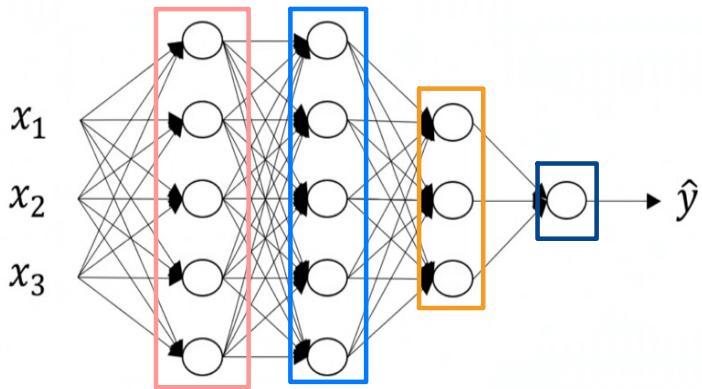


Forward propagation in a deep network



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

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

$$Z^{[1]} = W^{[1]} a^{[0]} + b^{[1]}$$

$$a^{[1]} = g^{[1]}(Z^{[1]})$$

$$Z^{[2]} = W^{[2]} a^{[1]} + b^{[2]}$$

$$a^{[2]} = g^{[2]}(Z^{[2]})$$

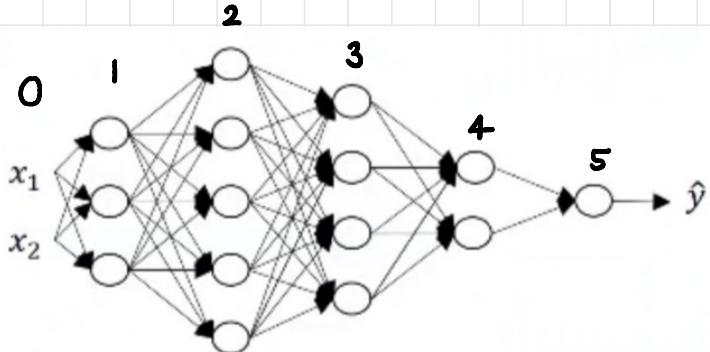
$$Z^{[3]} = W^{[3]} a^{[2]} + b^{[3]}$$

$$a^{[3]} = g^{[3]}(Z^{[3]})$$

$$Z^{[4]} = W^{[4]} a^{[3]} + b^{[4]}$$

$$a^{[4]} = g^{[4]}(Z^{[4]}) = \hat{y}$$

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



$$\begin{aligned}
 L &= 5 \\
 n^{[0]} &= n_x = 2 \\
 n^{[1]} &= 3 \\
 n^{[2]} &= 5 \\
 n^{[3]} &= 4 \\
 n^{[4]} &= 2 \\
 n^{[5]} &= 1
 \end{aligned}$$

$$\begin{aligned}
 W^{[0]} &: (3, 2) \rightarrow (n^{[0]}, n^{[1]}) \\
 W^{[1]} &: (5, 3) \rightarrow (n^{[1]}, n^{[2]}) \\
 W^{[2]} &: (4, 5) \rightarrow (n^{[2]}, n^{[3]}) \\
 W^{[3]} &: (2, 4) \rightarrow (n^{[3]}, n^{[4]}) \\
 W^{[4]} &: (1, 2) \rightarrow (n^{[4]}, n^{[5]})
 \end{aligned}$$

1개의 training set

$$\begin{aligned}
 Z^{[1]} &= W^{[1]} \cdot X + b^{[1]} \\
 (3, 1) &\leftarrow (3, 2) (2, 1) (3, 1) \\
 (n^{[1]}, 1) &\leftarrow (n^{[0]}, n^{[1]}) (n^{[1]}, 1) (n^{[2]}, 1)
 \end{aligned}$$

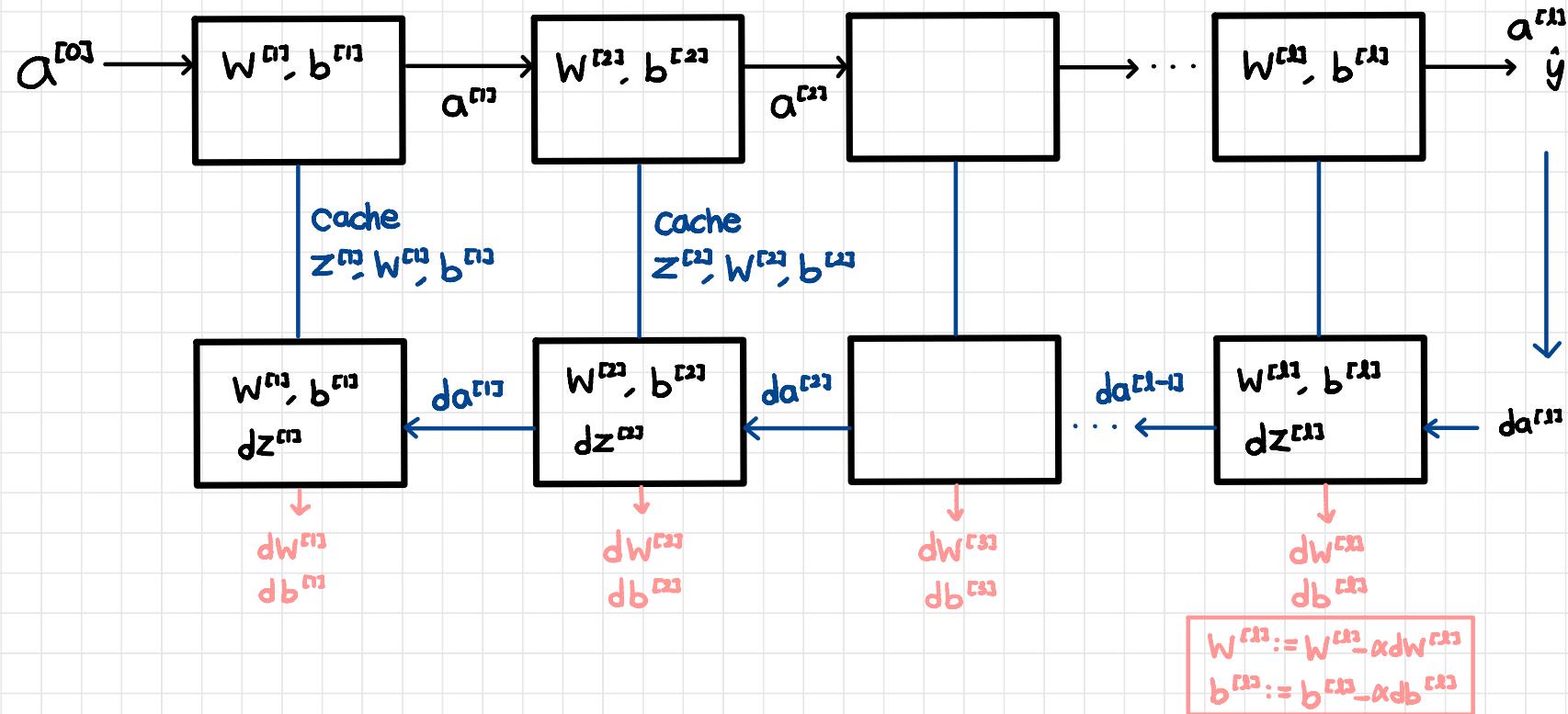
$$[:]=[:, :] [:]$$

m개의 training sets

$$\begin{aligned}
 Z^{[1]} &= W^{[1]} \cdot X + b^{[1]} \\
 (n^{[1]}, m) &\leftarrow (n^{[1]}, n^{[0]}) (n^{[0]}, m) (n^{[0]}, 1)
 \end{aligned}$$

$$\begin{aligned}
 W^{[l]} &: (n^{[l]}, n^{[l-1]}) \\
 b^{[l]} &: (n^{[l]}, 1) \\
 dW^{[l]} &: (n^{[l]}, n^{[l-1]}) \\
 db^{[l]} &: (n^{[l]}, 1) \\
 Z^{[l]}, A^{[l]} &: (n^{[l]}, m) \\
 dz^{[l]}, dA^{[l]} &: (n^{[l]}, m)
 \end{aligned}$$

Forward and backward functions



Forward Propagation / Backward propagation

Forward propagation

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

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

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

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

Backward propagation

Input $da^{[L]}$

Output $dz^{[L]}, dW^{[L]}, db^{[L]}$

$$dz^{[L]} = dA^{[L]} \times g^{[L]'}(z^{[L]})$$

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

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

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

What are hyperparameters ?

Parameters : $W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}, W^{(3)}, b^{(3)}, \dots$

Hyperparameters

- learning rate
- iteration
- layer L 수
- hidden unit 개수
- activation function
- momentum term
- mini batch size



정해진 것이 없으며, 어떤 값으로 설정하느냐에 따라서
모델의 성능에 차이가 생김