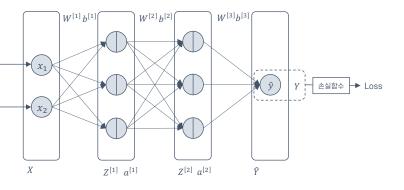
# 역전파Backpropagation 알고리즘 (1):

순전파를 하면서 각층의 기울기를 구하기





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

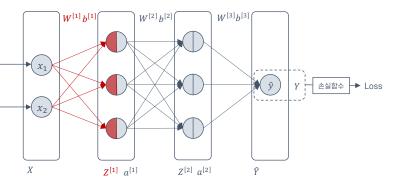
두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

순전파





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$  두번째

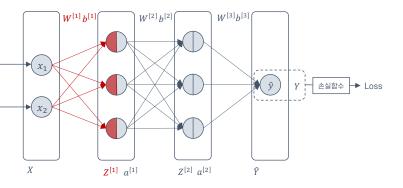
두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$





첫번째 은닉층의 가중치 : **W**<sup>[1]</sup>, b<sup>[1]</sup>

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

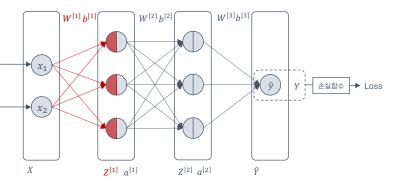
출력층 : Ŷ

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

순전파

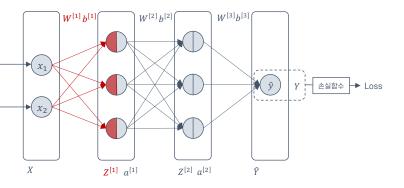
각층의 기울기

 $\frac{dZ^{[1]}}{dW^{[1]}} = X^T$ 

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

순전파

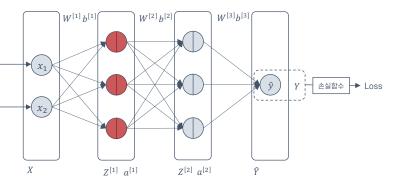
$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>,b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

#### 순전파

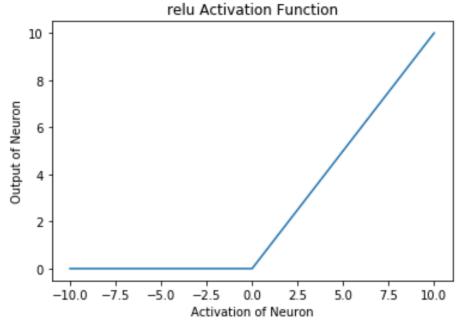
# $Z^{[1]} = XW^{[1]} + b^{[1]}$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

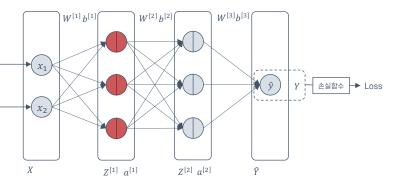
$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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







첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>,b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

#### 순전파

# $Z^{[1]} = XW^{[1]} + b^{[1]}$

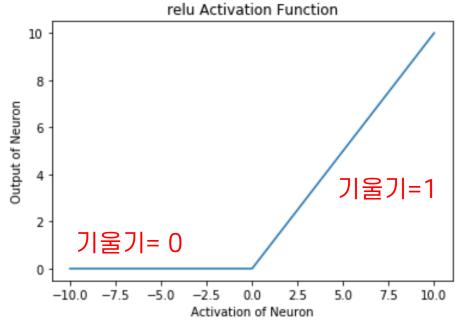
$$a^{[1]} = relu(Z^{[1]})$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

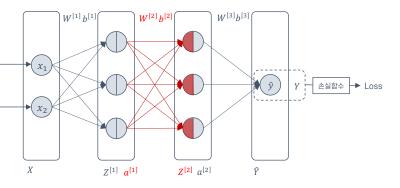
$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$







첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

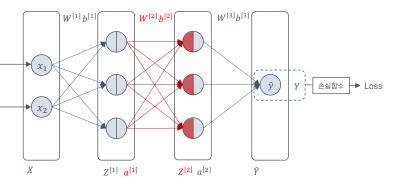
$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 : W<sup>[2]</sup>, b<sup>[2]</sup>

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

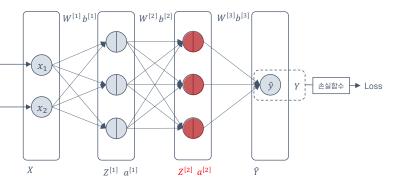
$$a^{[1]} = relu(Z^{[1]})$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>,b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값: **Z<sup>[2]</sup>** 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값: α<sup>[2]</sup>

출력층 :  $\hat{Y}$ 

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

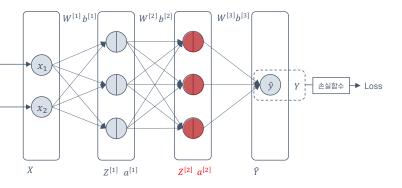
$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

1

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값: **Z<sup>[2]</sup>** 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

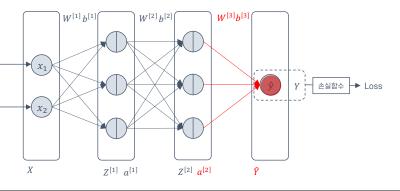
$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

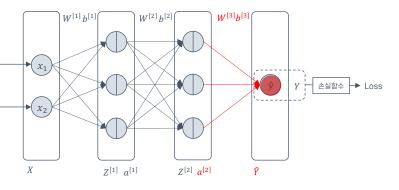
$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$  두번째 은

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$  출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값: <mark>a<sup>[2]</sup></mark>

출력층 :  $\hat{Y}$ 

# 순전파

# $Z^{[1]} = XW^{[1]} + b^{[1]}$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

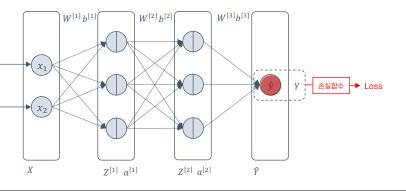
$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW[3]} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}}=1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $\mathbb{Z}^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

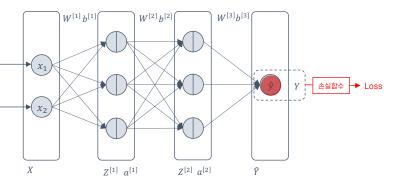
$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

 $\frac{dZ^{[1]}}{db^{[1]}} = 1$ 

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

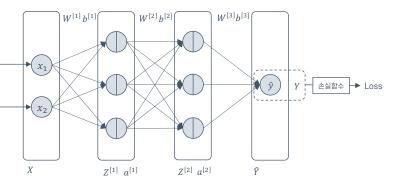
$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# $Z^{[1]} = XW^{[1]} + b^{[1]}$

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

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

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

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

### 각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

# 학습에 필요한 그래디언트

$$\frac{\partial Loss}{\partial W^{[1]}}, \frac{\partial Loss}{\partial W^{[2]}}, \frac{\partial Loss}{\partial W^{[3]}}$$

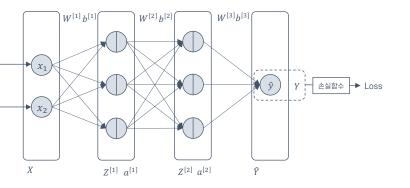
$$\frac{\partial Loss}{\partial b^{[1]}}, \frac{\partial Loss}{\partial b^{[2]}}, \frac{\partial Loss}{\partial b^{[3]}}$$



# 역전파Backpropagation 알고리즘 (2):

각 가중치의 손실에 대한 기울기를 구하기





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

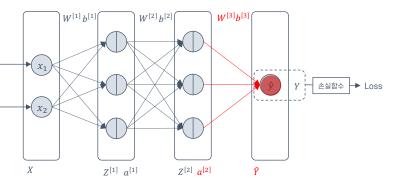
$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\widehat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

#### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}}$$

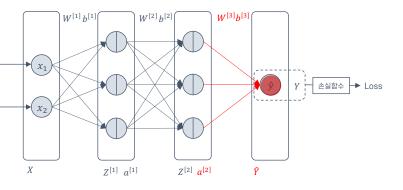
$$\frac{dL}{dh^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

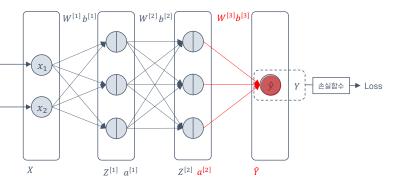
$$\frac{dL}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

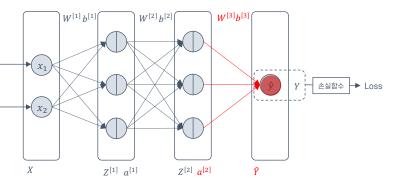
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

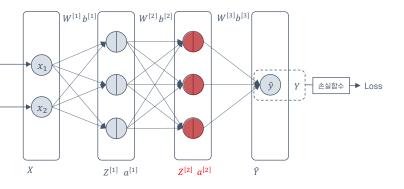
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

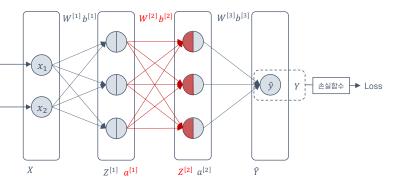
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

### 각층의 기울기

# 역전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}}$$

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

 $\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$ 

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$0 \le 0$$

 $\frac{d\widehat{Y}}{db^{[3]}} = 1$ 

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW[3]} = \frac{dL}{d\hat{V}} \frac{d\hat{Y}}{dW[3]}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

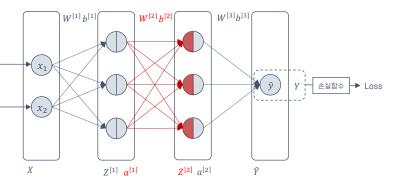
$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

 $\frac{d\widehat{Y}}{dW^{[3]}} = a^{T^{[2]}}$ 





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

### 각층의 기울기

# 역전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$dZ^{[2]}$$
  $T^{[1]}$ 

 $\frac{dZ^{[1]}}{dW^{[1]}} = X^T$ 

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{dh^{[2]}}$$

$$\frac{dL}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{d\hat{Y}}{d\hat{Y}} = W^{T[3]}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

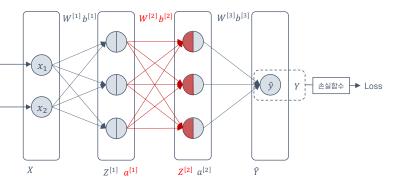
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dL}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

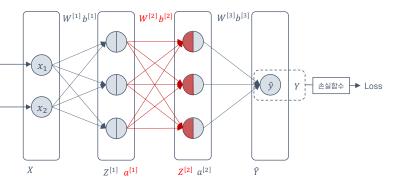
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

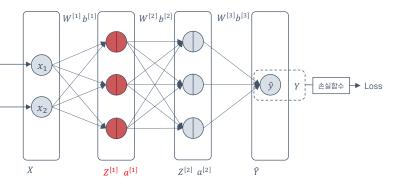
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

#### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dL}{da^{[1]}} =$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\widehat{Y}}{db^{[3]}} = 1$$

$$\frac{d\widehat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

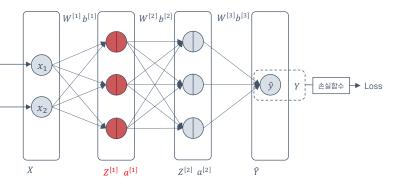
$$\frac{dL}{dh^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dh^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

순전파

# 각층의 기울기

역전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

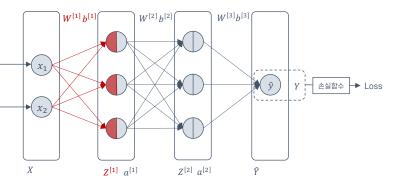
$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

 $dL dZ^{[2]}$ 

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

출력층 :  $\hat{Y}$ 

순전파

#### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{dL}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}}$$

$$\frac{dI}{d\lambda}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

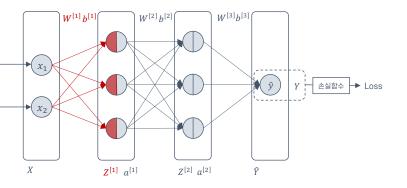
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{7}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}}$$

$$\frac{dL}{dX}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

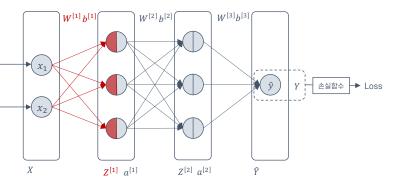
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X$$

$$\frac{dZ^{[1]}}{dh^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{T} \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1 \qquad \frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}} \qquad \frac{dL}{dz^{[1]}}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dZ}{dz^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

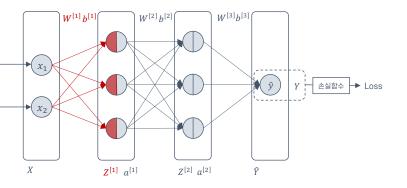
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

# 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dX^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{dh^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{T} \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1 \qquad \frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dZ^{[1]}}{dz^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{dh^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

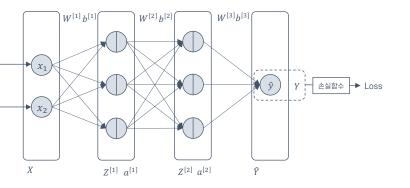
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 :  $\hat{Y}$ 

# 순전파

#### 각층의 기울기

#### 역전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{T} \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1 \qquad \frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}} \qquad \frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

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

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{dh^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{d}{dz^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

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

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\frac{d\widehat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\widehat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

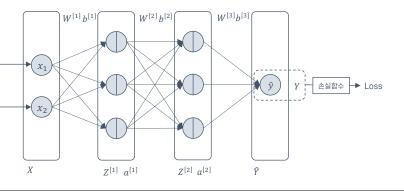
$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

인공신경망의 학습에 필요한 그래디언트 계산



# 역전파 알고리즘 개괄





첫번째 은닉층의 가중치 : W<sup>[1]</sup>, b<sup>[1]</sup>

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

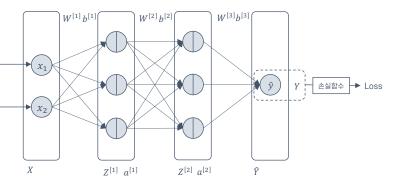
두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

#### 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

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

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

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

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dh^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0\\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

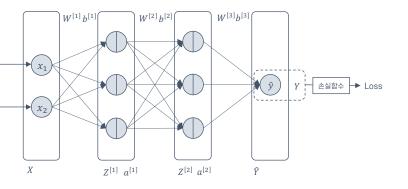
$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$



첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

#### 각층의 기울기

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

 $\frac{dZ}{dZ^{[1]}} = \frac{1}{da^{[1]}} \frac{1}{dZ^{[1]}}$ 

 $\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$ 

 $dL da^{[1]}$ 

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}} \qquad \frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

역전파

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

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

 $\frac{dZ^{[1]}}{dW^{[1]}} = X^T \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1$ 

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

$$\frac{dZ^{[2]}}{db^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

 $\frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$ 

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

 $\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$ 

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{du}{dZ^{[2]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

 $\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$ 

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

 $\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}}$ 

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

 $\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ}{da^{[1]}}$ 

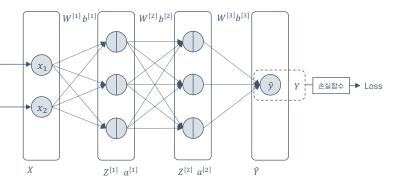
 $dL dZ^{[2]}$ 

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파를 통해 알아야 하는 점 (1) 학습 순서





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 : W<sup>[3]</sup>, b<sup>[3]</sup>

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

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

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

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

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

# 각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{T} \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1 \qquad \frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dZ^{[2]}}{W^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1 \qquad \frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{\partial}{\partial x_{[3]}} = a^{T[2]}$$
  $\frac{d\hat{Y}}{db^{[3]}} = 1$   $\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$ 

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}} \qquad \frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}} \qquad \frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

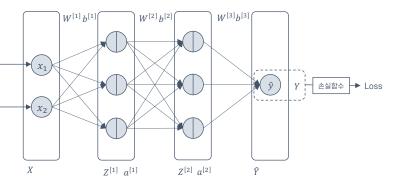
$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}} \qquad \frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}} \qquad \qquad \frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}} \qquad \qquad \frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

역전파를 통해 알아야 하는 점 (2) 메모리





첫번째 은닉층의 가중치 :  $W^{[1]}, b^{[1]}$ 

첫번째 은닉층의 로짓 값:  $Z^{[1]}$ 

첫번째 은닉층의 활성화 값:  $a^{[1]}$ 

두번째 은닉층의 가중치 :  $W^{[2]}, b^{[2]}$ 

출력층의 가중치 :  $W^{[3]}, b^{[3]}$ 

두번째 은닉층의 로짓 값:  $Z^{[2]}$ 

두번째 은닉층의 활성화 값:  $a^{[2]}$ 

출력층 : Ŷ

# 순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

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

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

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

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

# ㆍ 각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^{T} \qquad \frac{dZ^{[1]}}{db^{[1]}} = 1 \qquad \frac{dZ^{[1]}}{dX} = W^{T^{[1]}}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \ge 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T^{[1]}} \qquad \frac{dZ^{[2]}}{db^{[2]}} = 1 \qquad \frac{dZ^{[2]}}{da^{[1]}} = W^{T^{[2]}}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0\\ 1, & z^{[2]} \ge 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T^{[2]}} \qquad \frac{d\hat{Y}}{db^{[3]}} = 1 \qquad \frac{d\hat{Y}}{da^{[2]}} = W^{T^{[3]}}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}} \qquad \frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}} \qquad \frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}} \qquad \frac{dL}{db^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[1]}}{db^{[2]}} \qquad \frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

