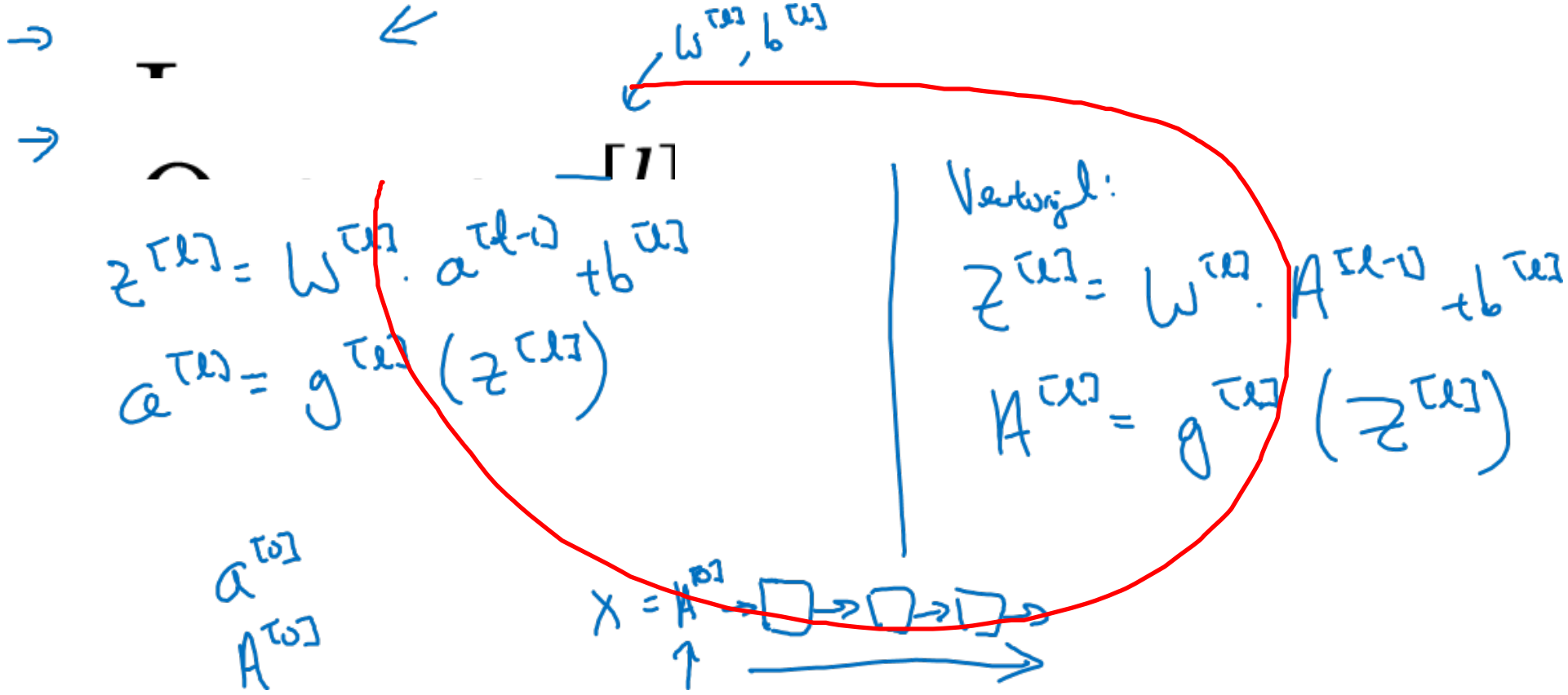


deeplearning.ai

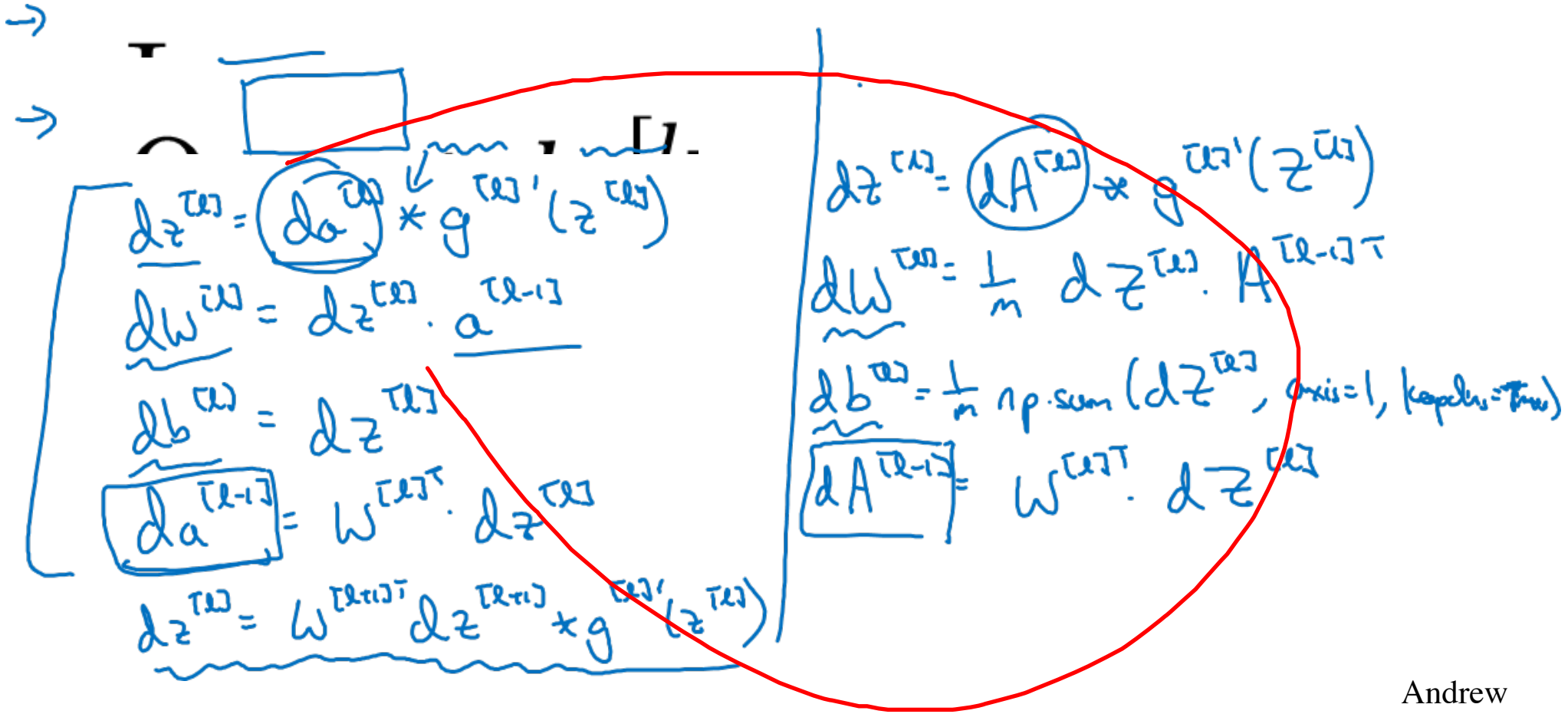
Deep Neural Networks

Forward and backward propagation

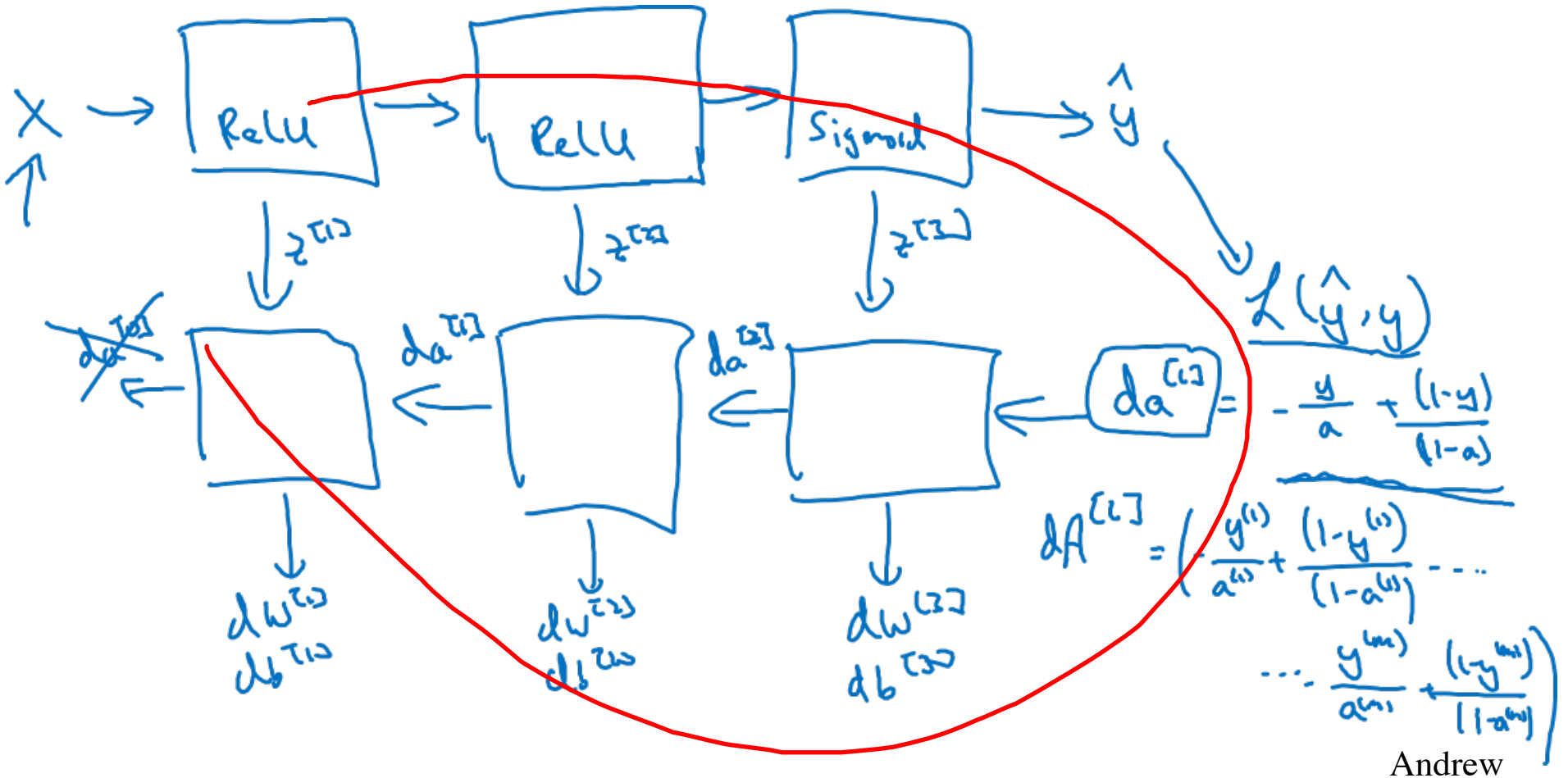
Forward propagation for layer l



Backward propagation for layer l



Summary

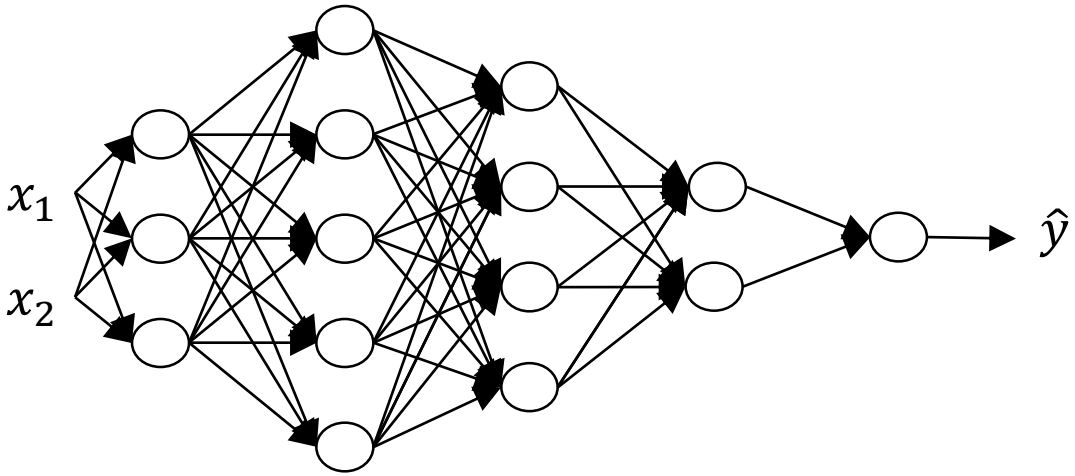

X
ReLU
ReLU
Sigmoid
ŷ
$z^{[1]}$
$z^{[2]}$
$z^{[3]}$
$da^{[2]}$
$da^{[1]}$
$da^{[0]}$
$dW^{[1]}$ $db^{[1]}$
$dW^{[2]}$ $db^{[2]}$
$dW^{[3]}$ $db^{[3]}$
$\mathcal{L}(\hat{y}, y)$
$da^{[3]} = -\frac{y}{a} + \frac{(1-y)}{(1-a)}$
$dA^{[3]} = \left(-\frac{y^{(1)}}{a^{(1)}} + \frac{(1-y^{(1)})}{(1-a^{(1)})} \cdots \right.$
$\left. \cdots \frac{y^{(m)}}{a^{(m)}} + \frac{(1-y^{(m)})}{(1-a^{(m)})} \right)$

deeplearning.ai

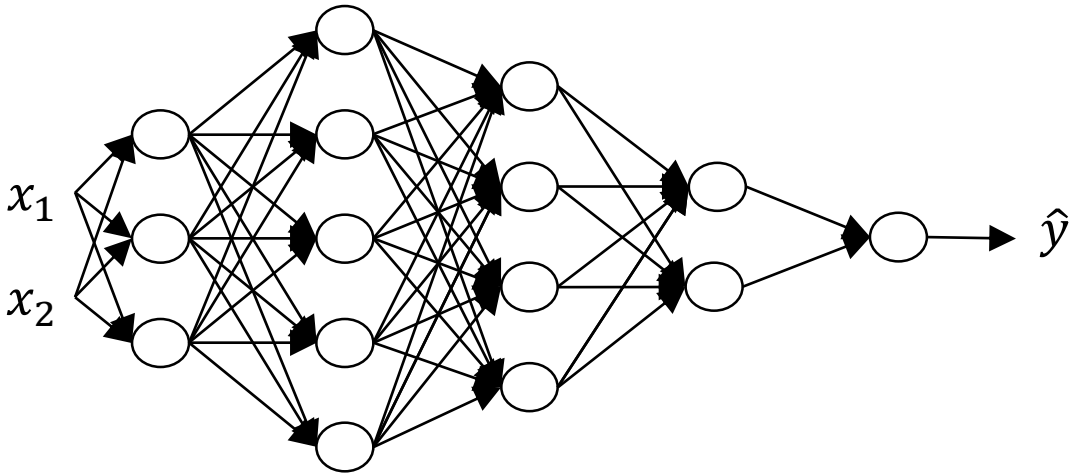
Deep Neural Networks

Getting your matrix dimensions right

Parameters $W^{[l]}$ and $b^{[l]}$



Vectorized implementation

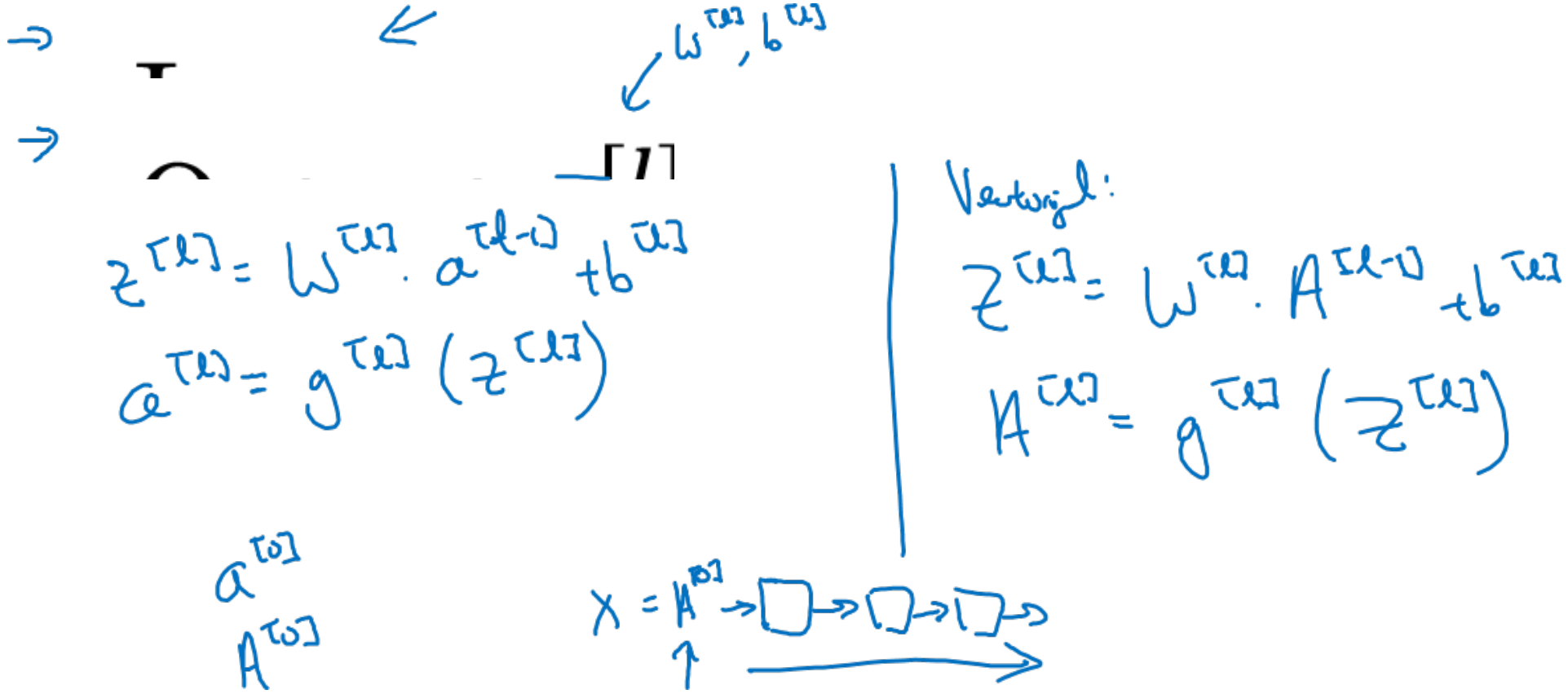


deeplearning.ai

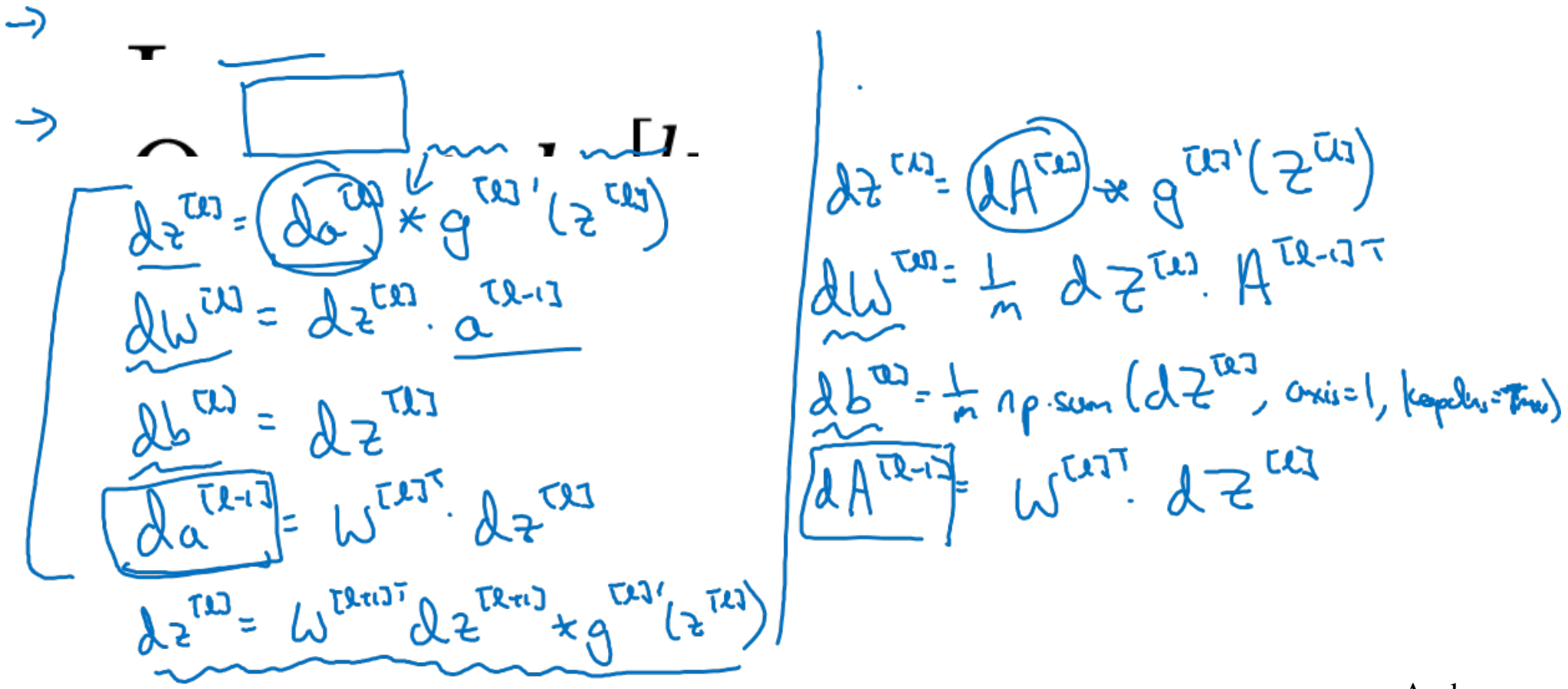
Deep Neural Networks

Forward and backward propagation

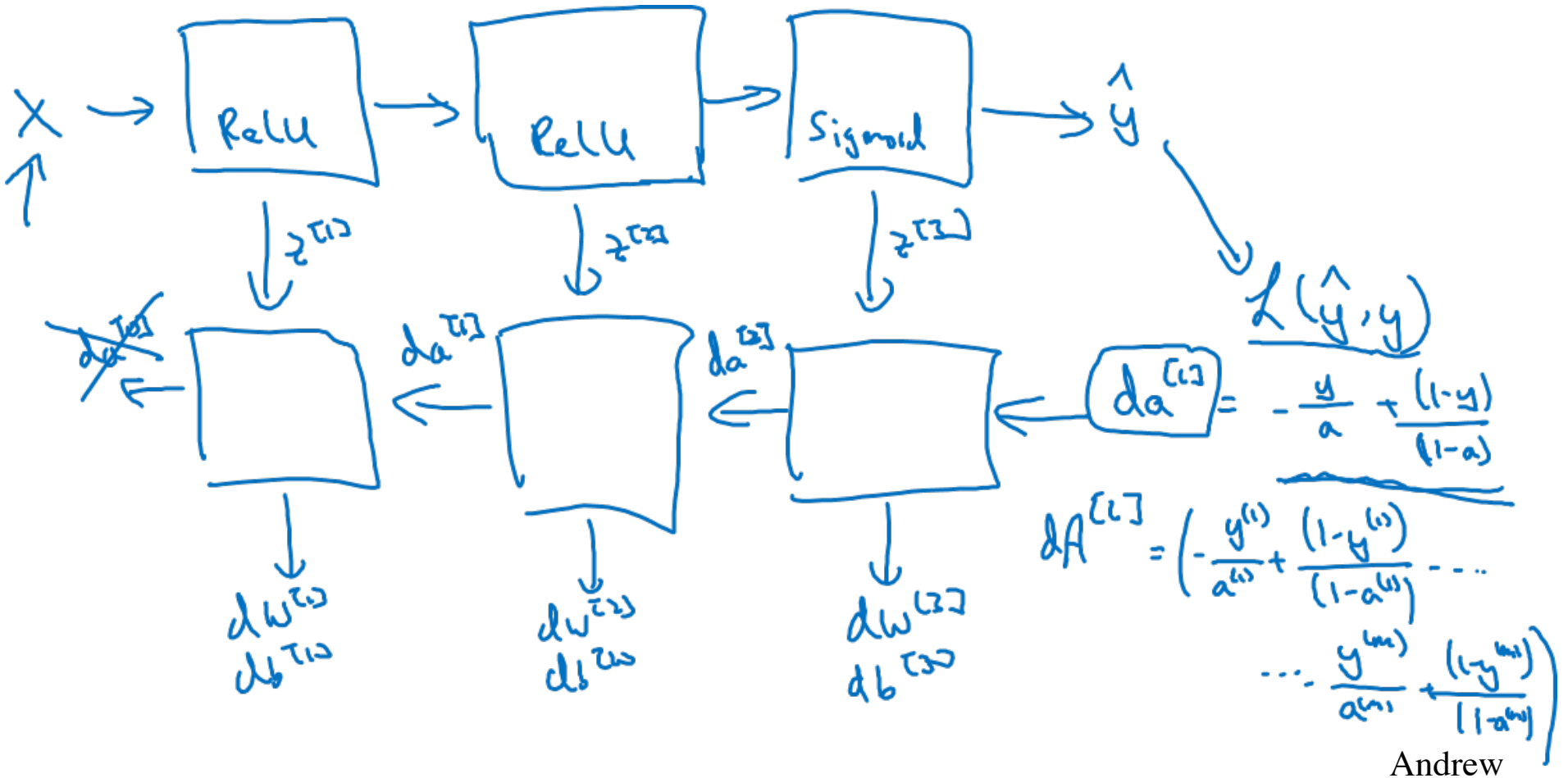
Forward propagation for layer l



Backward propagation for layer l



Summary


X
ReLU
ReLU
Sigmoid
$\hat{y}$
$z^{[1]}$
$z^{[2]}$
$z^{[3]}$
$da^{[0]}$
$da^{[1]}$
$da^{[2]}$
$da^{[3]}$
$\mathcal{L}(\hat{y}, y)$
$dA^{[1]} = \left(-\frac{y^{(1)}}{a^{(1)}} + \frac{(1-y^{(1)})}{(1-a^{(1)})} \cdots \right.$
$\left. \cdots - \frac{y^{(m)}}{a^{(m)}} + \frac{(1-y^{(m)})}{(1-a^{(m)})} \right)$
$dW^{[1]}$ $db^{[1]}$
$dW^{[2]}$ $db^{[2]}$
$dW^{[3]}$ $db^{[3]}$

deeplearning.ai

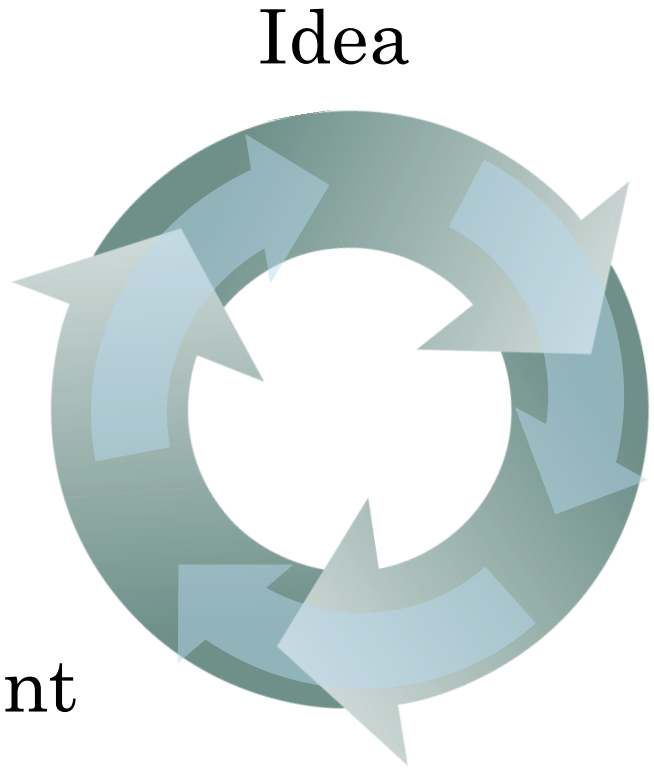
Deep Neural Networks

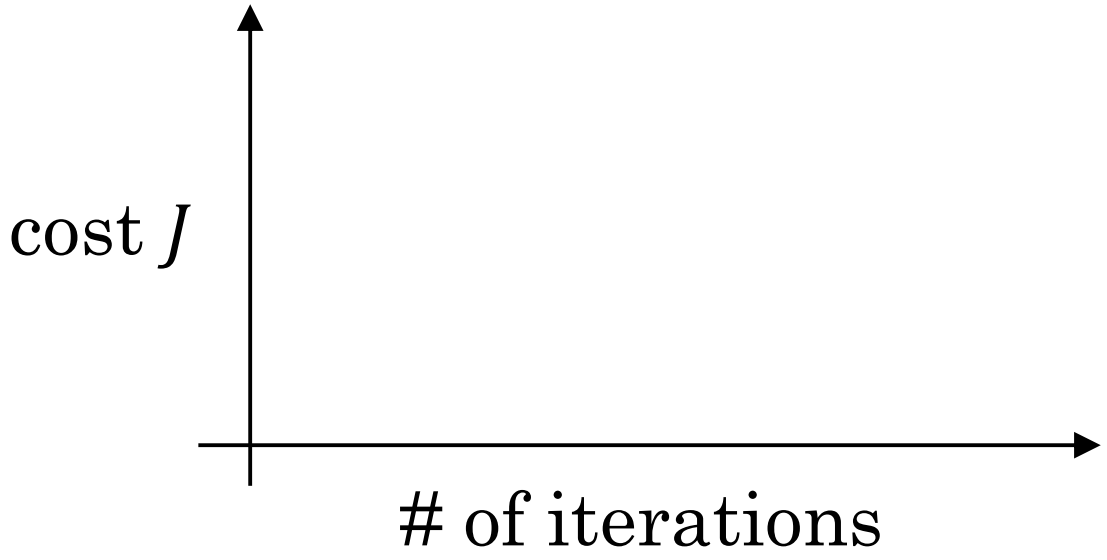
Parameters vs Hyperparameters

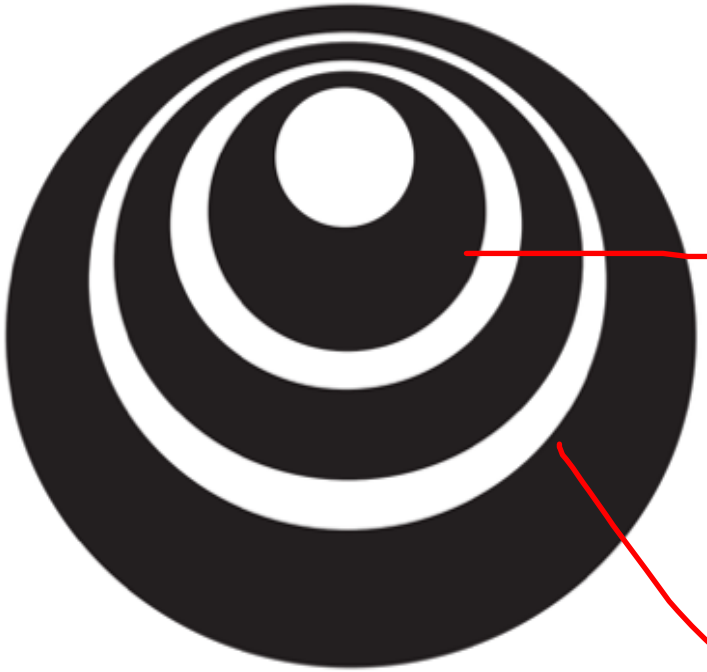
What are hyperparameters?

Parameters: $W^{[1]}$, $b^{[1]}$, $W^{[2]}$, $b^{[2]}$, $W^{[3]}$, $b^{[3]}$ ...

Applied deep learning is a very empirical process







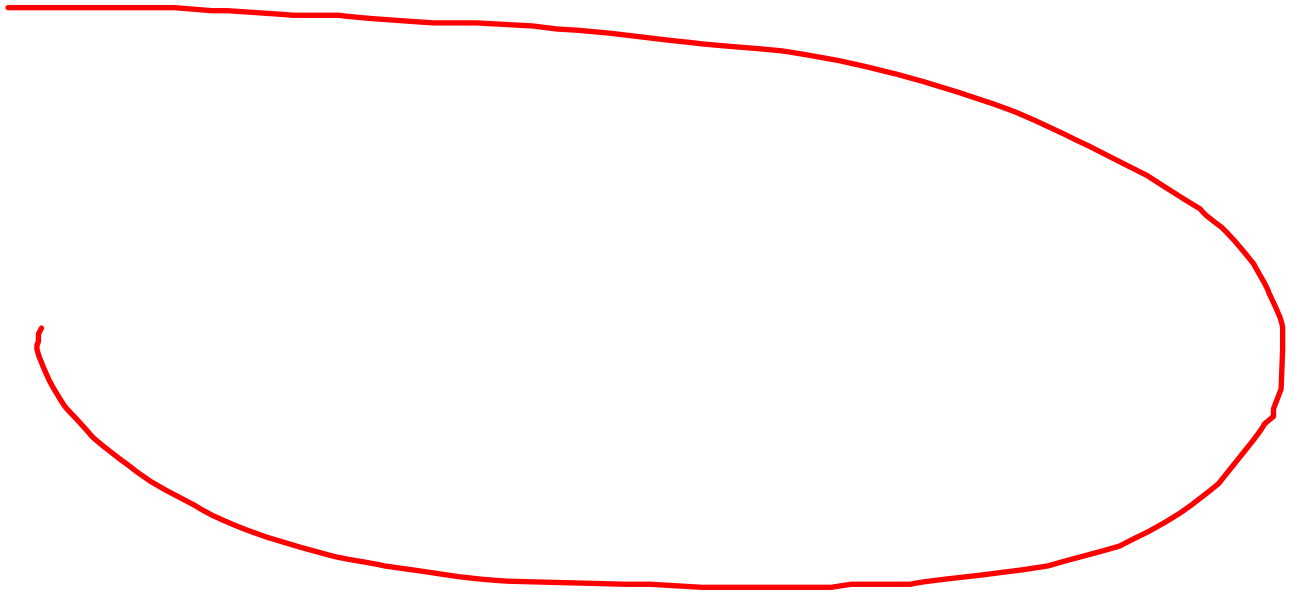
deeplearning.ai

Deep Neural Networks

Parameters vs Hyperparameters

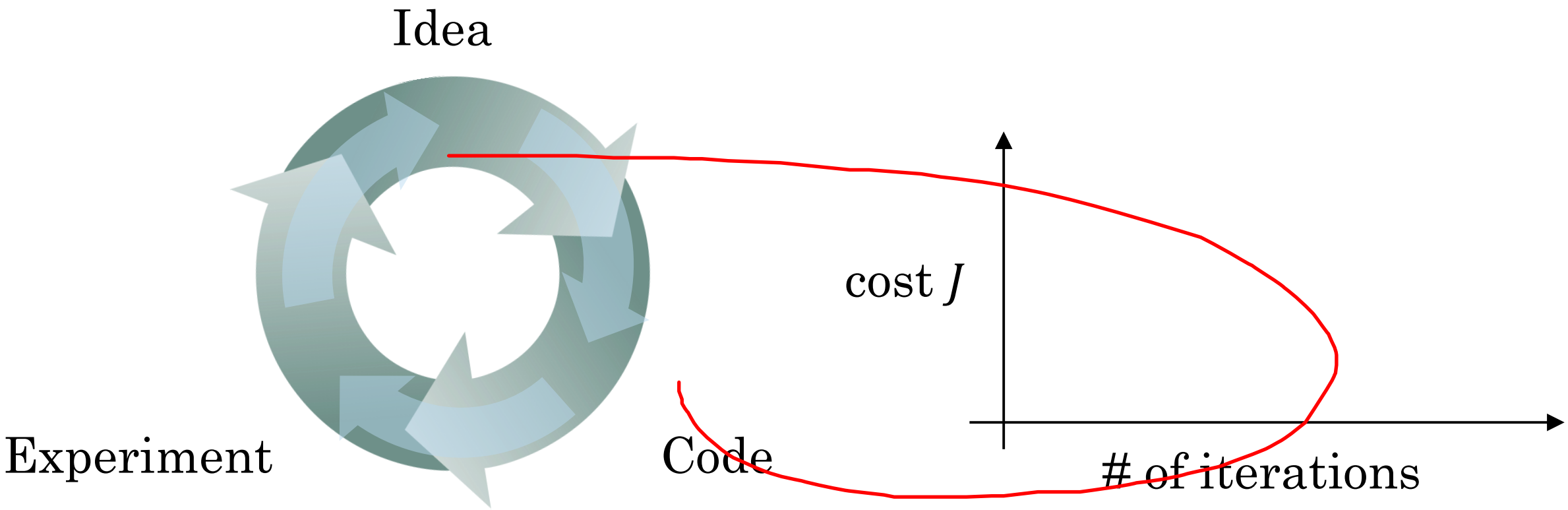
What are hyperparameters?

Parameters: $W^{[1]}$, $b^{[1]}$, $W^{[2]}$, $b^{[2]}$, $W^{[3]}$, $b^{[3]}$ ...



Andrew Ng

Applied deep learning is a very empirical process



deeplearning.ai

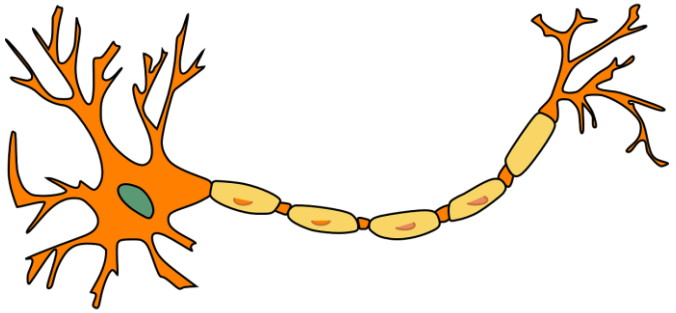
Deep Neural Networks

What does this have to do with the brain?

Forward and backward propagation

$$Z^{[1]} = W^{[1]}X + b^{[1]}$$
$$A^{[1]} = g^{[1]}(Z^{[1]})$$
$$Z^{[2]} = W^{[2]}A^{[1]} + b^{[2]}$$
$$A^{[2]} = g^{[2]}(Z^{[2]})$$
$$\vdots$$
$$A^{[L]} = g^{[L]}(Z^{[L]}) = \hat{Y}$$

$$dZ^{[L]} = A^{[L]} - Y$$
$$dW^{[L]} = \frac{1}{m} dZ^{[L]} A^{[L]^T}$$
$$db^{[L]} = \frac{1}{m} np.\mathrm{sum}(\mathrm{d}Z^{[L]}, axis = 1, keepdims = True)$$
$$dZ^{[L-1]} = dW^{[L]^T} dZ^{[L]} g'^{[L]}(Z^{[L-1]})$$
$$\vdots$$
$$dZ^{[1]} = dW^{[L]^T} dZ^{[2]} g'^{[1]}(Z^{[1]})$$
$$dW^{[1]} = \frac{1}{m} dZ^{[1]} A^{[1]^T}$$
$$db^{[1]} = \frac{1}{m} np.\mathrm{sum}(\mathrm{d}Z^{[1]}, axis = 1, keepdims = True)$$



Andrew Ng