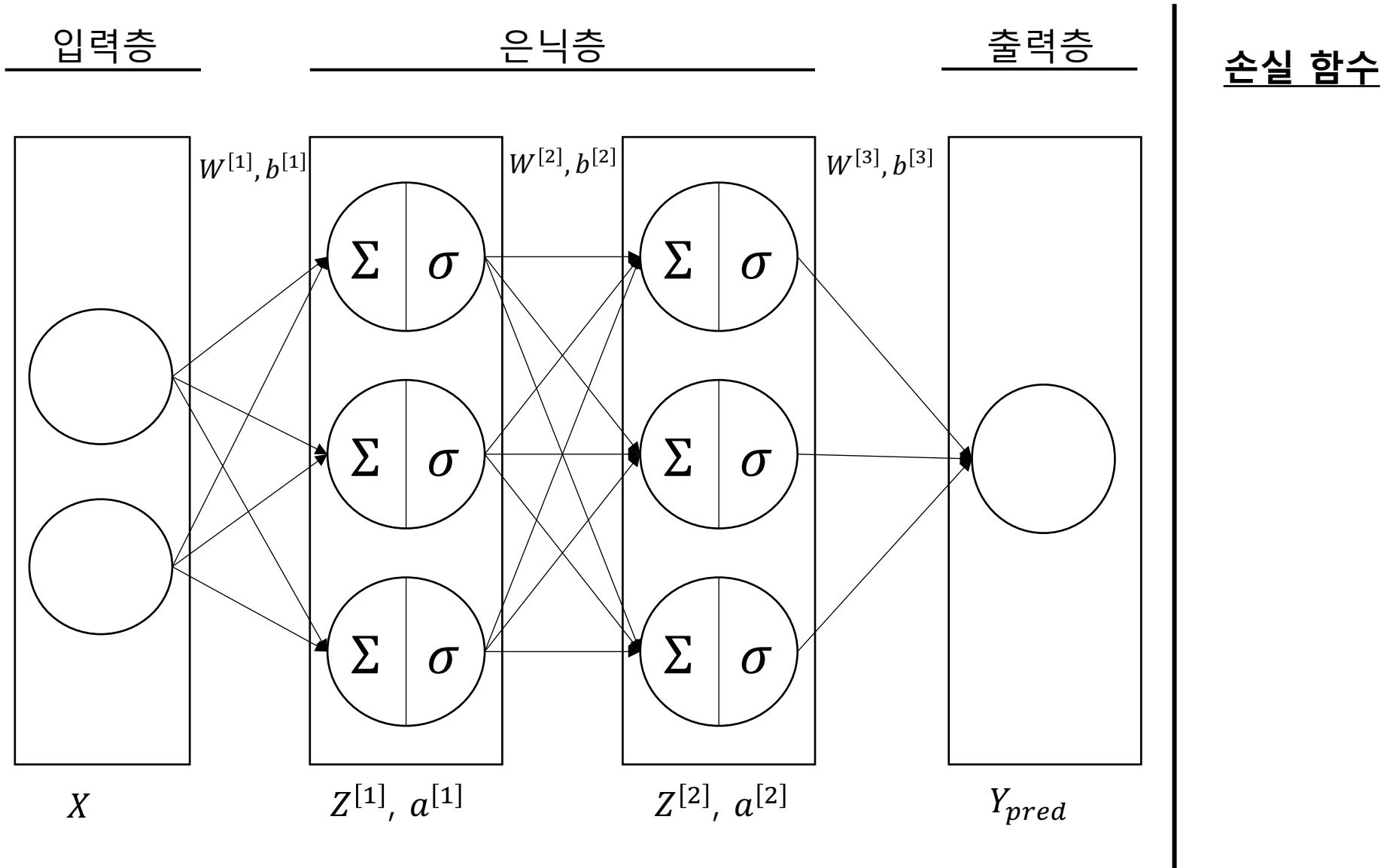
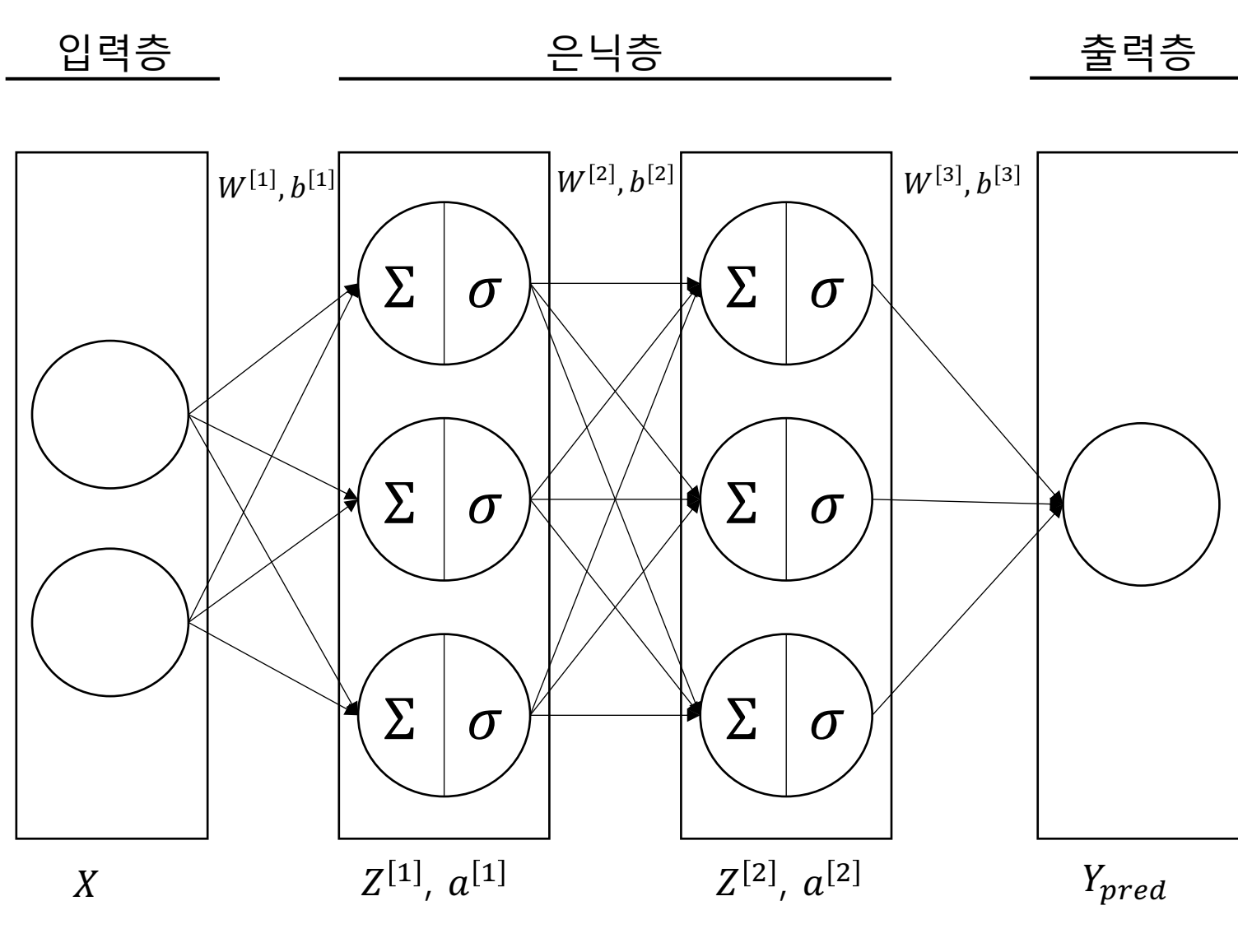


# 역전파 <Back Propagation>

# 네트워크를 학습한다는 것 = 가중치를 학습한다는 것



# 네트워크를 학습한다는 것 = 가중치를 학습한다는 것

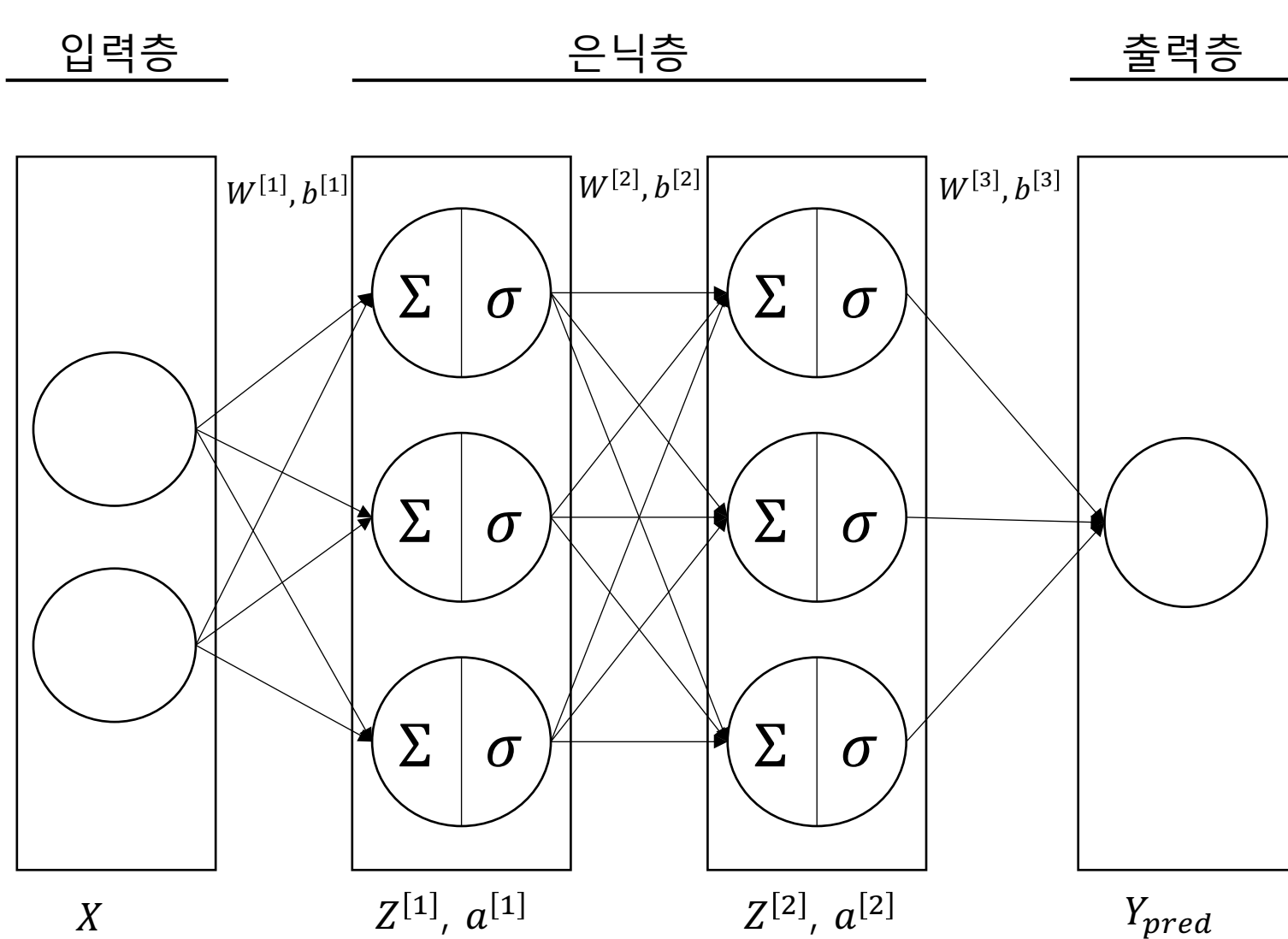


## 손실 함수

정답값과 예측값의 차이

$$Loss = \frac{1}{2} (Y_{pred} - Y_{true})^2$$

# 네트워크를 학습한다는 것 = 가중치를 학습한다는 것



## 손실 함수

정답값과 예측값의 차이

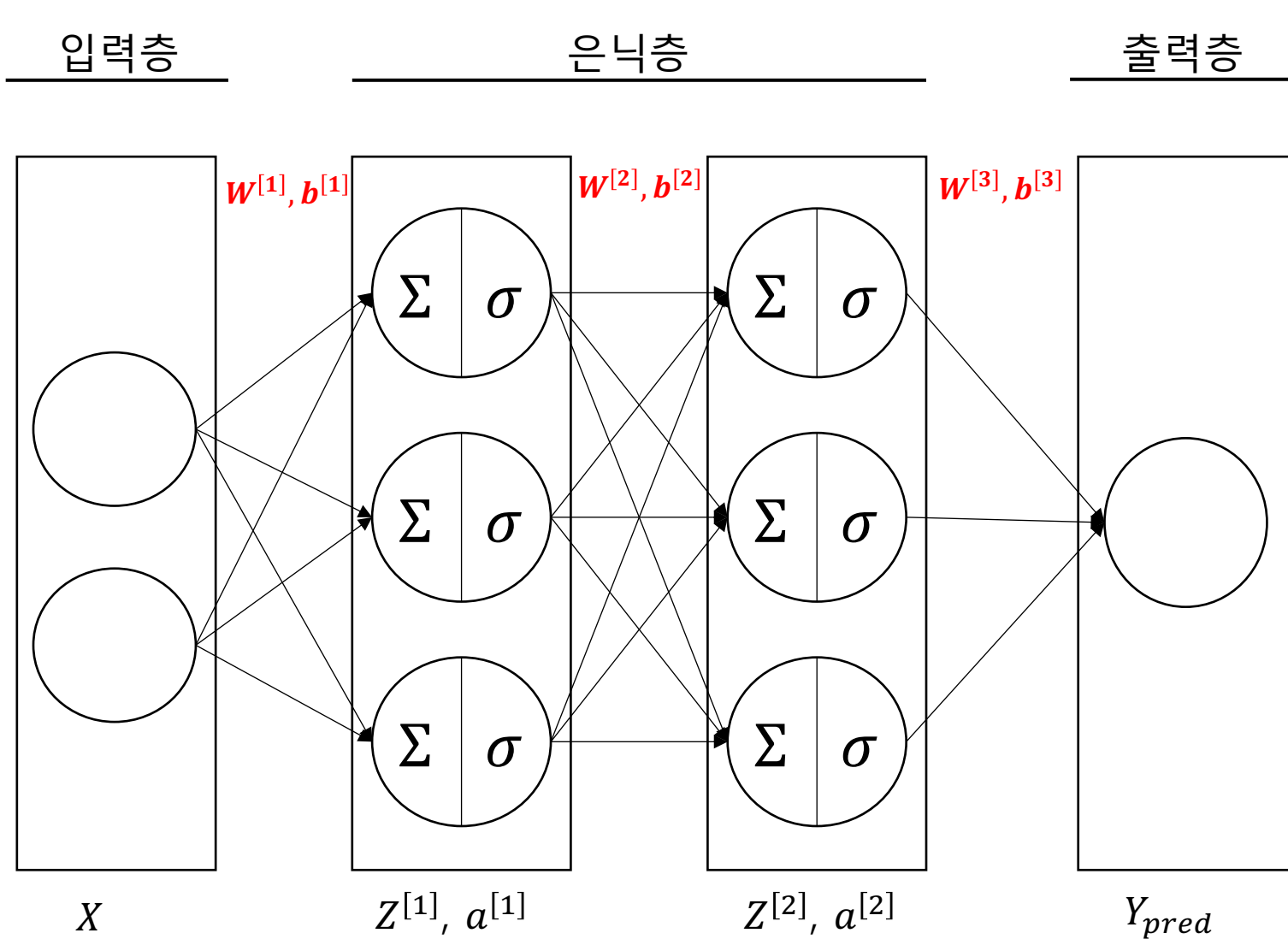
$$Loss = \frac{1}{2} (Y_{pred} - Y_{true})^2$$

## 경사하강법

손실함수를 통해 가중치를 찾는 방법

$$W := W - \alpha \frac{\partial Loss}{\partial W}$$

# 네트워크를 학습한다는 것 = 가중치를 학습한다는 것



## 손실 함수

정답값과 예측값의 차이

$$Loss = \frac{1}{2} (Y_{pred} - Y_{true})^2$$

## 경사하강법

손실함수를 통해 가중치를 찾는 방법

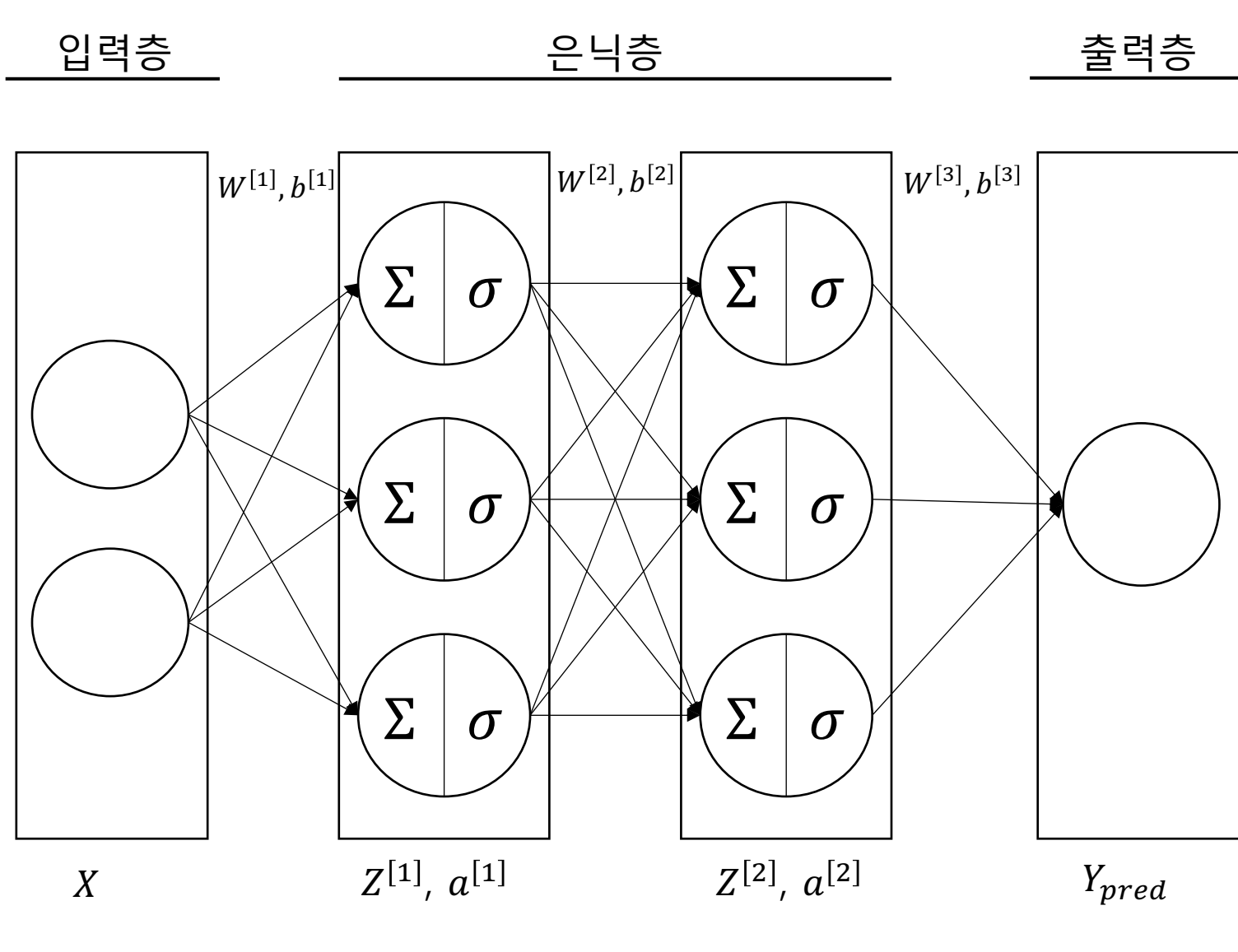
$$W := W - \alpha \frac{\partial Loss}{\partial W}$$

## 우리가 알아야 하는 것들

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

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

# 네트워크를 학습한다는 것 = 가중치를 학습한다는 것



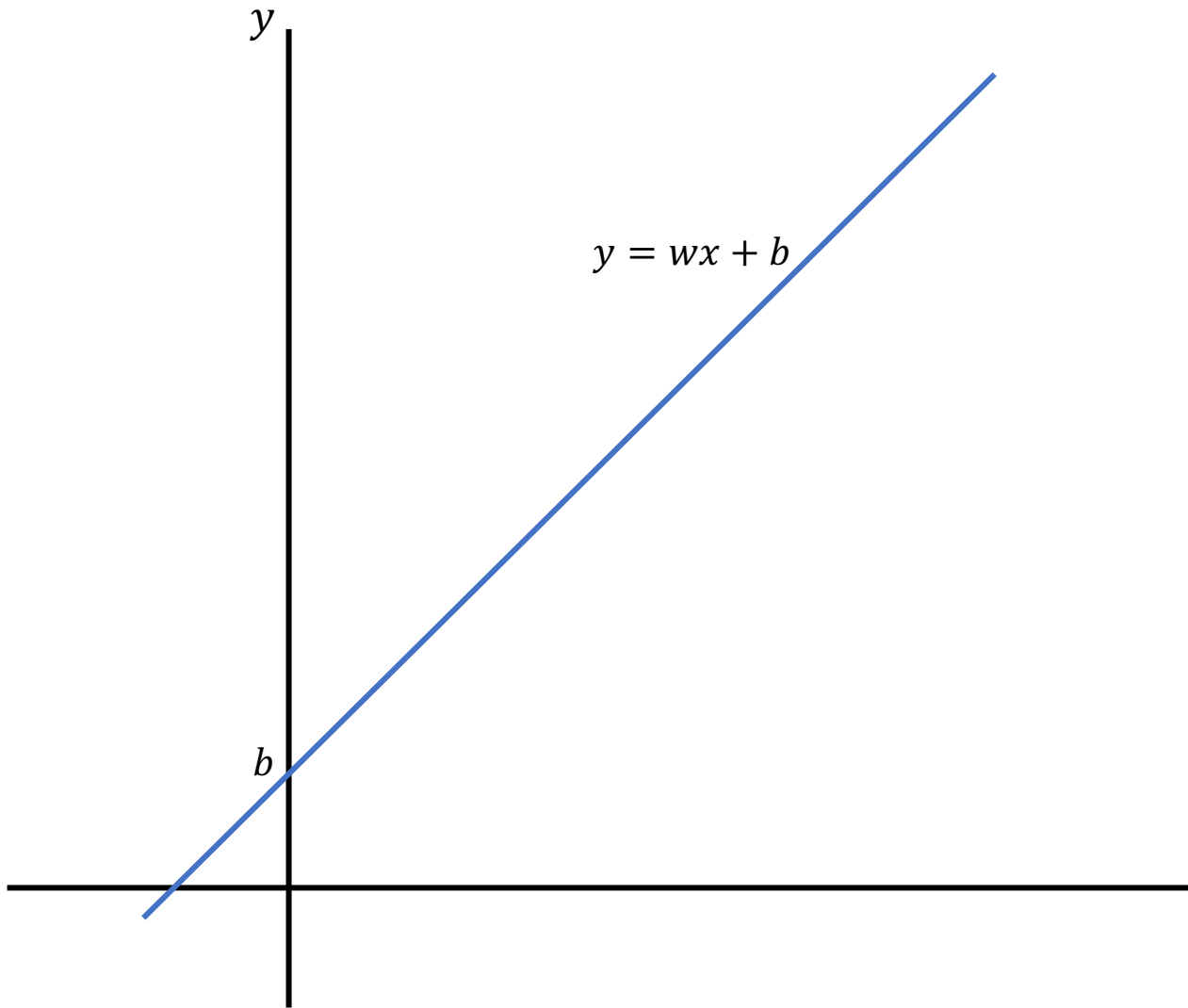
## 아래 기울기를 찾기 위한 방법

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

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

## 역전파를 이해하기 위한 수학 지식 (1) 기울기란(미분)

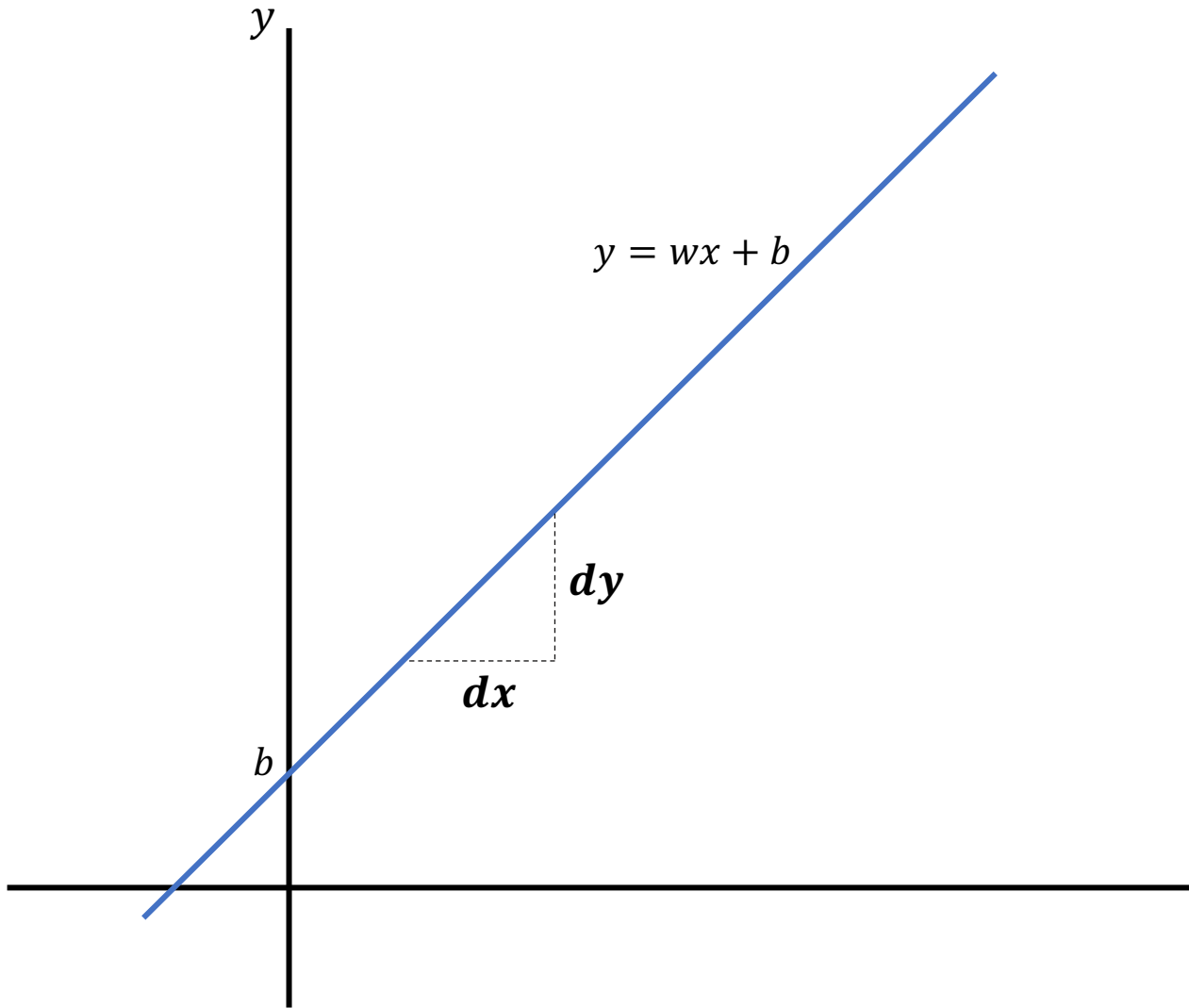
---



$$y = wx + b$$

## 역전파를 이해하기 위한 수학 지식 (1) 기울기란(미분)

---

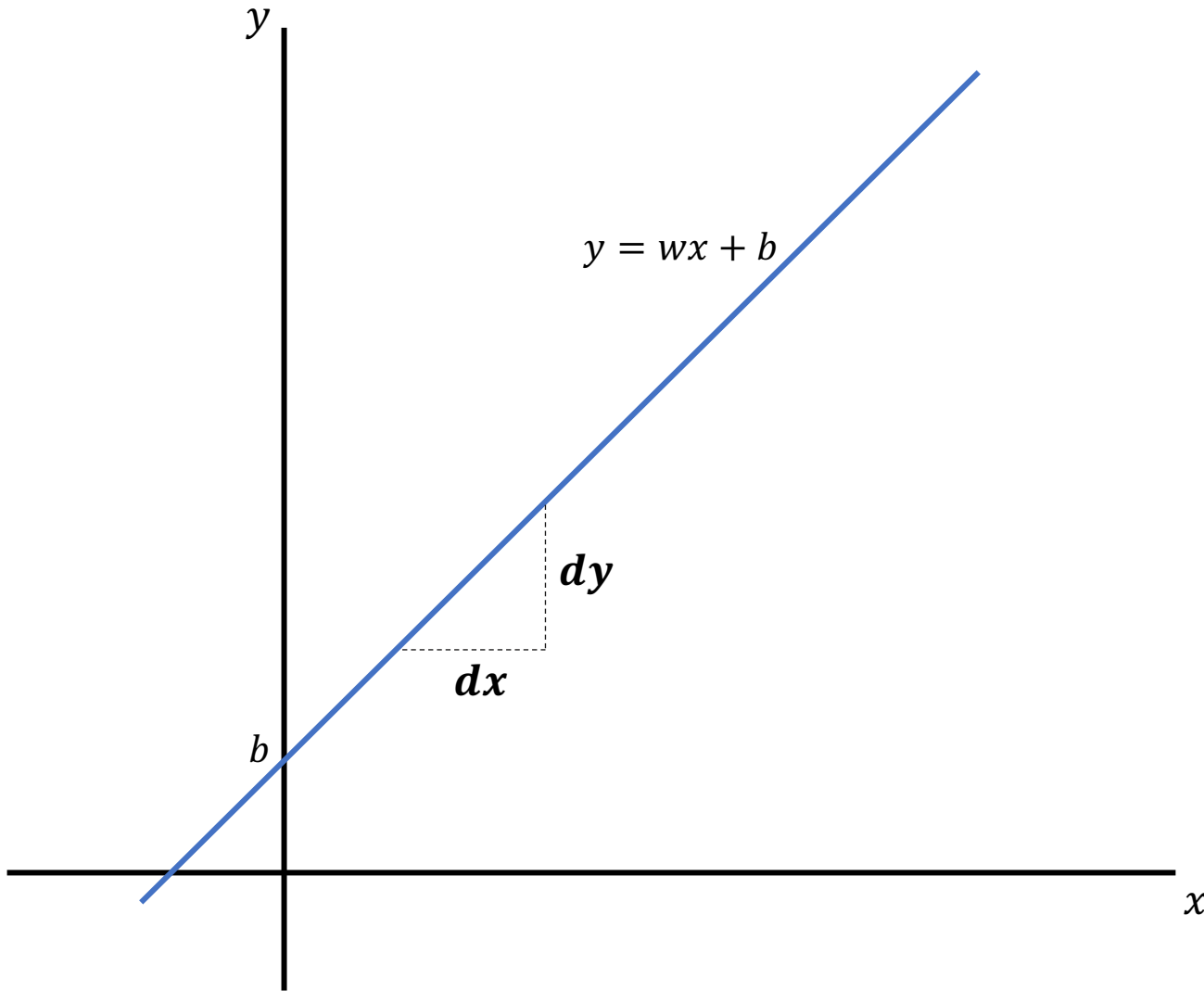


$$y = wx + b$$

$\frac{dy}{dx}$  :  $x$ 의 변화량( $dx$ )에 대한  $y$ 의 변화량 ( $dy$ )



## 역전파를 이해하기 위한 수학 지식 (1) 기울기란(미분)

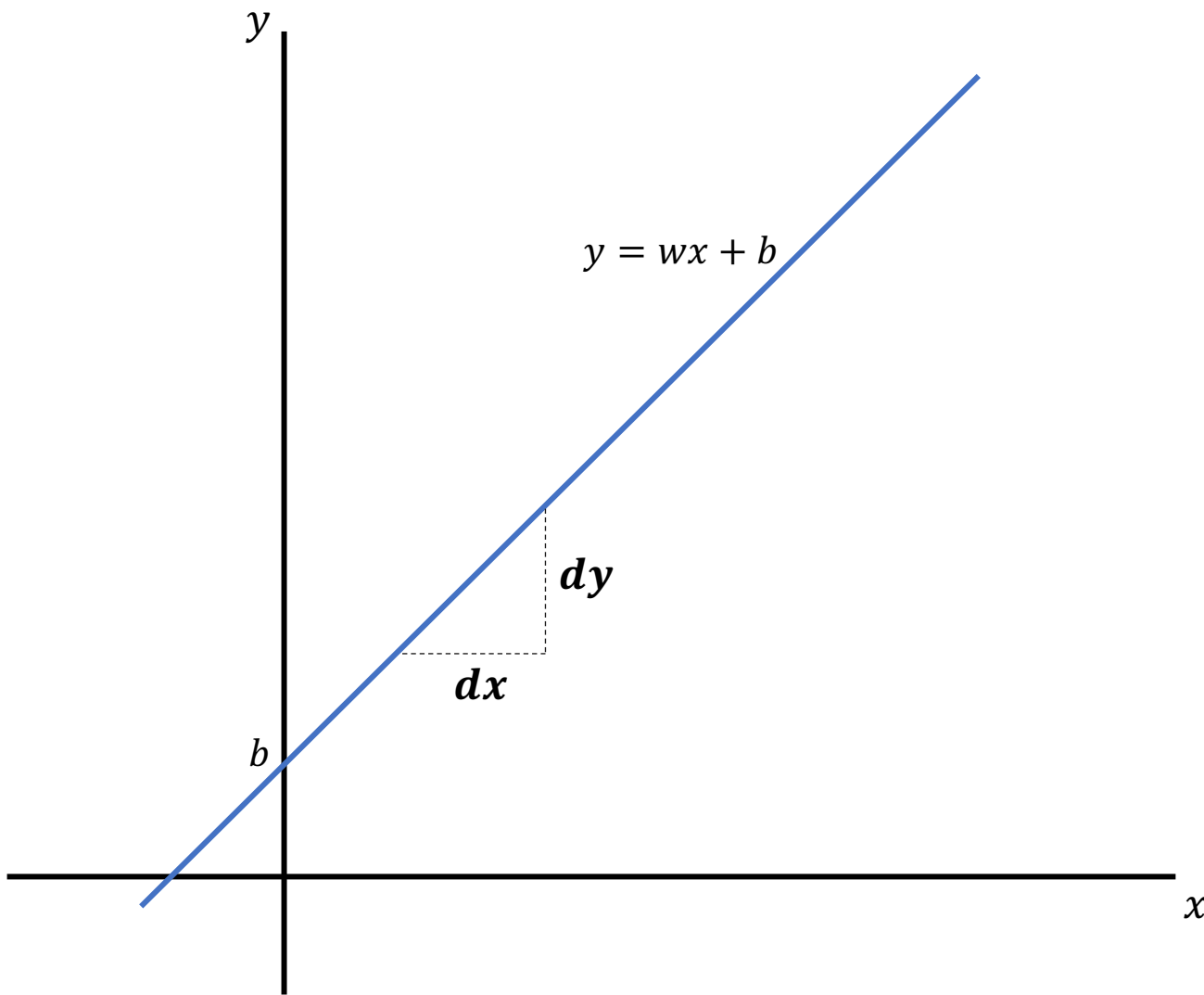


$$y = wx + b$$

$\frac{dy}{dx}$  :  $x$ 의 변화량( $dx$ )에 대한  $y$ 의 변화량 ( $dy$ )

Q)  $x$ 가 1이 증가하면  $y$ 는 얼마나 변화할까?

# 역전파를 이해하기 위한 수학 지식 (1) 기울기란(미분)



$$y = wx + b$$

$\frac{dy}{dx}$  :  $x$ 의 변화량( $dx$ )에 대한  $y$ 의 변화량 ( $dy$ )

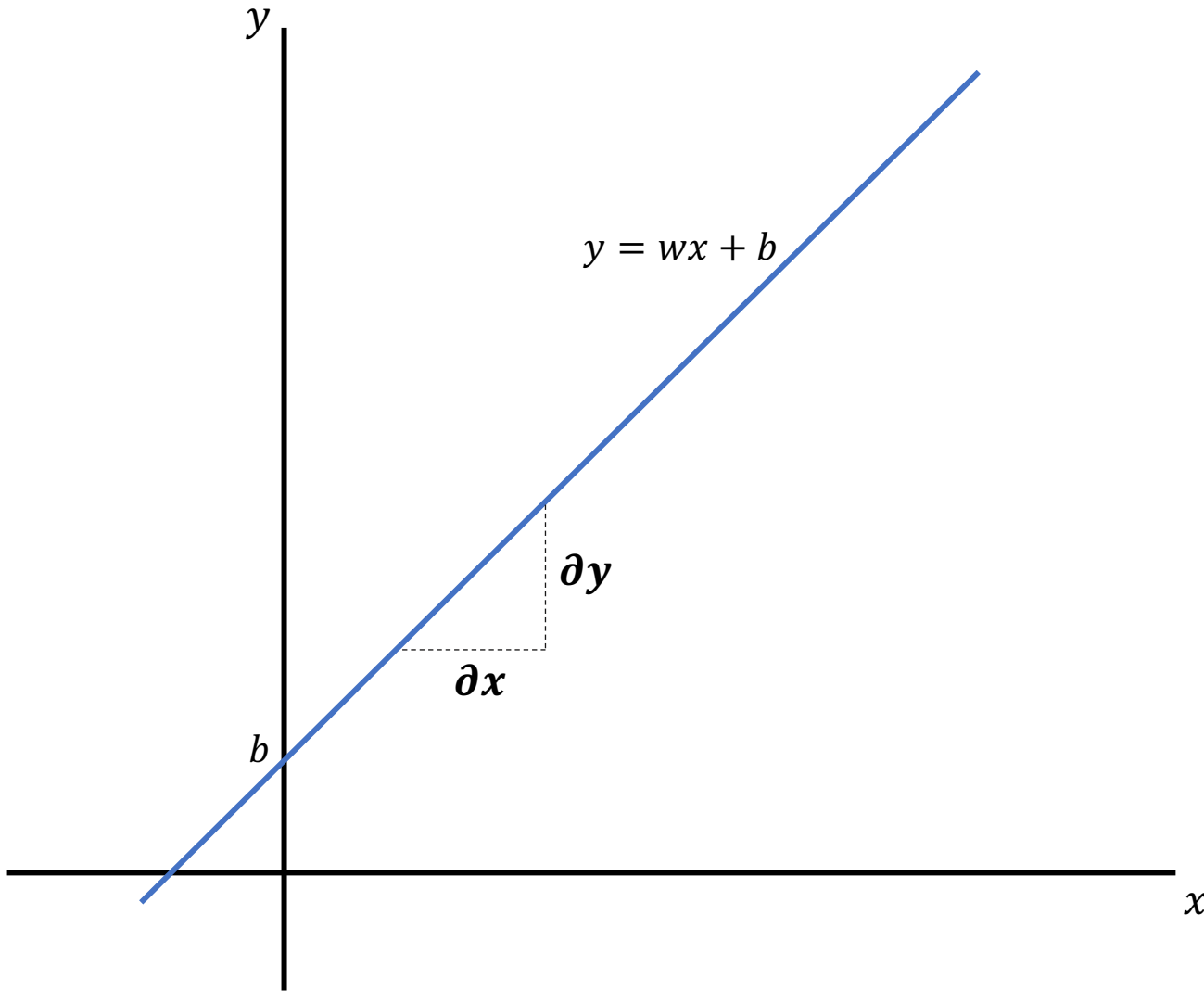
Q)  $x$ 가 1이 증가하면  $y$ 는 얼마나 변화할까?

$w$ 만큼 변화한다

$$\frac{dy}{dx} = w$$

## 역전파를 이해하기 위한 수학 지식 (2) 편미분

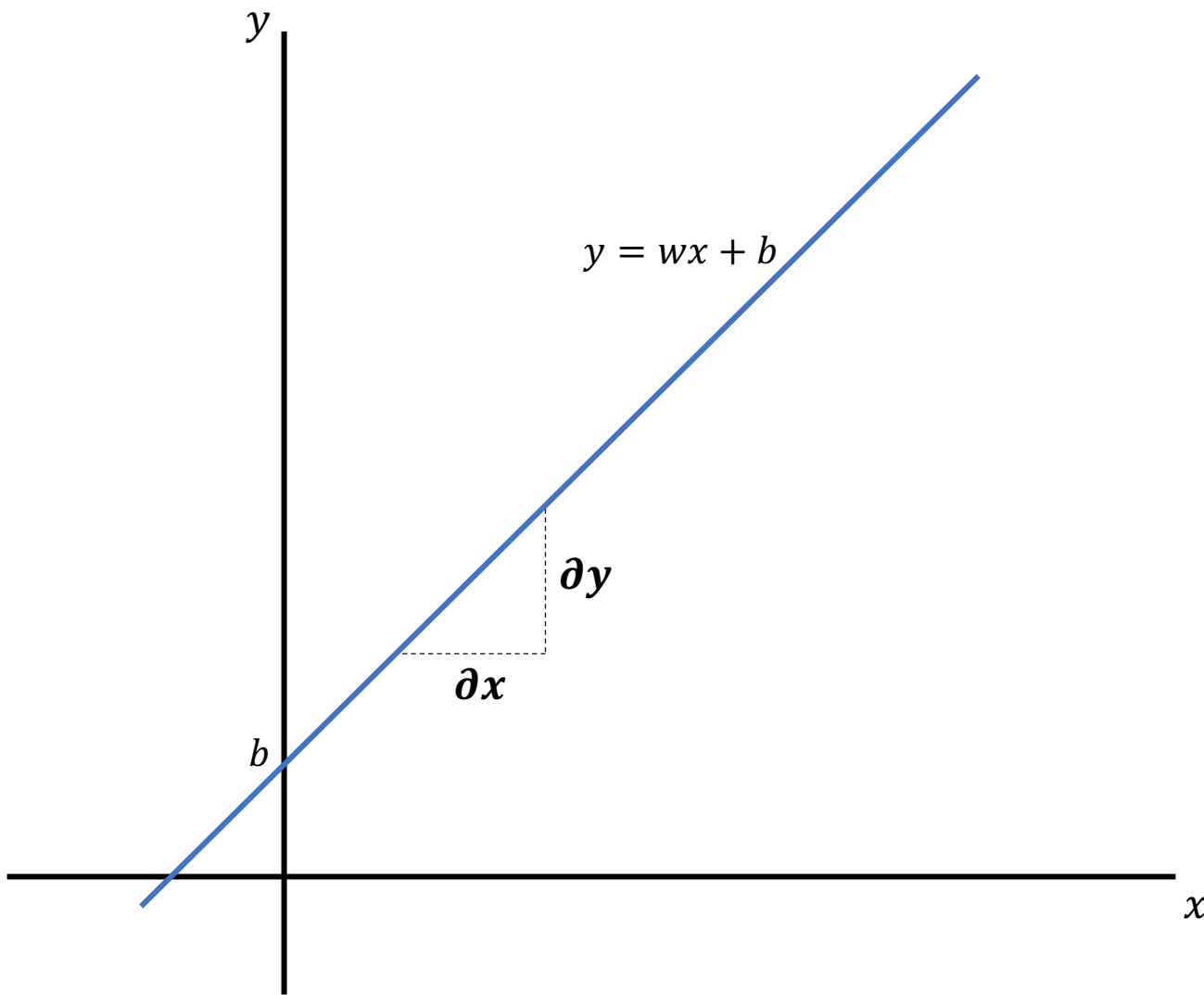
---



$$\underline{y = wx + b}$$

각 변수( $w$ ,  $x$ ,  $b$ ) 별로 기울기를 구해보자  
-> 편미분

## 역전파를 이해하기 위한 수학 지식 (2) 편미분

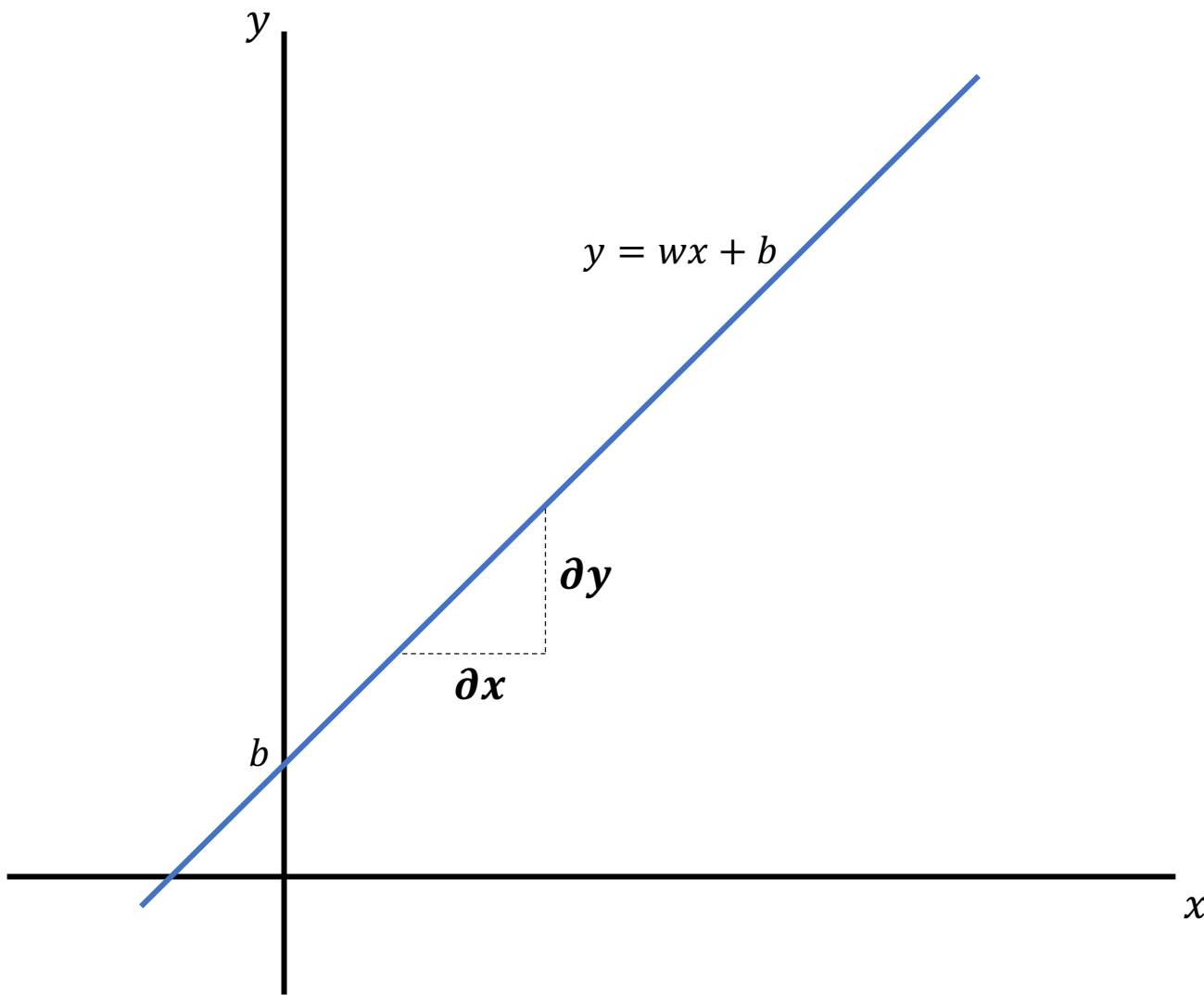


$$\underline{y = wx + b}$$

$\frac{\partial y}{\partial x}$  :  $x$ 의 변화량( $\partial x$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial x} = w$$

## 역전파를 이해하기 위한 수학 지식 (2) 편미분



$$\underline{y = wx + b}$$

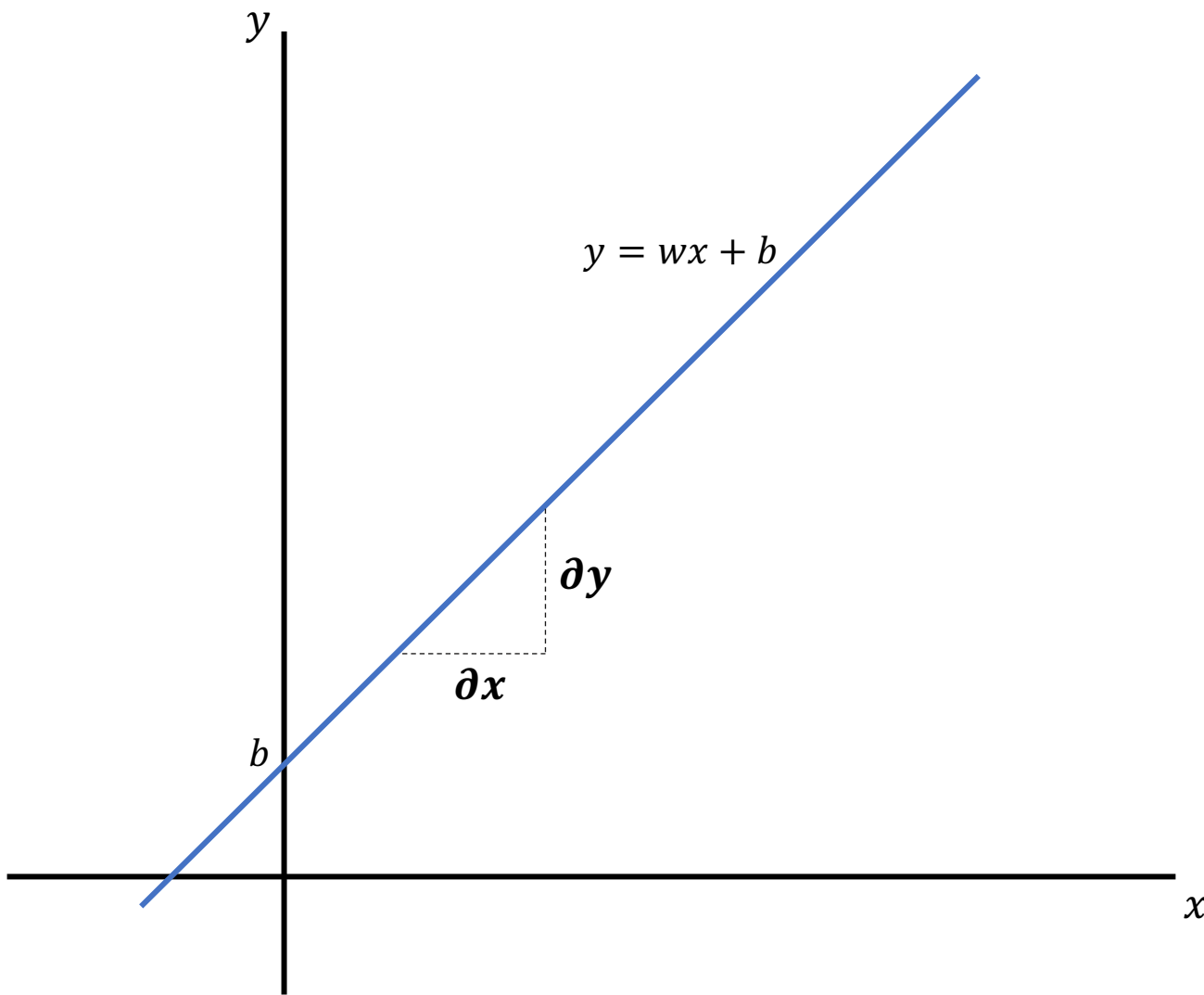
$\frac{\partial y}{\partial x}$  :  $x$ 의 변화량( $\partial x$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial x} = w$$

$\frac{\partial y}{\partial w}$  :  $w$ 의 변화량( $\partial w$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial w} = x$$

## 역전파를 이해하기 위한 수학 지식 (2) 편미분



$$\underline{y = wx + b}$$

$\frac{\partial y}{\partial x}$  :  $x$ 의 변화량( $\partial x$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial x} = w$$

$\frac{\partial y}{\partial w}$  :  $w$ 의 변화량( $\partial w$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial w} = x$$

$\frac{\partial y}{\partial b}$  :  $b$ 의 변화량( $\partial b$ )에 대한  $y$ 의 변화량 ( $\partial y$ )

$$\frac{\partial y}{\partial b} = 1$$

## 역전파를 이해하기 위한 수학 지식 (3) 행렬미분

---

$$\underline{Y = XW + b}$$

$\frac{dY}{dX}$  :  $X$ 의 각 원소 별 변화량( $dx_1, dx_2, \dots$ )에 대한  $y$ 의 각 원소 별 변화량 ( $dy_1, dy_2, \dots$ )

## 역전파를 이해하기 위한 수학 지식 (3) 행렬미분

$$\underline{Y = XW + b}$$

$\frac{dY}{dX}$  :  $X$ 의 각 원소 별 변화량( $dx_1, dx_2, \dots$ )에 대한  $y$ 의 각 원소 별 변화량 ( $dy_1, dy_2, \dots$ )

$$\frac{dY}{dX} = \left( \frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n} \right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left( \frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n} \right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$

벡터의 미분이 전치행렬이 되는 이유

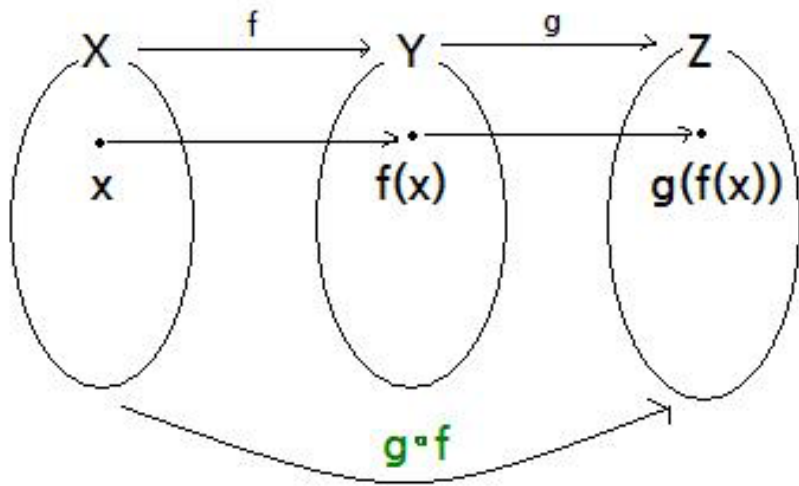
$$Y = XW = (x_1, x_2, \dots, x_n) \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

그러므로,  $Y$ 에 대한  $W$ 의 기울기를 구하면,

$$\frac{dY}{dW} = \begin{bmatrix} \frac{dY}{dw_1} \\ \frac{dY}{dw_2} \\ \dots \\ \frac{dY}{dw_n} \end{bmatrix} = \begin{bmatrix} \frac{d(w_1 x_1 + w_2 x_2 + \dots + w_n x_n)}{dw_1} \\ \frac{d(w_1 x_1 + w_2 x_2 + \dots + w_n x_n)}{dw_2} \\ \dots \\ \frac{d(w_1 x_1 + w_2 x_2 + \dots + w_n x_n)}{dw_n} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = X^T$$



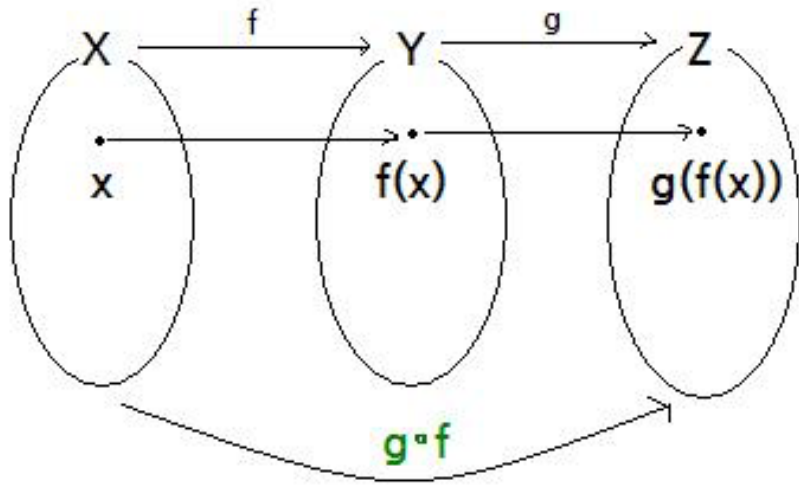
## 역전파를 이해하기 위한 수학 지식 (4) 합성함수



함수

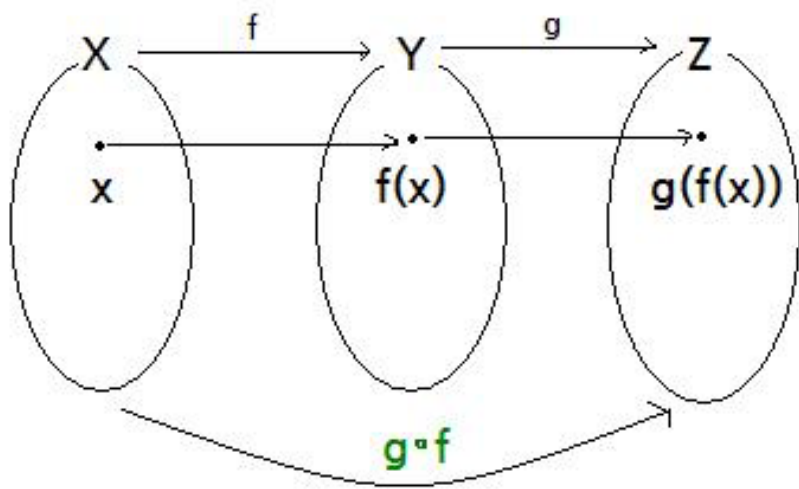
$Y = f(X)$	
$Z = g(Y)$	
$Z = g(f(X))$	

## 역전파를 이해하기 위한 수학 지식 (4) 합성함수



함수	미분
$Y = f(X)$	$\frac{dY}{dX} = f'(X)$
$Z = g(Y)$	$\frac{dZ}{dY} = g'(Y)$
$Z = g(f(X))$	$\frac{dZ}{dX} = g'(f(X))f'(X)$

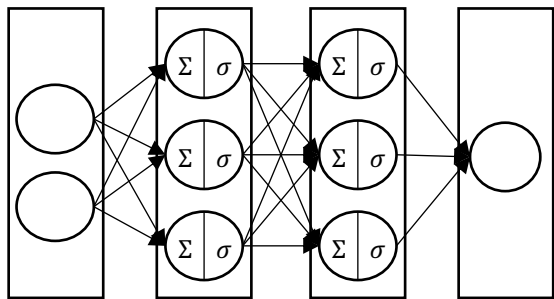
## 역전파를 이해하기 위한 수학 지식 (4) 합성함수



함수	미분
$Y = f(X)$	$\frac{dY}{dX} = f'(X)$
$Z = g(Y)$	$\frac{dZ}{dY} = g'(Y)$
$Z = g(f(X))$	$\frac{dZ}{dX} = g'(f(X))f'(X) = \frac{dZ}{dY} \frac{dY}{dX}$

각 함수의 기울기로 합성함수의 기울기를 계산할 수 있음

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

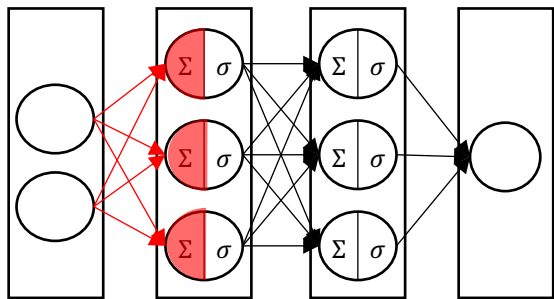
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

각층의 기울기

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $y_{pred}$

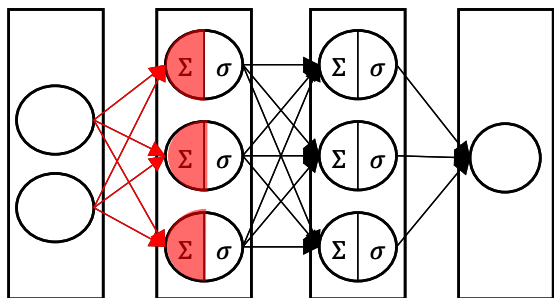
출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

각층의 기울기

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

각층의 기울기

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

행렬곱 미분 Remind

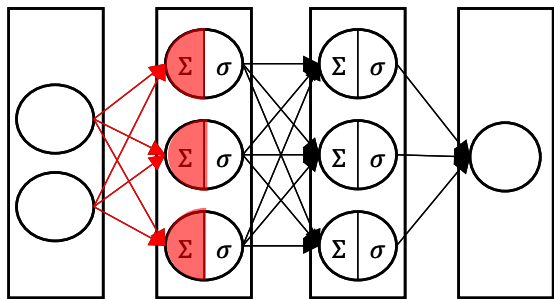
$$Y = XW + b$$

$$\frac{dY}{dX} = \left( \frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n} \right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left( \frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n} \right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

각층의 기울기

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

행렬곱 미분 Remind

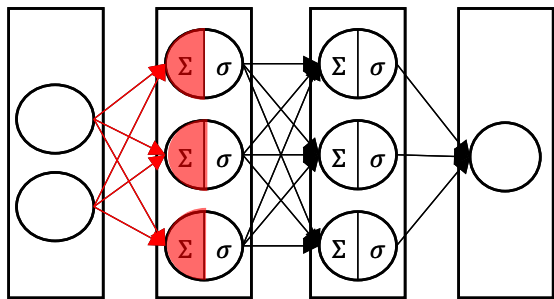
$$Y = XW + b$$

$$\frac{dY}{dX} = \left( \frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n} \right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left( \frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n} \right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

각층의 기울기

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

행렬곱 미분 Remind

$$Y = XW + b$$

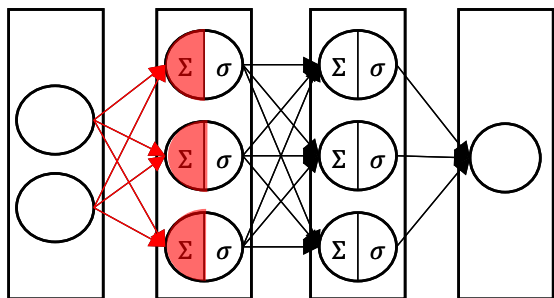
$$\frac{dY}{dX} = \left( \frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n} \right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left( \frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n} \right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$



# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

$$Z^{[1]} = \textcolor{red}{X}W^{[1]} + b^{[1]}$$

각층의 기울기

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

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

$$\frac{dZ^{[1]}}{d\textcolor{red}{X}} = W^{T[1]}$$

행렬곱 미분 Remind

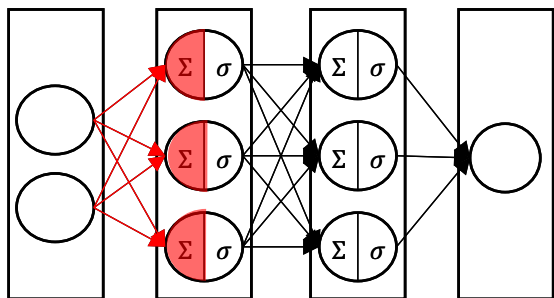
$$Y = XW + b$$

$$\frac{dY}{dX} = \left(\frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n}\right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left(\frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n}\right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

각층의 기울기

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

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

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

행렬곱 미분 Remind

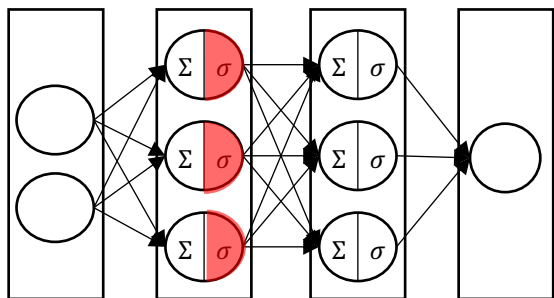
$$Y = XW + b$$

$$\frac{dY}{dX} = \left( \frac{dy_1}{dx_1}, \frac{dy_2}{dx_2}, \dots, \frac{dy_n}{dx_n} \right) = (w_1^T, w_2^T, \dots, w_n^T) = W^T$$

$$\frac{dY}{dW} = \left( \frac{dy_1}{dw_1}, \frac{dy_2}{dw_2}, \dots, \frac{dy_n}{dw_n} \right) = (x_1^T, x_2^T, \dots, x_n^T) = X^T$$

$$\frac{dY}{db} = 1$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

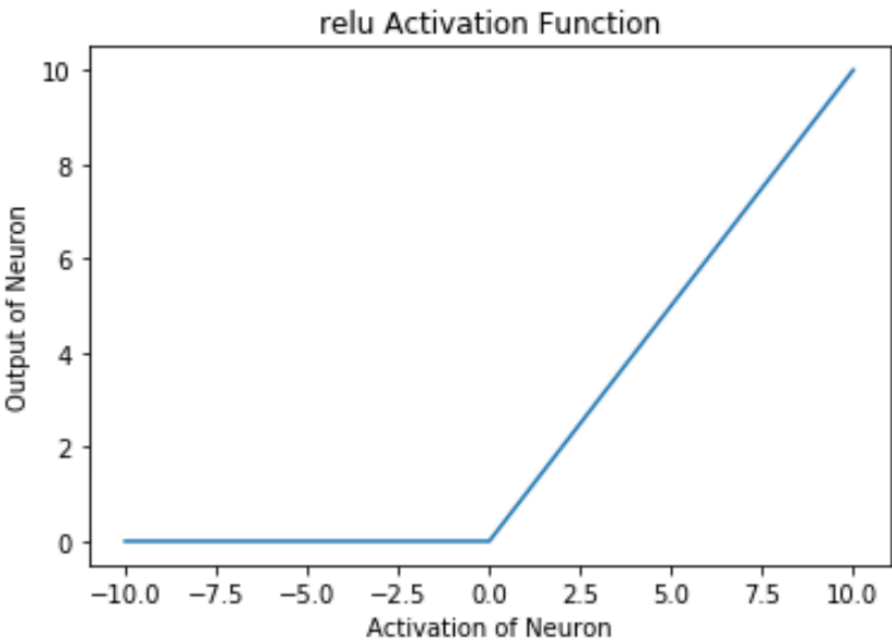
$$a^{[1]} = \text{relu}(Z^{[1]})$$

각층의 기울기

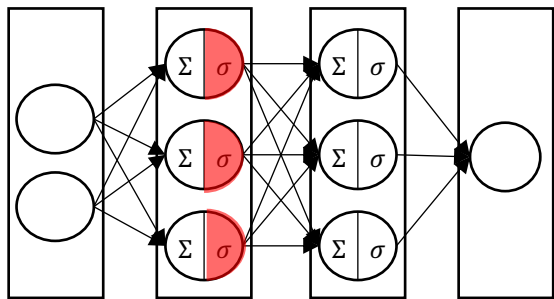
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

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

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



# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

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

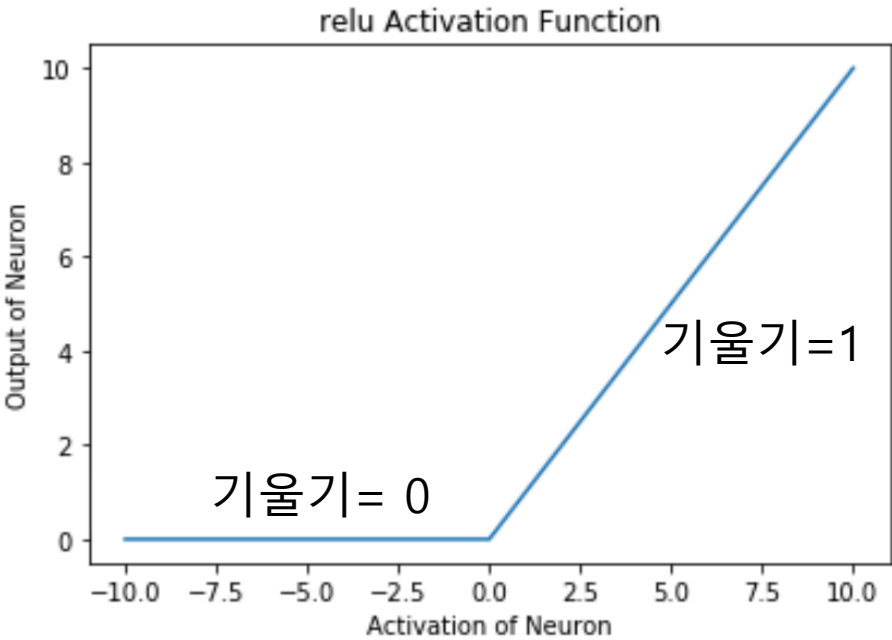
각층의 기울기

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

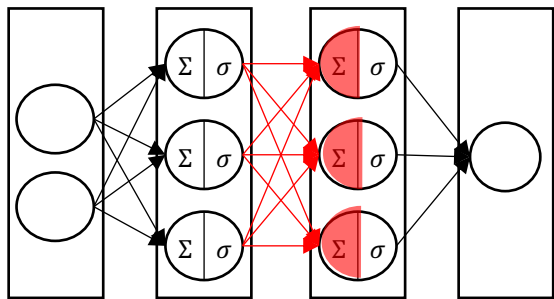
$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

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

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



# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

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

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

각층의 기울기

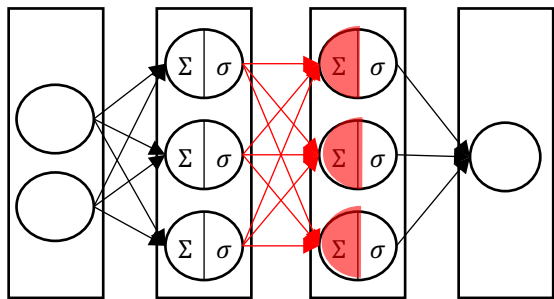
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

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

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

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파

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

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

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

각층의 기울기

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

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

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

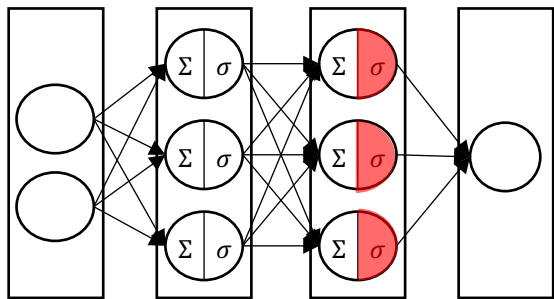
$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

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

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

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

## 순전파

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

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

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

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

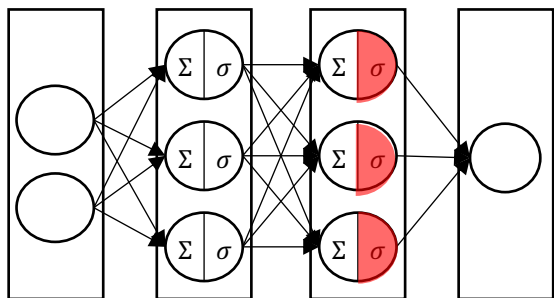
## 각층의 기울기

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

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

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

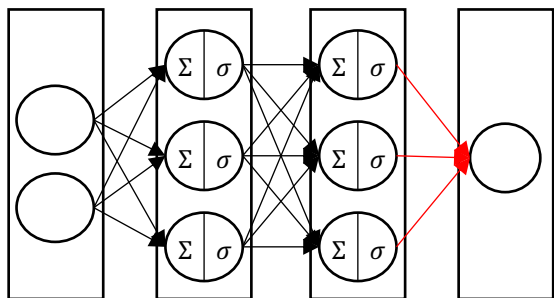
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$		



# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

## 순전파

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

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

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

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

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

## 각층의 기울기

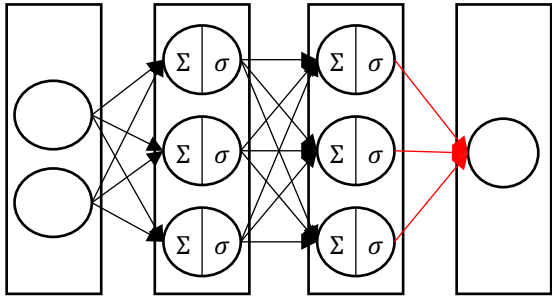
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

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

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

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

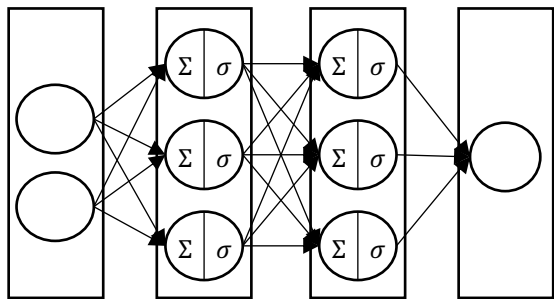
첫번째 은닉층의 로짓 값:  $z^{[1]}$     첫번째 은닉층의 활성화 값:  $a^{[1]}$     첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$     두번째 은닉층의 활성화 값:  $a^{[2]}$     두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$     출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$		
$Z^{[1]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[1]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

## 순전파

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

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

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

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

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

$$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$$

## 각층의 기울기

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

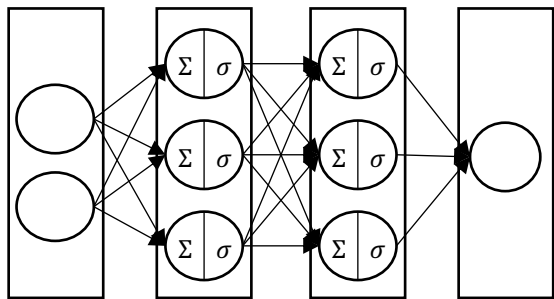
$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

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

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

$$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]} \quad \frac{dY_{pred}}{db^{[3]}} = 1 \quad \frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

## 순전파

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

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

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

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

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

$$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$$

## 각층의 기울기

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

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

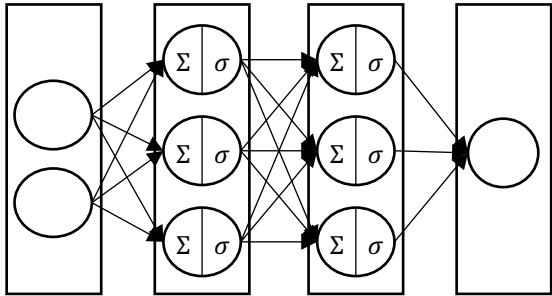
$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

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

$$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]} \quad \frac{dY_{pred}}{db^{[3]}} = 1 \quad \frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$$

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

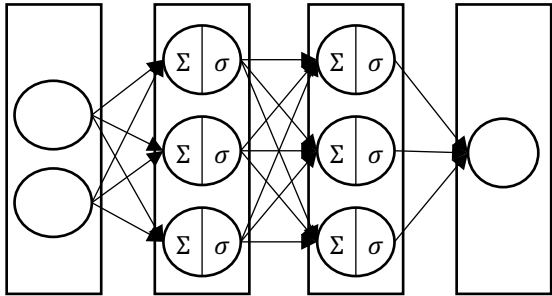
순전파	각층의 기울기		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$		

## 우리가 찾고자 하는 값

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

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

# 역전파 알고리즘 (1) 순전파를 하면서 각층의 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$		

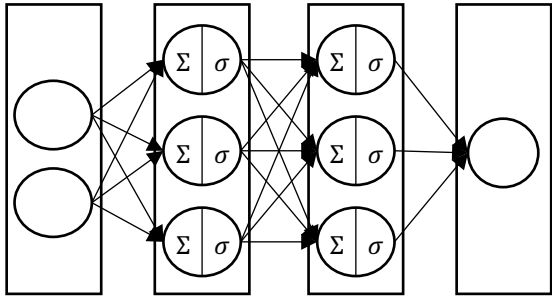
## 우리가 찾고자 하는 값

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

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

각층의 기울기 정보를 바탕으로,  
우리가 찾고자 하는 손실함수에 대한  
기울기를 찾자!

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$

첫번째 은닉층의 활성화 값:  $a^{[1]}$

첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$

두번째 은닉층의 활성화 값:  $a^{[2]}$

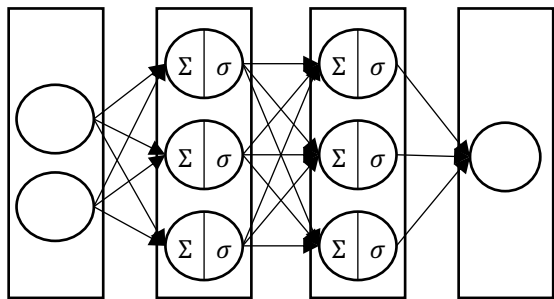
두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} =$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

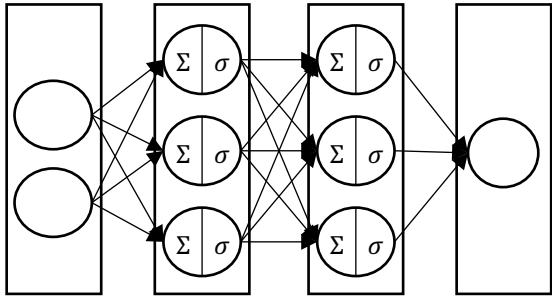
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			



# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

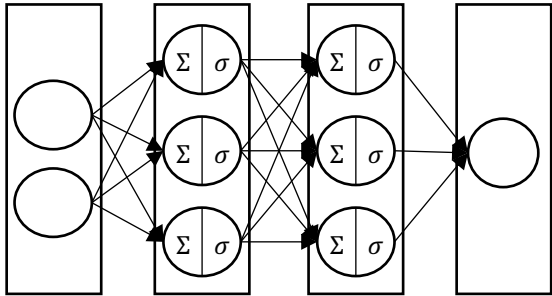
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$ <p>합성함수의 미분 법칙</p>

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

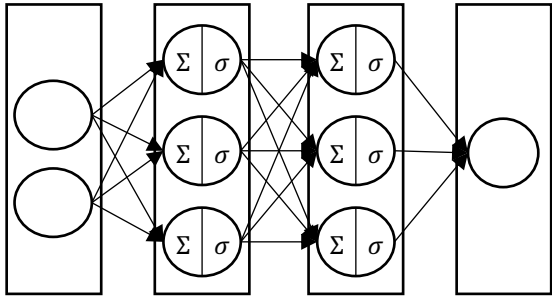
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			$\frac{dL}{db^{[3]}}$

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

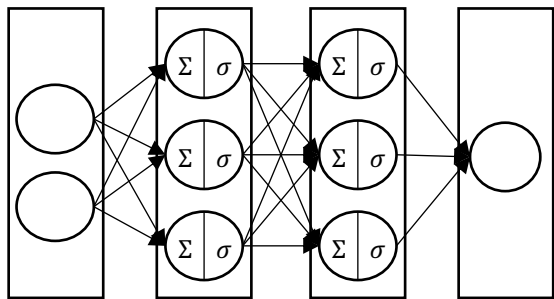
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}} \quad \frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

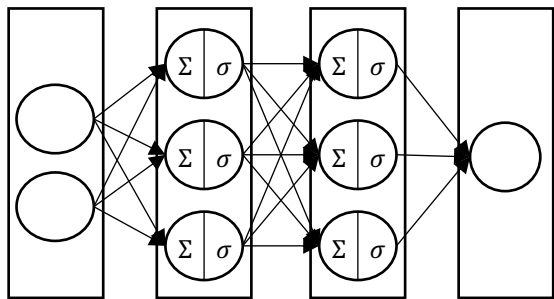
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$			
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$					
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$			
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$					
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

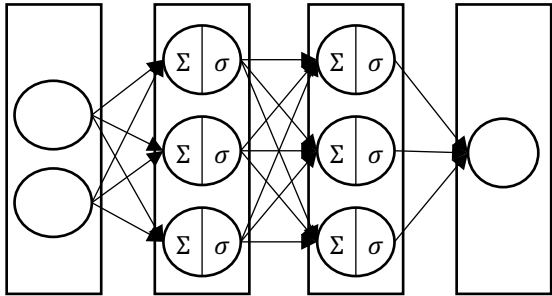
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$			
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$					
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$			
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$					
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

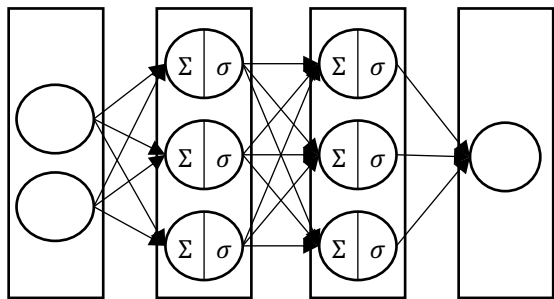
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}}$
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}} \quad \frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}} \quad \frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

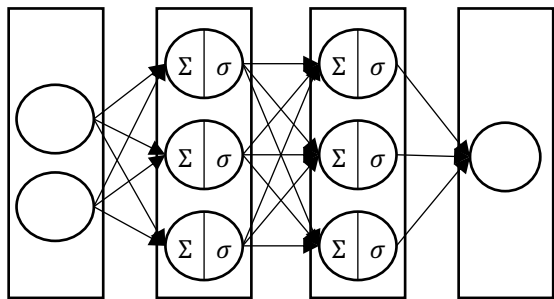
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$			
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$					
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$			
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

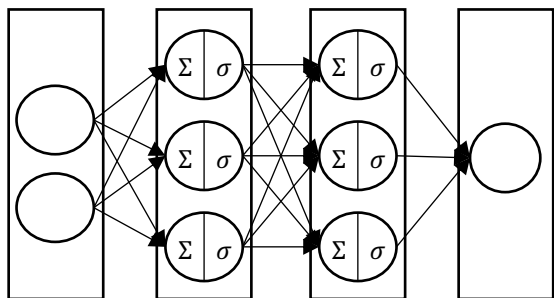
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}} \quad \frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}} \quad \frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$			



# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

## 순전파

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

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

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

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

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

$$\text{Loss} = \frac{1}{2}(y_{pred} - y_{true})^2$$

## 각층의 기울기

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

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

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

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

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

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

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

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

$$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{dY_{pred}}{db^{[3]}} = 1$$

$$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$$

## 역전파

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

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

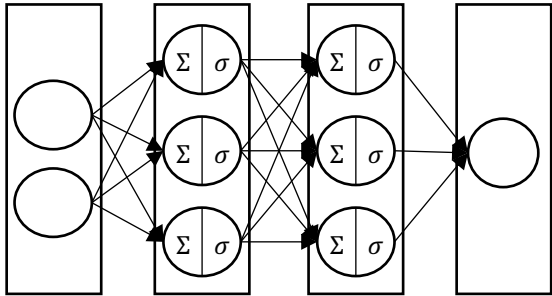
$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$$

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

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

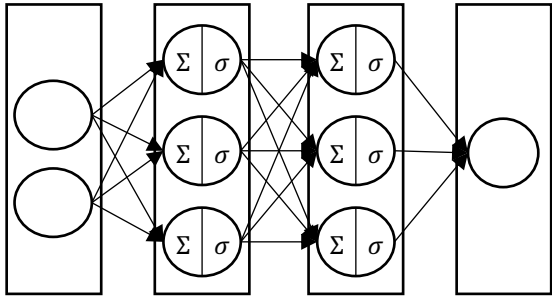
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$			
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$					
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

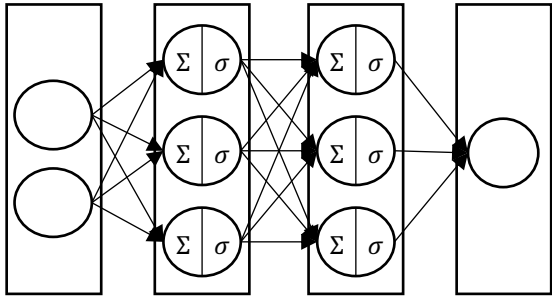
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$			
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

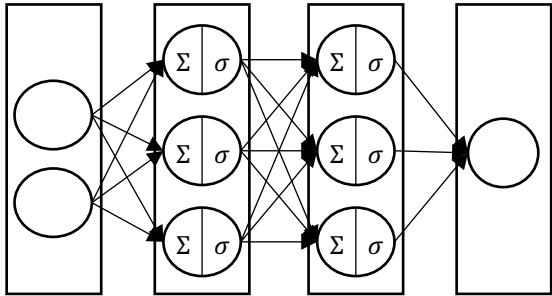
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$		
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$


첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

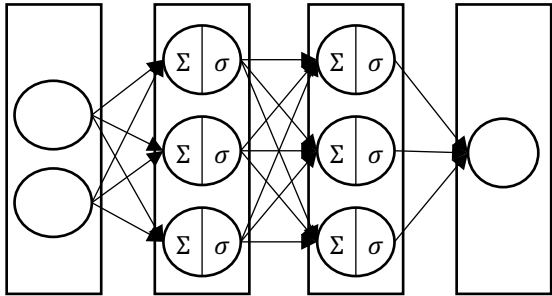
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$	$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$	
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					



# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

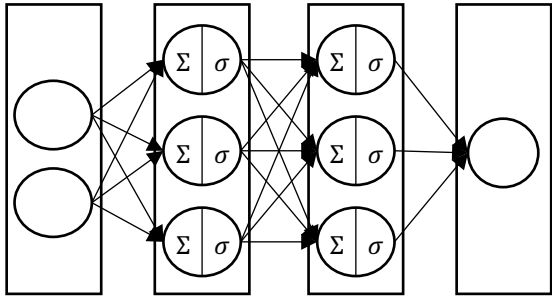
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$	$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$	$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 (2) 각 가중치의 손실에 대한 기울기를 구하기



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

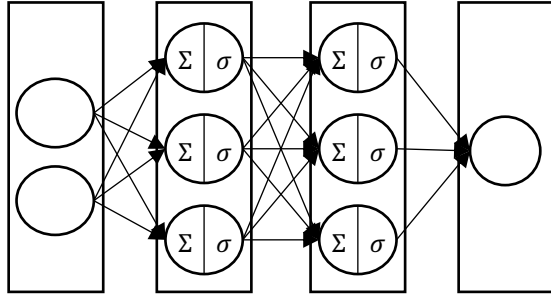
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$	$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$	$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

# 역전파 알고리즘 개괄



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

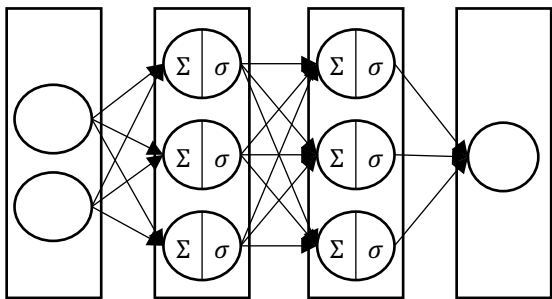
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$



# 역전파 알고리즘 개괄



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$     첫번째 은닉층의 활성화 값:  $a^{[1]}$     첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$     두번째 은닉층의 활성화 값:  $a^{[2]}$     두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

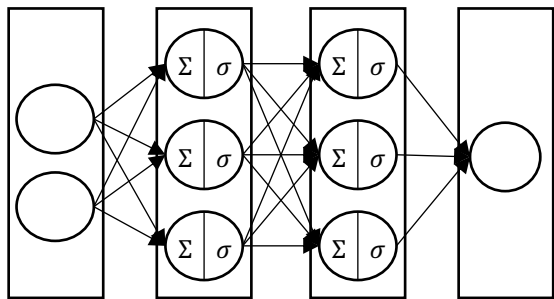
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$		

(1) 순전파를 진행하면서 각 층의 기울기 및 손실함수 값을 구하고

# 역전파 알고리즘 개괄



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

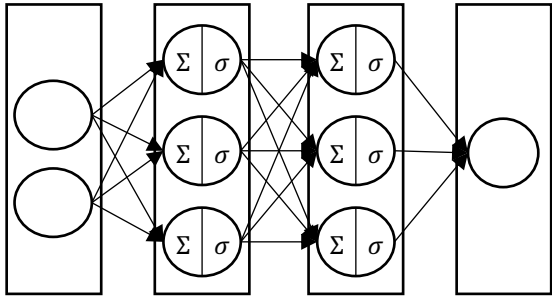
출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$	$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$	$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					

(2) 각층의 기울기를 통해 역으로 각 가중치의 기울기를 구함

# 역전파를 통해 알아야 하는 점 (1) 학습 순서



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

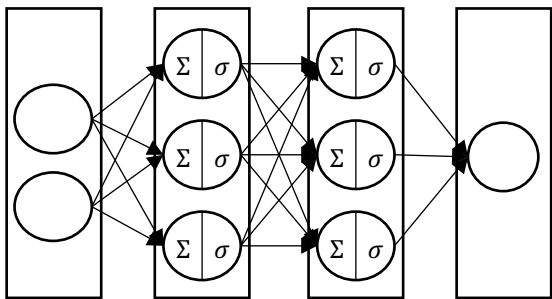
두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기			역전파		
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$	$\frac{dZ^{[1]}}{db^{[1]}} = 1$	$\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$	$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$	$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$
$a^{[1]} = relu(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$		
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$	$\frac{dZ^{[2]}}{db^{[2]}} = 1$	$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$	$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$	$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = relu(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$			$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$		
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = a^{T[2]}$	$\frac{dY_{pred}}{db^{[3]}} = 1$	$\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$	$\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$	$\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$					
(1) 순전파				(2) 역전파		

# 역전파를 통해 알아야 하는 점 (2) 메모리



입력층 :  $X$

첫번째 은닉층의 로짓 값:  $z^{[1]}$  첫번째 은닉층의 활성화 값:  $a^{[1]}$  첫번째 은닉층의 가중치 :  $w^{[1]}, b^{[1]}$

두번째 은닉층의 로짓 값:  $z^{[2]}$  두번째 은닉층의 활성화 값:  $a^{[2]}$  두번째 은닉층의 가중치 :  $w^{[2]}, b^{[2]}$

출력층 :  $Y_{pred}$

출력층의 가중치 :  $w^{[3]}, b^{[3]}$

순전파	각층의 기울기	역전파
$Z^{[1]} = XW^{[1]} + b^{[1]}$	$\frac{dZ^{[1]}}{dW^{[1]}} = \textcolor{red}{X}^T$ $\frac{dZ^{[1]}}{db^{[1]}} = 1$ $\frac{dZ^{[1]}}{dX} = W^{T[1]}$	$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$ $\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$ $\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$
$a^{[1]} = \text{relu}(Z^{[1]})$	$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$	$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$
$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$	$\frac{dZ^{[2]}}{dW^{[2]}} = \textcolor{red}{a^{T[1]}}$ $\frac{dZ^{[2]}}{db^{[2]}} = 1$ $\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$	$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$ $\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$ $\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$
$a^{[2]} = \text{relu}(Z^{[2]})$	$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$	$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$
$Y_{pred} = a^{[2]}W^{[3]} + b^{[3]}$	$\frac{dY_{pred}}{dW^{[3]}} = \textcolor{red}{a^{T[2]}}$ $\frac{dY_{pred}}{db^{[3]}} = 1$ $\frac{dY_{pred}}{da^{[2]}} = W^{T[3]}$	$\frac{dL}{dW^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{dW^{[3]}}$ $\frac{dL}{db^{[3]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{db^{[3]}}$ $\frac{dL}{da^{[2]}} = \frac{dL}{dY_{pred}} \frac{dY_{pred}}{da^{[2]}}$
$Loss = \frac{1}{2}(y_{pred} - y_{true})^2$	$\frac{dL}{dY_{pred}} = y_{pred} - y_{true}$	

각 층에서의 기울기 정보(특히 **출력 값**)을 저장해야 함 -> 메모리 많이 필요