Deep Neural Networks

deeplearning.ai

Building blocks of deep neural networks

Forward and backward functions
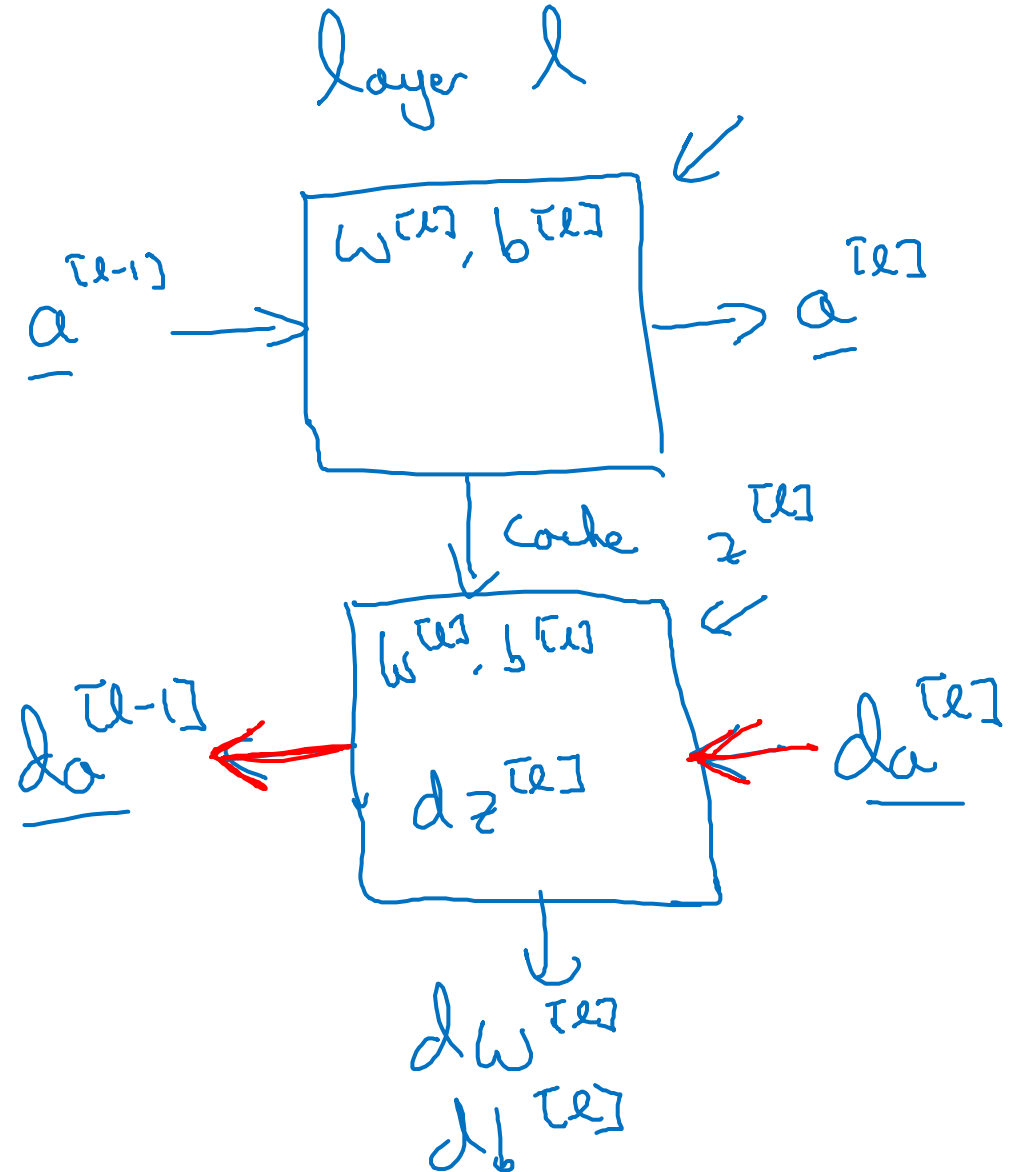

x_1
x_2
x_3
x_4

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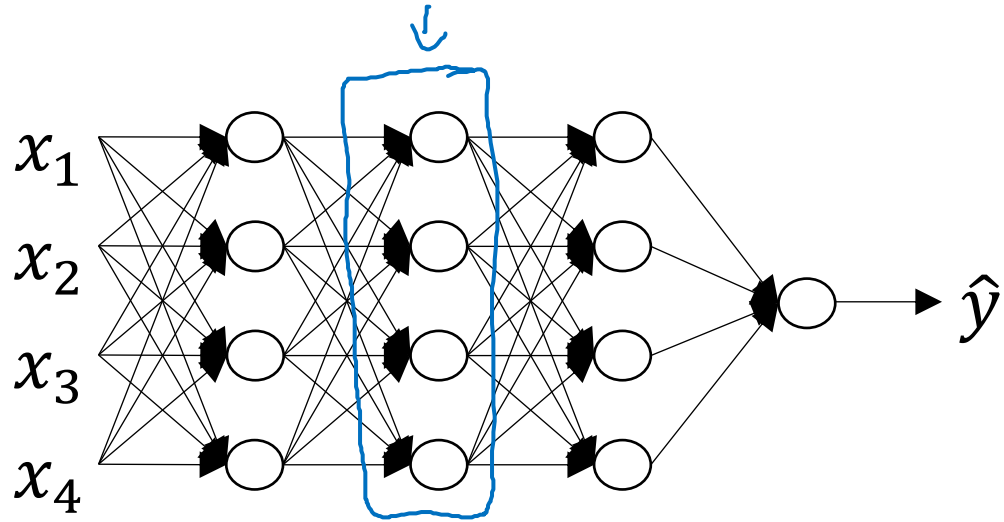
Layer l: $W^{[l]}, b^{[l]}$

$\rightarrow$ Forward: Input $a^{[l-1]}$, output $a^{[l]}$
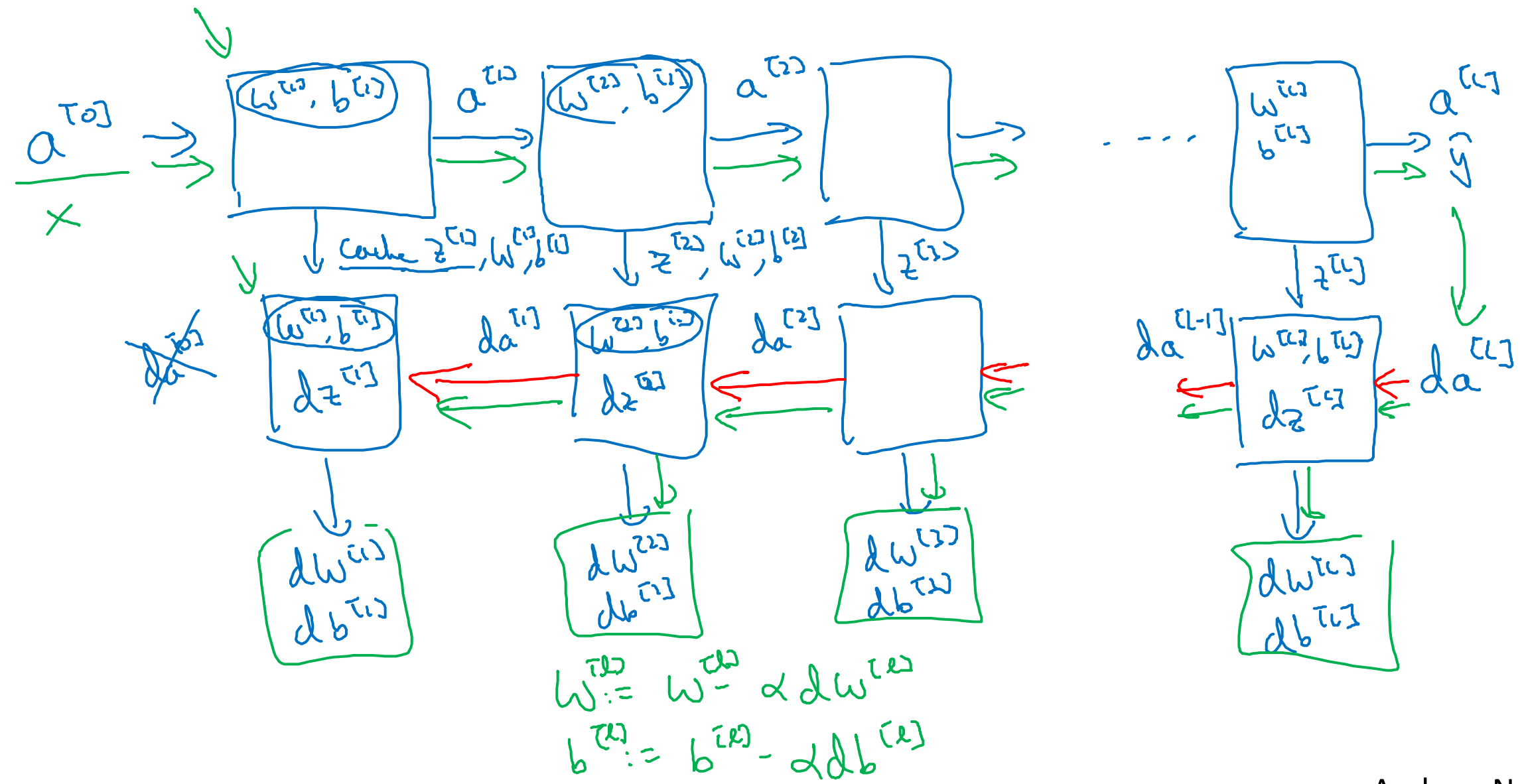
$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$ cache $z^{[l]}$

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

$\rightarrow$ Backward: Input $da^{[l]}$, output $da^{[l-1]}$

cache $(z^{[l]})$

$dW^{[l]}$

$db^{[l]}$

layer l

$a^{[l-1]}$ $W^{[l]}, b^{[l]}$ $a^{[l]}$

cache $z^{[l]}$

$da^{[l-1]}$ $W^{[l]}, b^{[l]}$ $da^{[l]}$

$dz^{[l]}$

$dW^{[l]}$

$db^{[l]}$

Andrew Ng

Forward and backward functions

$$W^{[\ell]} := W^{[\ell]} - \alpha \, dW^{[\ell]}$$
$$b^{[\ell]} := b^{[\ell]} - \alpha \, db^{[\ell]}$$

Andrew Ng

Forward propagation for layer l

$\rightarrow$ Input $a^{[l-1]}$ $\Leftarrow$

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

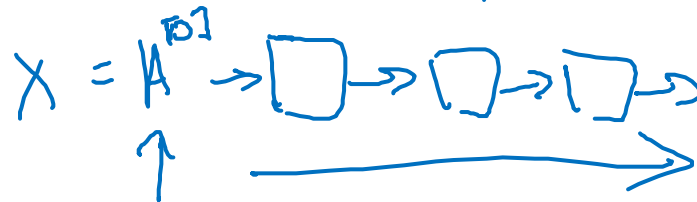$\rightarrow$ Output $a^{[l]}$, cache $(z^{[l]})$

$$z^{[l]} = W^{[l]} \cdot a^{[l-1]} + b^{[l]}$$

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

Vectorized:

$$z^{[l]} = W^{[l]} \cdot A^{[l-1]} + b^{[l]}$$

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

$a^{[0]}$

$A^{[0]}$

$X = A^{[0]} \rightarrow \square \rightarrow \square \rightarrow \square \rightarrow$

Backward propagation for layer l

$\rightarrow$ Input $da^{[l]}$

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

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

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

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

$$\boxed{da^{[l-1]}} = W^{[l]T} \cdot dz^{[l]}$$

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

$$dz^{[l]} = \underbrace{\bigcirc dA^{[l]}} * g^{[l]'}(z^{[l]})$$

$$dW^{[l]} = \frac{1}{m} dz^{[l]} \cdot A^{[l-1]T}$$

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

$$\boxed{dA^{[l-1]}} = W^{[l]T} \cdot dz^{[l]}$$

Summary

$X \rightarrow$ ReLU $\rightarrow$ ReLU $\rightarrow$ Sigmoid $\rightarrow \hat{y} \rightarrow \mathcal{L}(\hat{y}, y)$

$z^{[1]}$, $z^{[2]}$, $z^{[3]}$

$da^{[0]}$ (crossed out) $\leftarrow \boxed{} \xleftarrow{da^{[1]}} \boxed{} \xleftarrow{da^{[2]}} \boxed{} \leftarrow \boxed{da^{[L]}} = -\frac{y}{a} + \frac{(1-y)}{(1-a)}$

$dW^{[1]}$
$db^{[1]}$

$dW^{[2]}$
$db^{[2]}$

$dW^{[3]}$
$db^{[3]}$

$$dA^{[L]} = \left(-\frac{y^{(1)}}{a^{(1)}} + \frac{(1-y^{(1)})}{(1-a^{(1)})} \cdots \right.$$

$$\left. \cdots -\frac{y^{(m)}}{a^{(m)}} + \frac{(1-y^{(m)})}{(1-a^{(m)})} \right)$$