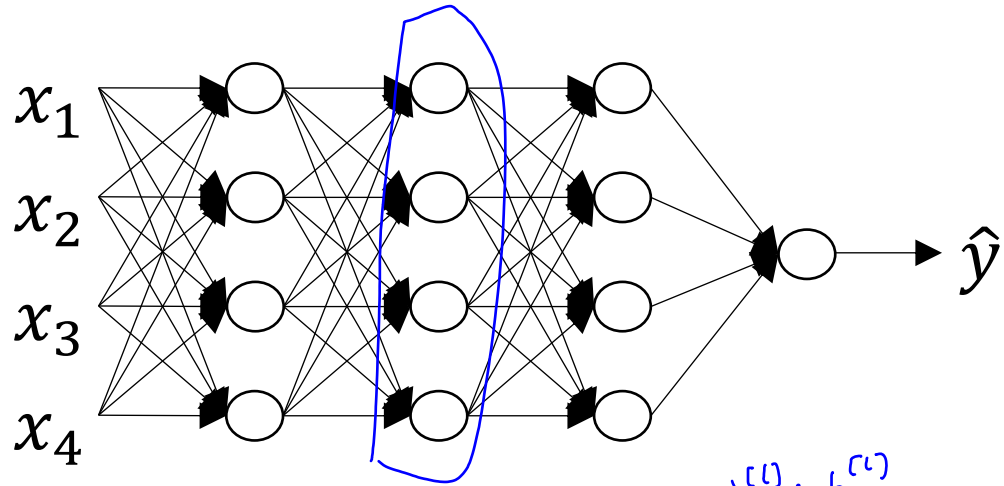# Deep Neural Networks

---

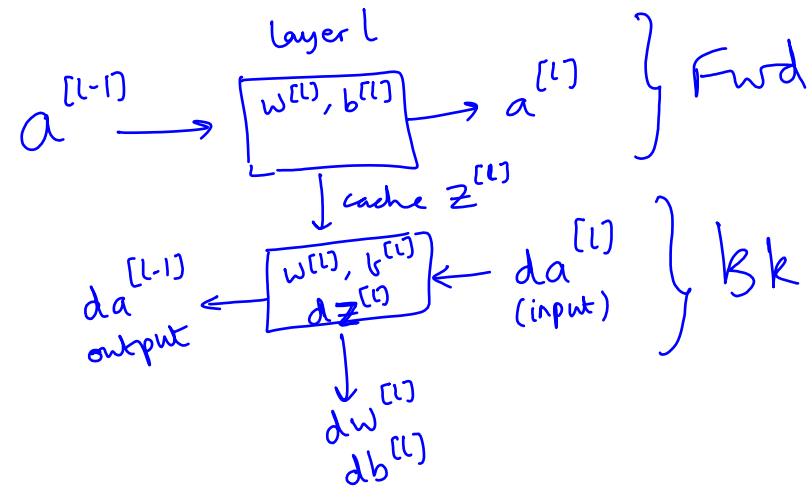## Building blocks of deep neural networks
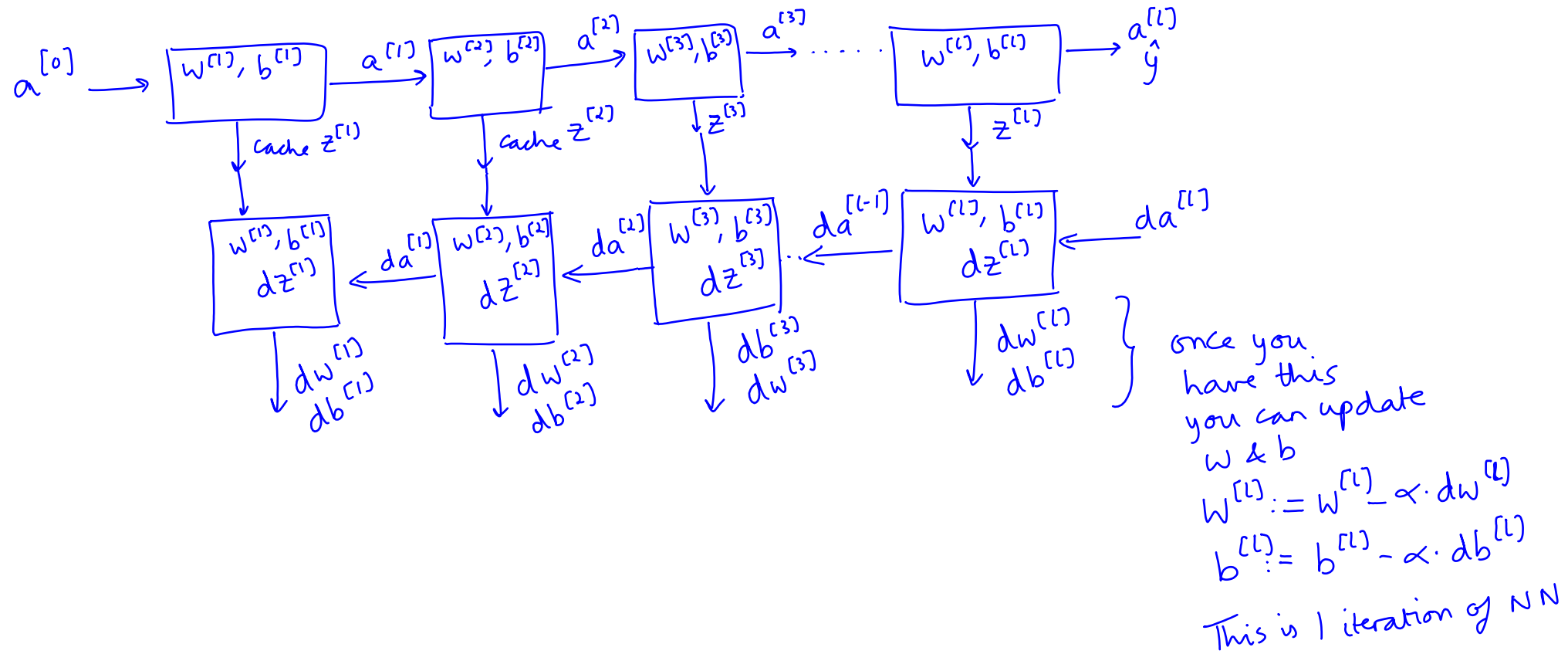
Forward and backward functions

consider a single layer L, with $W^{[l]}$ & $b^{[l]}$
lets write the fwd & bkwd prop for these

Fwd prop
Input $a^{[l-1]}$, output $a^{[l]}$ and cache $Z^{[l]}$
$$Z^{[l]} = W^{[l]} \cdot A^{[l-1]} + \mathbf{b}^{[l]}$$
$$A^{[l]} = g^{[l]}(Z^{[l]})$$

Bkwd prop
Input $da^{[l]}$, output $da^{[l-1]}$, $dw^{[l]}$, $db^{[l]}$
& cache $Z^{[l]}$

Forward and backward functions

$a^{[0]} \rightarrow \boxed{W^{[1]}, b^{[1]}} \xrightarrow{a^{[1]}} \boxed{W^{[2]}, b^{[2]}} \xrightarrow{a^{[2]}} \boxed{W^{[3]}, b^{[3]}} \xrightarrow{a^{[3]}} \cdots \cdots \boxed{W^{[l]}, b^{[l]}} \rightarrow \begin{matrix} a^{[l]} \\ \hat{y} \end{matrix}$

cache $z^{[1]}$ cache $z^{[2]}$ $z^{[3]}$ $z^{[l]}$

$\boxed{\begin{matrix} W^{[1]}, b^{[1]} \\ dz^{[1]} \end{matrix}} \xleftarrow{da^{[1]}} \boxed{\begin{matrix} W^{[2]}, b^{[2]} \\ dz^{[2]} \end{matrix}} \xleftarrow{da^{[2]}} \boxed{\begin{matrix} W^{[3]}, b^{[3]} \\ dz^{[3]} \end{matrix}} \xleftarrow{da^{[l-1]}} \cdots \boxed{\begin{matrix} W^{[l]}, b^{[l]} \\ dz^{[l]} \end{matrix}} \xleftarrow{da^{[l]}}$

$\downarrow \begin{matrix} dW^{[1]} \\ db^{[1]} \end{matrix}$ $\downarrow \begin{matrix} dW^{[2]} \\ db^{[2]} \end{matrix}$ $\downarrow \begin{matrix} db^{[3]} \\ dW^{[3]} \end{matrix}$ $\downarrow \begin{matrix} dW^{[l]} \\ db^{[l]} \end{matrix} \Big\}$

once you have this you can update W & b

$W^{[l]} := W^{[l]} - \alpha \cdot dW^{[l]}$

$b^{[l]} := b^{[l]} - \alpha \cdot db^{[l]}$

This is 1 iteration of NN

Andrew Ng

Deep Neural Networks

Forward and backward propagation

deeplearning.ai

Forward propagation for layer l

Input $a^{[l-1]}$

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

Can cache
$W^{[l]}$ & $b^{[l]}$
as well

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

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

Vectorized

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

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

Backward propagation for layer l

Input $da^{[l]}$

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

element wise

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

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

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

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

Need $W^{[l]}, b^{[l]}, z^{[l]}$ cached

vectorized

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

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

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

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

Andrew Ng

Summary

PTO

Andrew Ng