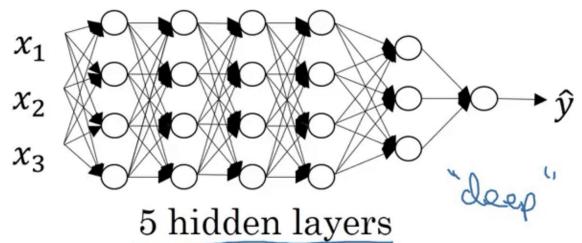
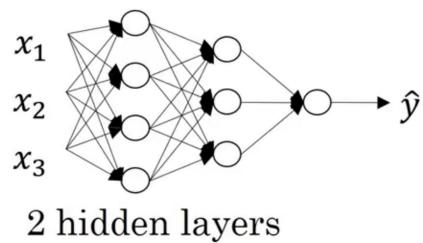
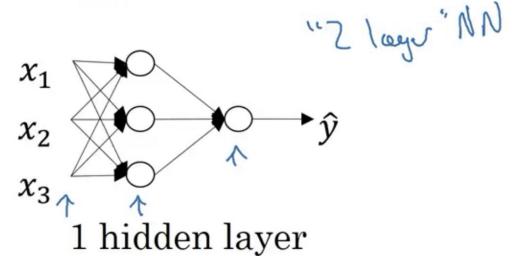
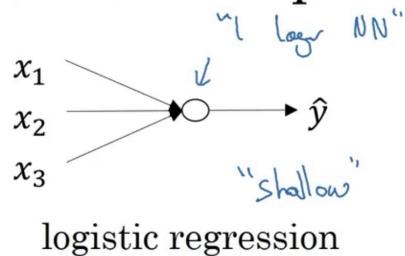


# Week 4 - Deep Neural Networks

Wednesday, August 12, 2020 1:02 PM

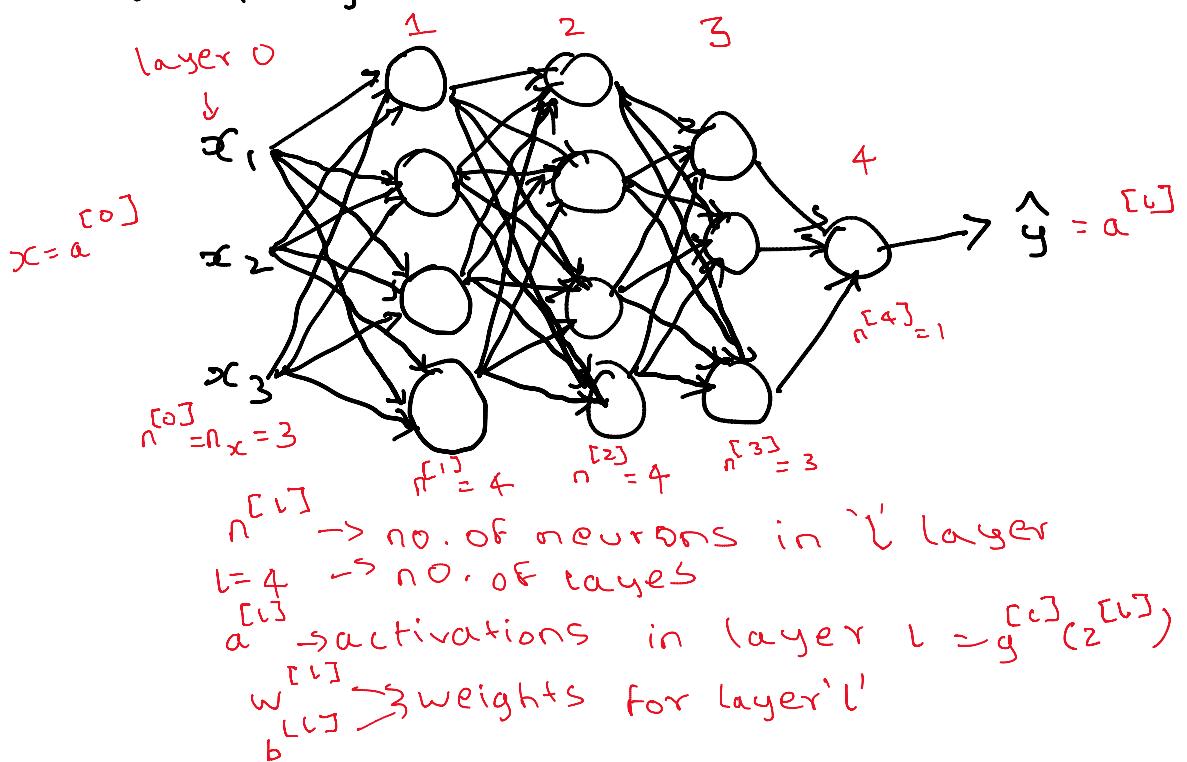
## ① Deep L-layer Neural Network

What is a deep neural network?



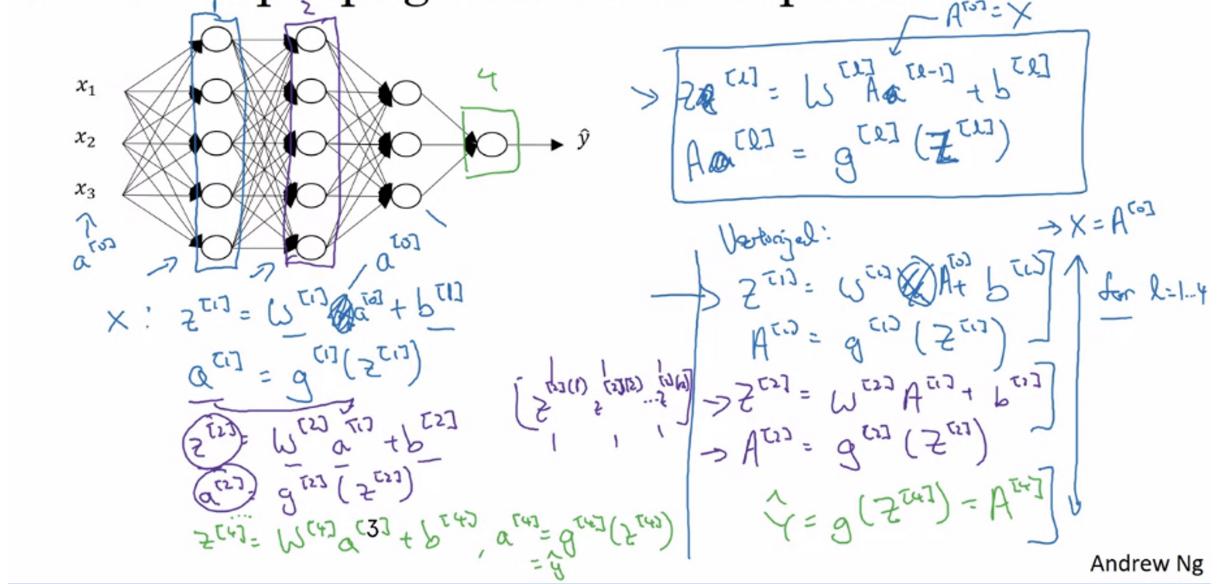
Andrew Ng

e.g.: 4 layer NN



## ② Forward propagation in Deep Neural Network

# Forward propagation in a deep network



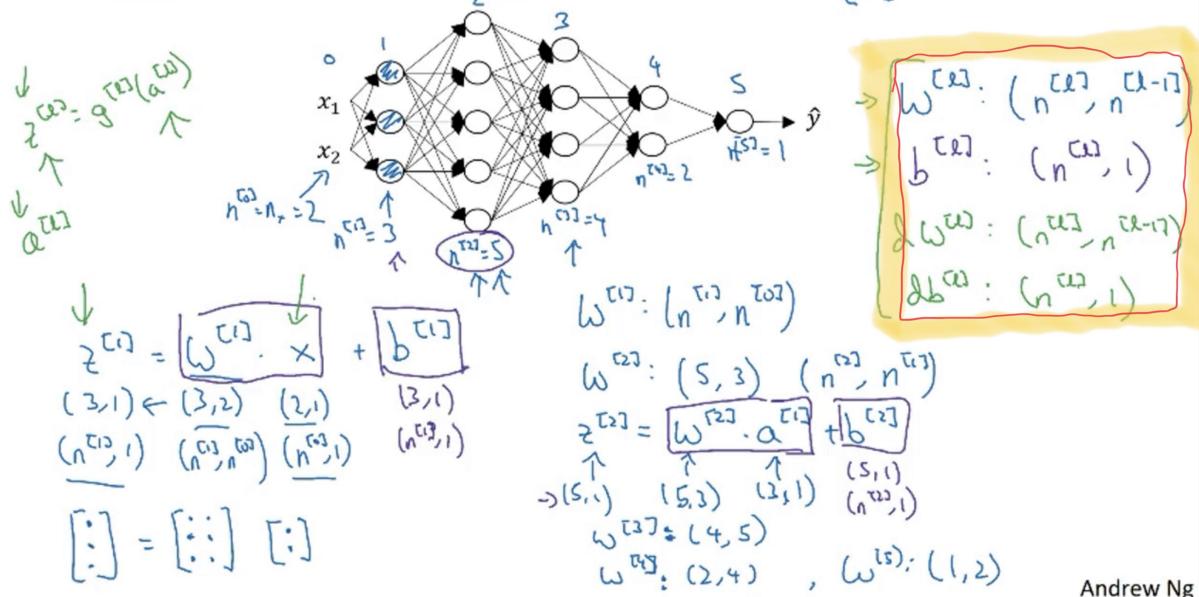
→ General formula

$$z^{[l]} = W^{[l]} A^{[l-1]} + b^{[l]} \quad \left| \begin{array}{l} \text{looped through} \\ 1 \text{ to } l \text{ (no. of layers)} \end{array} \right.$$

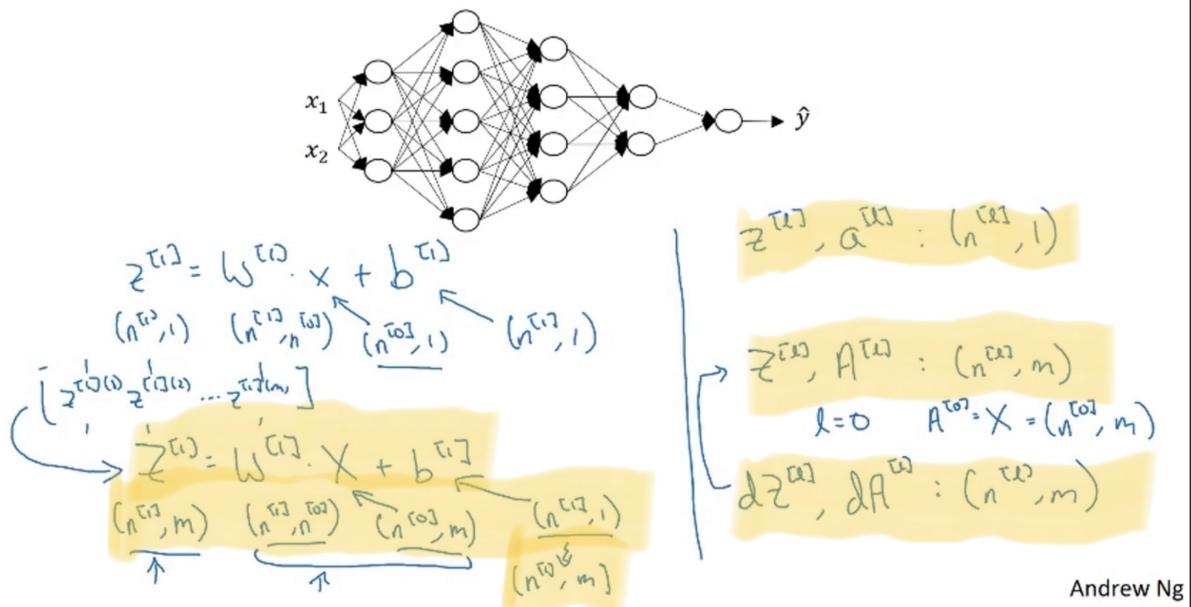
$$A^{[l]} = g^{[l]}(z^{[l]})$$

## ③ Getting the matrix dimensions right

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



## Vectorized implementation

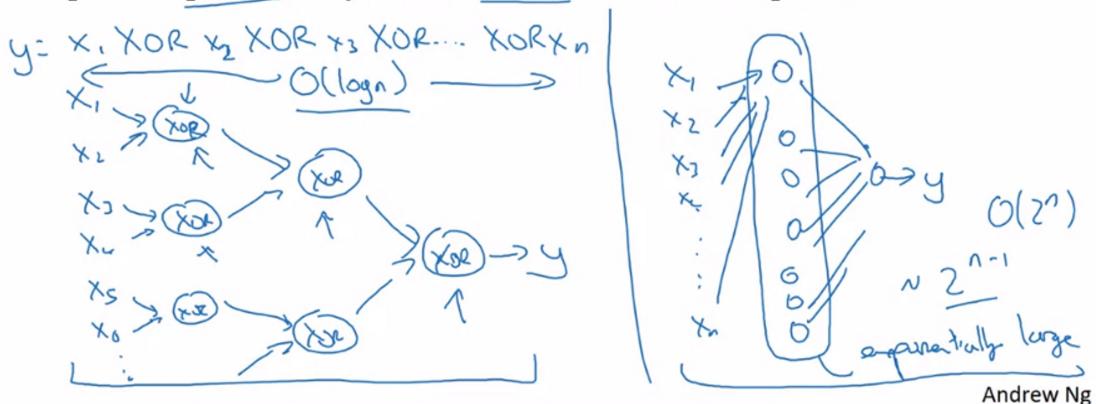


## ④ Why deep representations

- The earlier layers compute simple things
- The deeper layer compute complex objects.

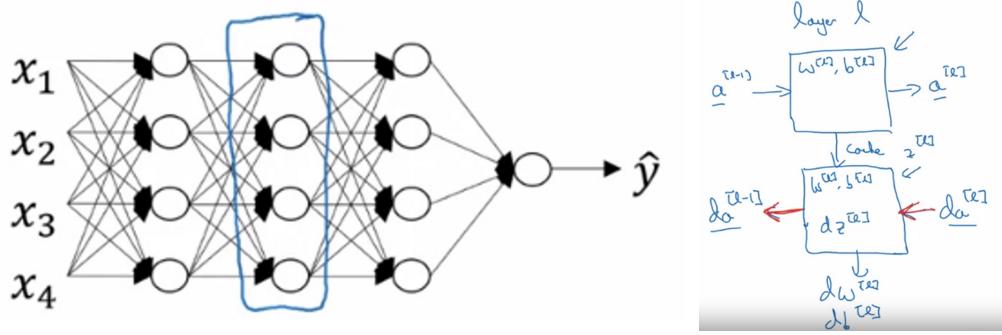
## Circuit theory and deep learning

Informally: There are functions you can compute with a “small” L-layer deep neural network that shallower networks require exponentially more hidden units to compute.



## ⑤ Building blocks of deep neural networks

- forward & backward functions



$\rightarrow$  layer  $l$ :  $w^{[l]}, b^{[l]}$

$\rightarrow$  Forward function : Input  $\rightarrow a^{[l-1]}$

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

$a^{[l]} = g^{[l]}(z^{[l]})$

cache  $z^{[l]}$  for backprop

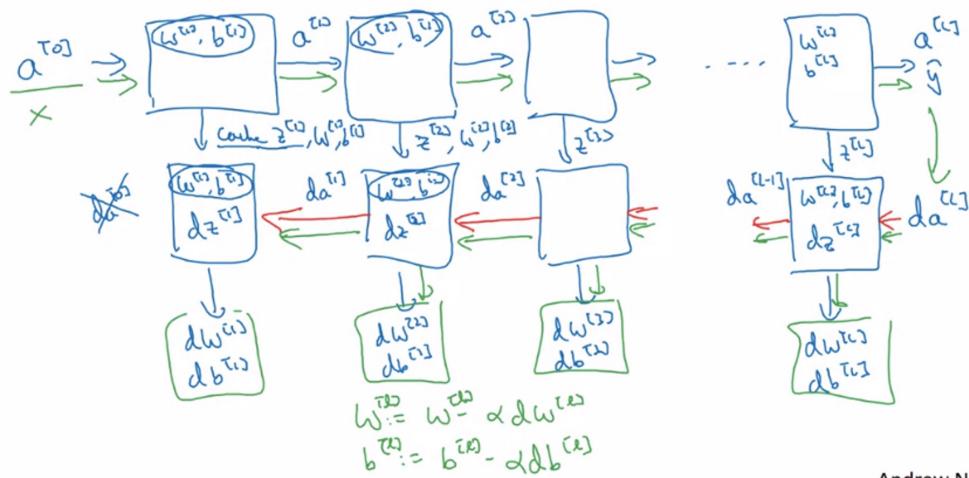
$\rightarrow$  Backward Function : Input  $\rightarrow da^{[l]}$

cache  $z^{[l]}$

Output  $\rightarrow da^{[l-1]}, dw^{[l]}, db^{[l]}$

1 layer  
1 iter

## Forward and backward functions



Andrew Ng

## ⑥ Forward and Backward propagation

- Forward propagation for layer ' $l$ '

i. Input  $A^{[l-1]}$

ii. Output  $A^{[l]}$ , cache ( $z^{[l]}$ )

$$z^{[l]} = w^{[l]}A^{[l-1]} + b^{[l]}$$

$$A^{[l]} = g^{[l]}(z^{[l]})$$

- Backward propagation for layer ' $l$ '

i. Input  $da^{[l]}$

.. . . . . [l-1] . [l] . [l]

i. Input  $\delta A^{[l]}$

ii. Output  $\delta a^{[l-1]}, \delta W^{[l]}, \delta b^{[l]}$

$$\delta Z^{[l]} = \delta A^{[l]} * g^{[l]}'(z^{[l]})$$

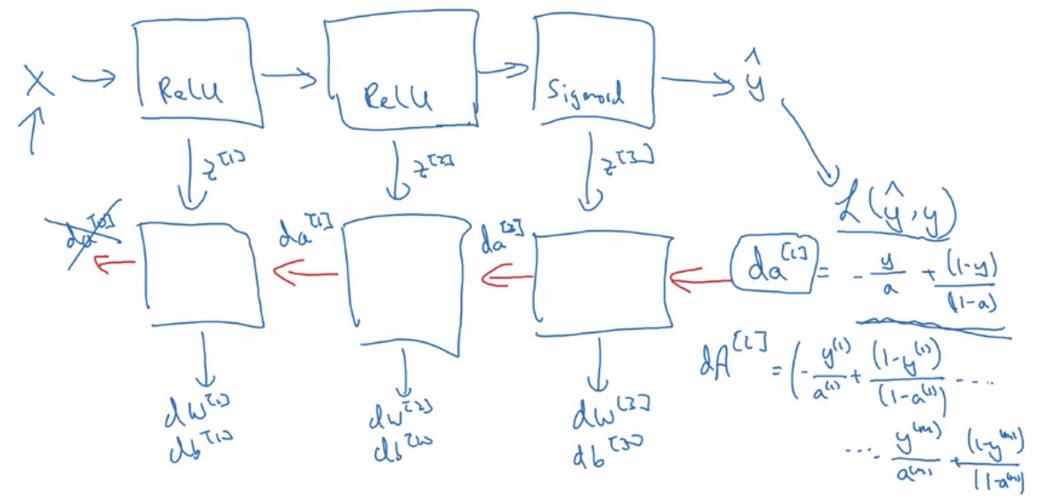
$$\delta W^{[l]} = \gamma_m \delta Z^{[l]} \cdot A^{[l-1]T}$$

$$\delta b^{[l]} = (\gamma_m) \text{np.sum}(\delta Z^{[l]}), \text{axis}=1, \text{keepdims=True})$$

$$\boxed{\delta A^{[l-1]} = W^{[l]} - \delta Z^{[l]}}$$

$$\delta Z^{[l+1]} = (W^{[l+1]})^T \delta Z^{[l+1]} * g^{[l+1]}'(z^{[l+1]})$$

## Summary



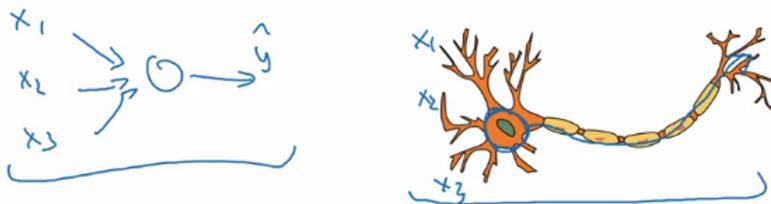
## ⑦ Parameters vs Hyperparameters

- Parameters:  $W^{[l]}, b^{[l]}$
- Hyperparameters:
  - learning rate ( $\alpha$ )
  - no. of iterations (epochs)
  - no. of hidden layers ( $L$ )
  - no. of hidden units ( $n^{[l]}$ )
  - activation fn. choice ( $g^{[l]}$ )
  - others: Momentum, minibatch size, regularization parameters...?
- Deep learning is a very empirical process
  - idea
  - experiment code

experiment  
code

- The values may change over time

## ⑧ Deep learning and Human Brain



- While they may look similar they are fundamentally very different
- Analogy is misused and is an oversimplification