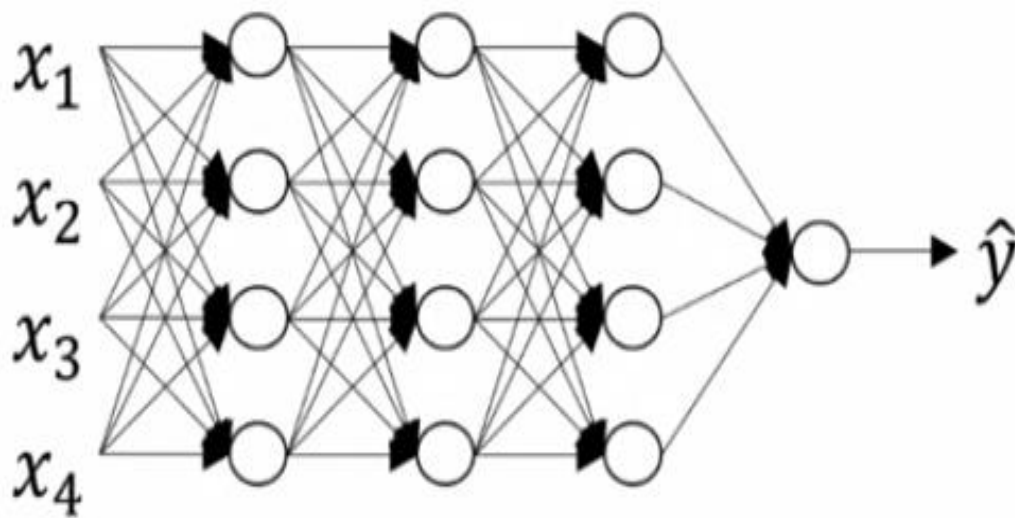


AUTHOR: SYEDA DARAQSHAN

WEEK 4

NEURAL NETWORKS AND DEEP LEARNING

BUILDING BLOCKS OF DEEP NEURAL NETWORK



notes

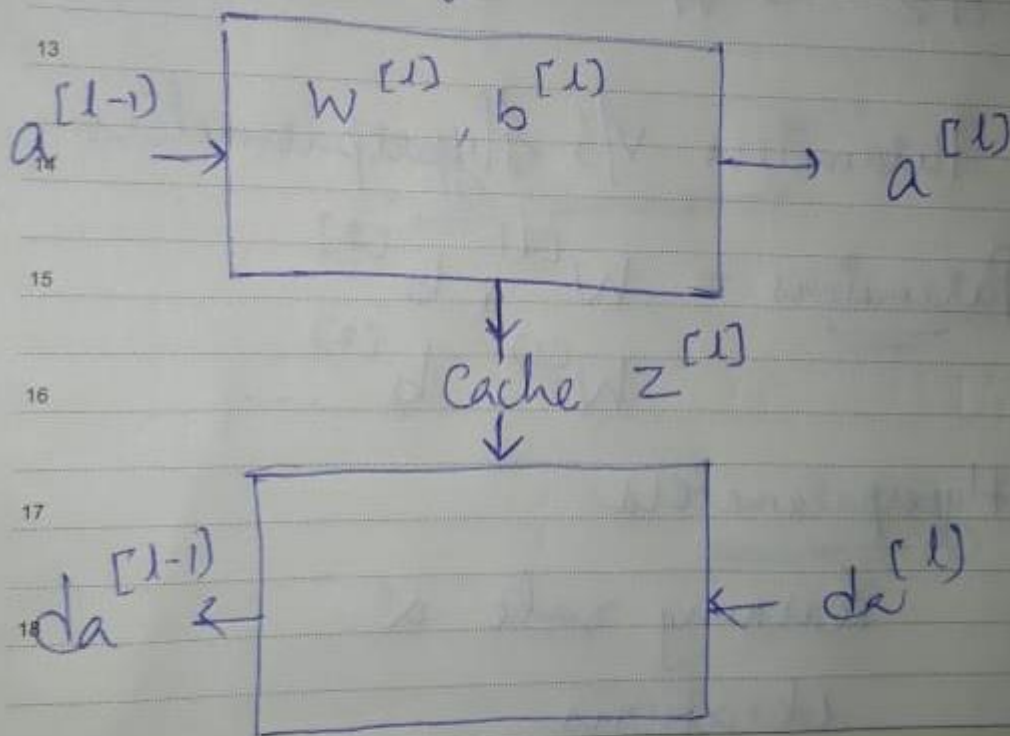
Layer l : $W^{[l]}, b^{[l]}$

Forward prop : i/p $a^{[l-1]}$
o/p $a^{[l]}$
Cache $z^{[l]}$

Backward prop

$\parallel p$ $da^{[l]}$
 $\otimes p$ $da^{[l-1]}$ $Cache\ z^{[l]}$
 $, dw^{[l]}, db^{[l]}$

layer l



Notes

$$dz^{[l]} = da^{[l]} * g'^{[l]}(z^{[l]})$$

$$dZ^{[L]} = A^{[L]} - Y$$

$$dW^{[L]} = \frac{1}{m} dZ^{[L]} A^{[L-1]T}$$

$$db^{[L]} = \frac{1}{m} \text{np.sum}(dZ^{[L]}, \text{axis} = 1, \text{keepdims} = \text{True})$$

$$dZ^{[L-1]} = W^{[L]T} dZ^{[L]} * g'^{[L-1]}(Z^{[L-1]})$$

Note that $*$ denotes element-wise multiplication)

\vdots

$$dZ^{[1]} = W^{[2]} dZ^{[2]} * g'^{[1]}(Z^{[1]})$$

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} A^{[0]T}$$

Note that $A^{[0]T}$ is another way to denote the input features, which is also written as X^T

$$db^{[1]} = \frac{1}{m} \text{np.sum}(dZ^{[1]}, \text{axis} = 1, \text{keepdims} = \text{True})$$

Handwritten mathematical derivations for backpropagation through a neural network layer:

$$dW^{[l]} = dZ^{[l]} \cdot a^{[l-1]}$$

$$db^{[l]} = dZ^{[l]}$$

$$da^{[l-1]} = W^{[l]T} dZ^{[l]}$$

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

PARAMETERS V/S HYPERPARAMETERS

Hyperparameters control the parameters, hence the name given!

