# Backpropagation

기계학습/딥러닝

임 경 태

| Week | Chapter | Contents |
|---|---|---|
| 1 | 1, 2장 | 강의 소개, 파이썬 복습 |
| 2 | 1, 3장 | 파이썬 복습, Numpy, Pandas |
| 3 | 1, 4장 | 딥러닝을 위한 미분 |
| 4 | 5장 | 회귀 |
| 5 | 5장 | 분류 |
| 6 | 6장 | XOR문제 |
| 7 | 7장 | 딥러닝 |
| 8 | 1~7장 | 중간고사 |
| 9 | 8장 | MNIST 필기체 구현 및 팀 프로젝트 소개 |
| 10 | 9장 | 오차역전파 |
| **11** | 11장 | Jetbot 자율주행 (Collision Avoidance, Transfer Learning) |
| 12 | 12장 | 특강 (인공지능 활용 연구) + 팀프로젝트 자율 실습 |
| **13** | 10장 | Jetbot 자율주행 (Road Following) |
| 14 | 11장 | 합성곱 신경망(CNN), 순환 신경망(RNN) + 팀프로젝트 자율 실습 |
| **15** | 8~12장 | 기말고사 (or 프로젝트 발표) |

# CONTENTS

목적 : 오차역전파를 이용하는 이유와 오차역전파에 대한 이해

목표 : 각 층에서의 오차역전파 진행 과정 이해

내용 : 수치 미분 문제점, 오차역전파 원리, MNIST 검증

# CONTENTS

# 수치 미분 문제점

## 수치 미분을 사용한 가중치 업데이트 과정

1. Training Data 입력

2. Feed Forward

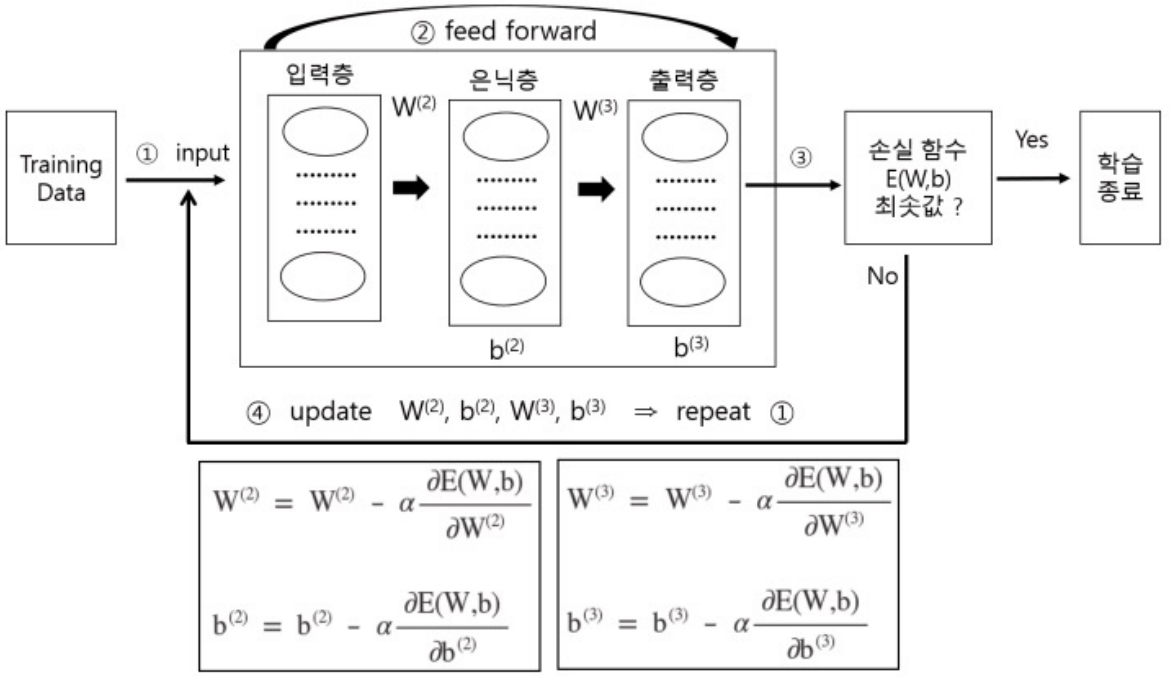
3. 손실함수 계산

4. 수치 미분을 통하여 가중치 업데이트

5. 2~4 반복

   - 많은 시간 소요
   - 수치 미분에서의 오차 포함

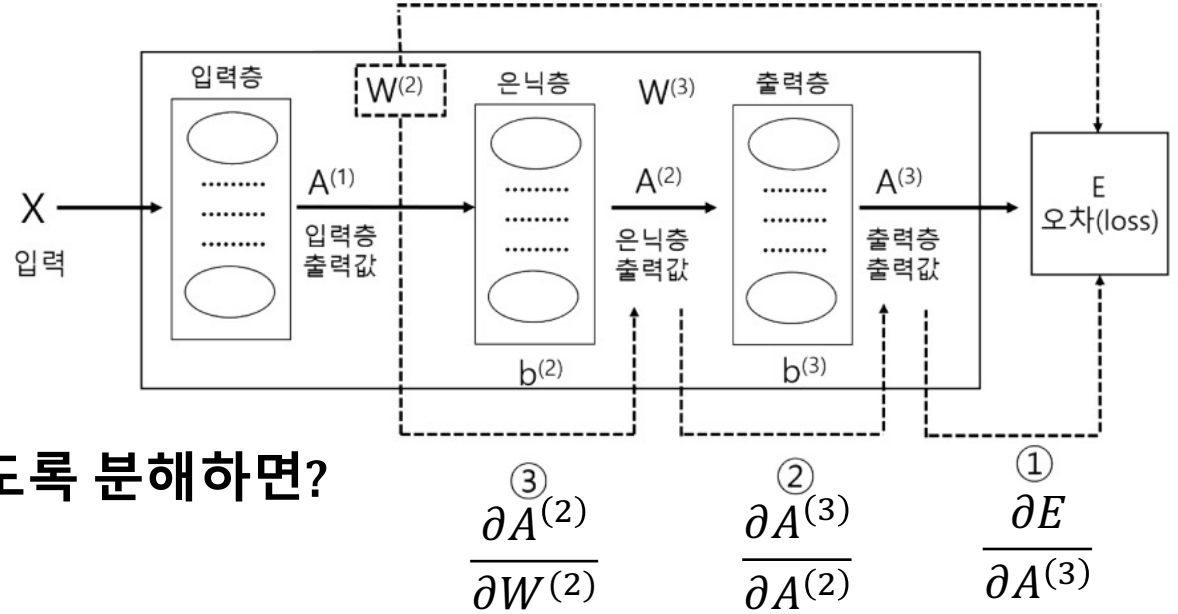
   → Chain Rule을 사용하여 분해, 오차역전파



# 오차역전파 개념 및 동작 원리

## Chain Rule을 적용하여 수식 분해

$A^{(1)}$: 입력층 출력 값
$A^{(2)}$: 은닉층 출력 값
$A^{(3)}$: 출력층 출력 값




$\frac{\partial E}{\partial W^{(2)}}$ 를 chain Rule을 사용하여 $A^{(2)}, A^{(3)}$를 포함하도록 분해하면?

→ $\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial W^{(2)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial A^{(2)}} \cdot \frac{\partial A^{(2)}}{\partial W^{(2)}}$

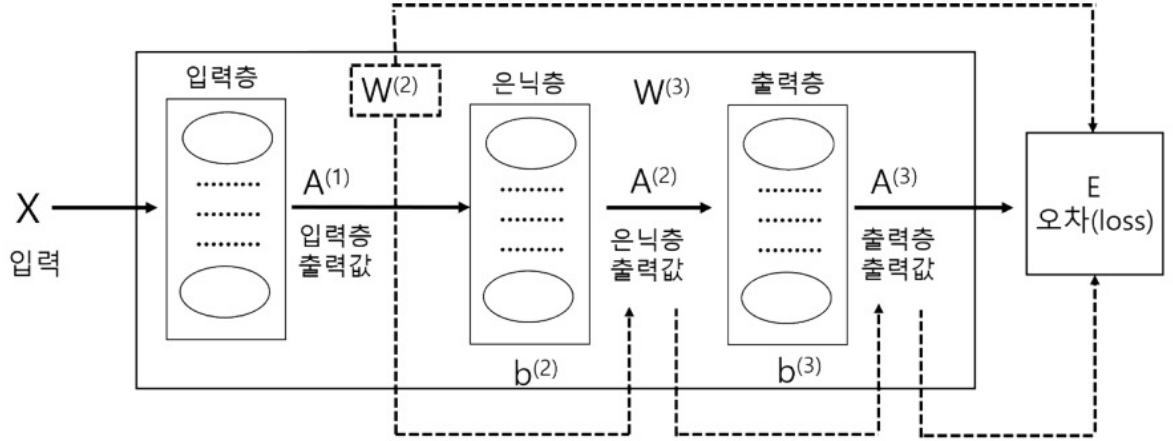
Chain Rule을 사용하여 $A^{(2)}, A^{(3)}$를 포함하도록 가중치 업데이트 식을 분해하면?

→ $W^{(2)} := W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}} = W^{(2)} - \alpha \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial A^{(2)}} \cdot \frac{\partial A^{(2)}}{\partial W^{(2)}}$

# 오차역전파 개념 및 동작 원리

## Chain Rule을 적용하여 수식 분해

$A^{(1)}$: 입력층 출력 값
$A^{(2)}$: 은닉층 출력 값
$A^{(3)}$: 출력층 출력 값




$\frac{\partial E}{\partial W^{(2)}}$은 chain Rule에 의해 국소적인 미분의 곱셈으로 나타낼 수 있다!

③ $\frac{\partial A^{(2)}}{\partial W^{(2)}}$ ② $\frac{\partial A^{(3)}}{\partial A^{(2)}}$ ① $\frac{\partial E}{\partial A^{(3)}}$

✓ **국소적인 미분?**

- 이전 layer의 출력 값 변화에 따른 다음 layer의 출력 값 변화율을 나타낼 수 있다.
- 계산 전체에서 어떤 일이 벌어지든 상관 없이 자신과 직접 관계된 정보만으로 결과를 출력한다.
- $\frac{\partial E}{\partial A^{(3)}}, \frac{\partial A^{(3)}}{\partial A^{(2)}}, \frac{\partial A^{(2)}}{\partial W^{(2)}}$ 와 같은 계산 결과를 모두 보관할 수 있다.

$\frac{\partial A^{(2)}}{\partial W^{(2)}}$ : $W^{(2)}$에 대한 $A^{(2)}$의 변화율

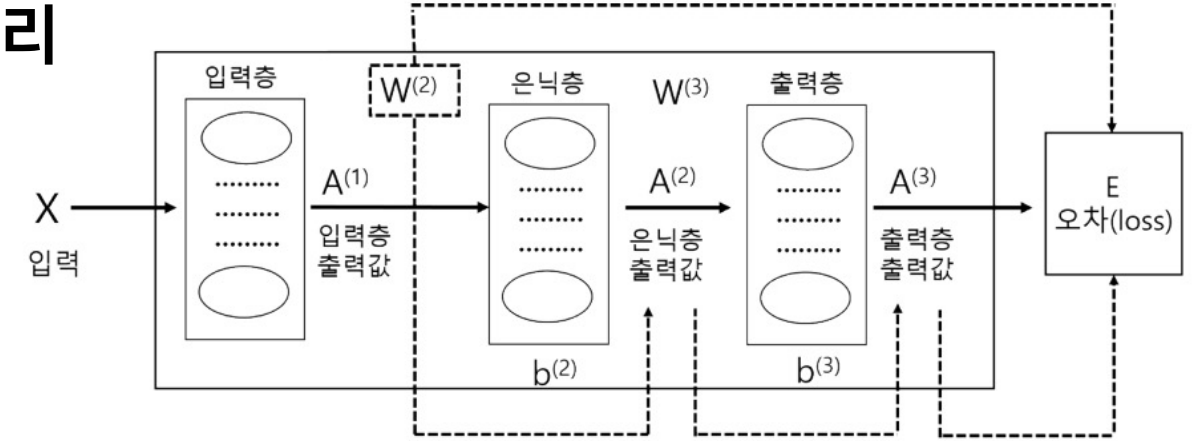
$\frac{\partial A^{(3)}}{\partial A^{(2)}}$ : $A^{(2)}$에 대한 $A^{(3)}$의 변화율

$\frac{\partial E}{\partial A^{(3)}}$ : $A^{(3)}$에 대한 $E$의 변화율

# 오차역전파 개념 및 동작 원리

## Chain Rule을 적용하여 수식 분해 - 정리

$A^{(1)}$: 입력층 출력 값
$A^{(2)}$: 은닉층 출력 값
$A^{(3)}$: 출력층 출력 값




- 가중치, 편향의 변화에 따른 오차가 얼마나 변하는지 나타내는 식을 Chain Rule을 통하여 다음과 같이 국소 미분으로 분리할 수 있다.

$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial A^{(2)}} \cdot \frac{\partial A^{(2)}}{\partial W^{(2)}} \qquad \frac{\partial E}{\partial W^{(3)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial W^{(3)}}$$

$$\frac{\partial E}{\partial b^{(2)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial A^{(2)}} \cdot \frac{\partial A^{(2)}}{\partial b^{(2)}} \qquad \frac{\partial E}{\partial b^{(3)}} = \frac{\partial E}{\partial A^{(3)}} \cdot \frac{\partial A^{(3)}}{\partial b^{(3)}}$$

- 이 국소 미분으로 분리된 식을 이용하여 최종적으로 수치 미분이 아닌 곱셈 형태의 산술 식으로 계산하여 오차역전파에 적용한다.

③ $\frac{\partial A^{(2)}}{\partial W^{(2)}}$ ② $\frac{\partial A^{(3)}}{\partial A^{(2)}}$ ① $\frac{\partial E}{\partial A^{(3)}}$

$\frac{\partial A^{(2)}}{\partial W^{(2)}}$ : $W^{(2)}$에 대한 $A^{(2)}$의 변화율

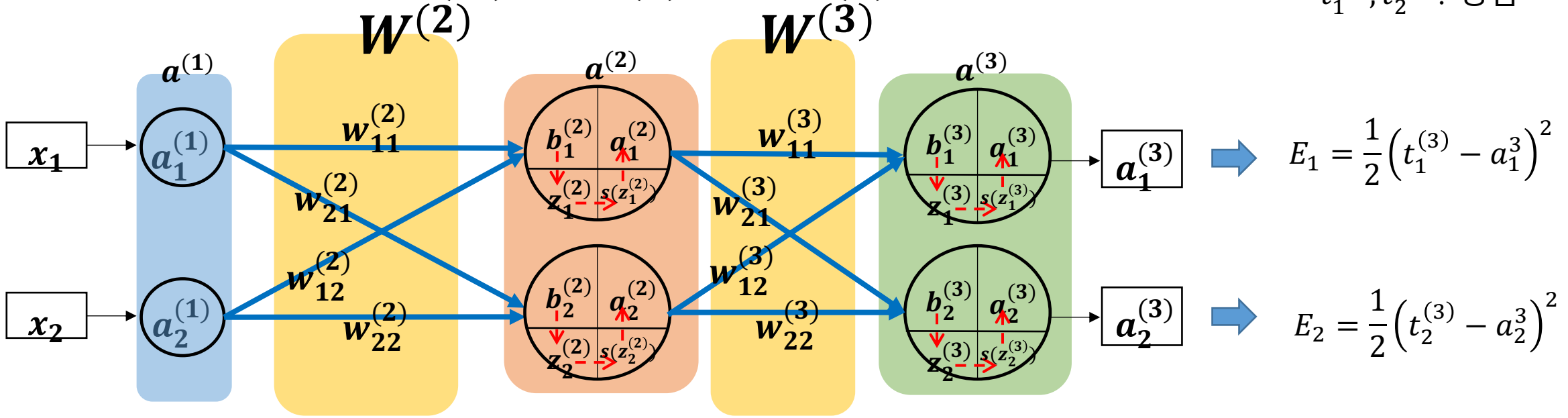
$\frac{\partial A^{(3)}}{\partial A^{(2)}}$ : $A^{(2)}$에 대한 $A^{(3)}$의 변화율

$\frac{\partial E}{\partial A^{(3)}}$ : $A^{(3)}$에 대한 $E$의 변화율

# 오차역전파 개념 및 동작 원리

## 각 층에서의 가중치($W$), 편향($b$)과 오차($E$)

$t_1^{(3)}, t_2^{(3)}$ : 정답




$$E_1 = \frac{1}{2}\left(t_1^{(3)} - a_1^3\right)^2$$

$$E_2 = \frac{1}{2}\left(t_2^{(3)} - a_2^3\right)^2$$

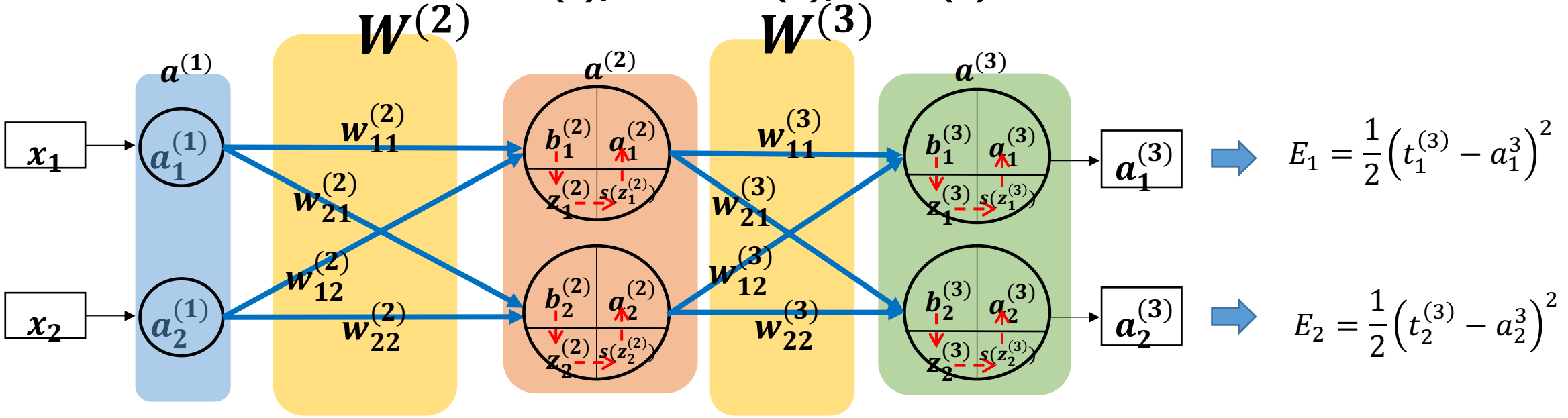
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(2)} = \begin{bmatrix} b_1^{(2)} & b_2^{(2)} \end{bmatrix}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$E = \frac{1}{2}\sum_{i=1}^{2}\left(t_i^{(3)} - a_i^{(3)}\right)^2 = \frac{1}{2}\left\{\left(t_1^{(3)} - a_1^{(3)}\right)^2 + \left(t_2^{(3)} - a_2^{(3)}\right)^2\right\} = E_1 + E_2$$

# 오차역전파 개념 및 동작 원리

## 각 층에서의 선형 회귀 값(z), 출력 값(a), 오차(E)




입력 층

$$a_1^{(1)} = x_1$$
$$a_2^{(1)} = x_2$$

은닉 층

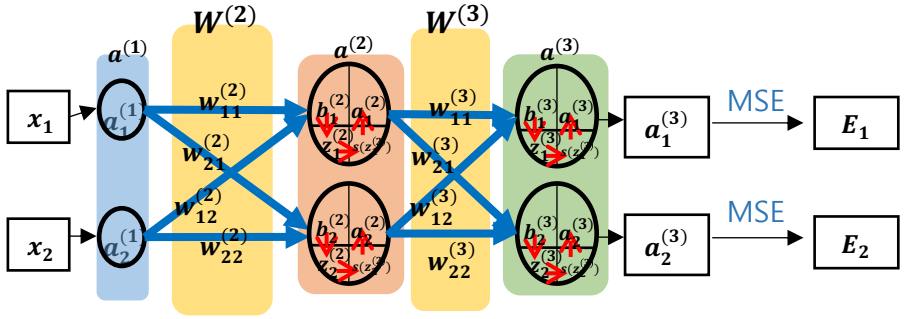
$$z_1^{(2)} = a_1^{(1)} w_{11}^{(2)} + a_2^{(1)} w_{12}^{(2)} + b_1^{(2)}$$
$$z_2^{(2)} = a_2^{(1)} w_{21}^{(2)} + a_2^{(1)} w_{22}^{(2)} + b_2^{(2)}$$
$$a_1^{(2)} = sigmoid\left(z_1^{(2)}\right)$$
$$a_2^{(2)} = sigmoid\left(z_2^{(2)}\right)$$

출력 층

$$z_1^{(3)} = a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}$$
$$z_2^{(3)} = a_2^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}$$
$$a_1^{(3)} = sigmoid\left(z_1^{(3)}\right)$$
$$a_2^{(3)} = sigmoid\left(z_2^{(3)}\right)$$

# 오차역전파 개념 및 동작 원리

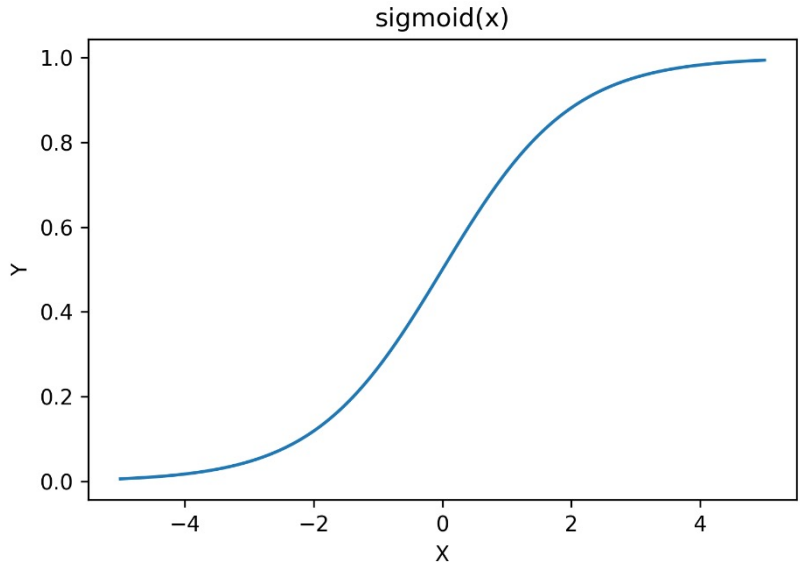
## Sigmoid 함수 미분

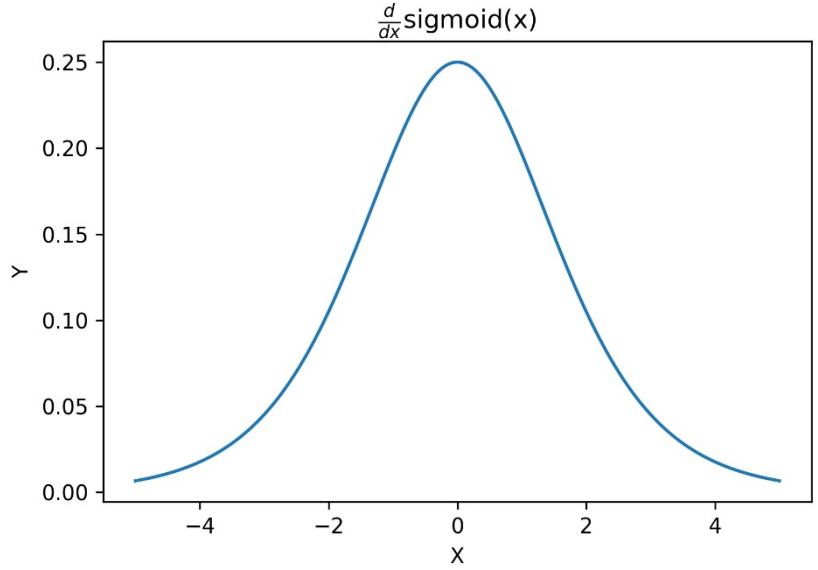



$$S(x) = \frac{1}{1+e^{-x}}$$

$$\frac{d}{dx}S(x) = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= S(x)\big(1 - S(x)\big)$$



# CONTENTS

# 출력 층에서의 오차역전파

## $w_{11}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(3)}}$

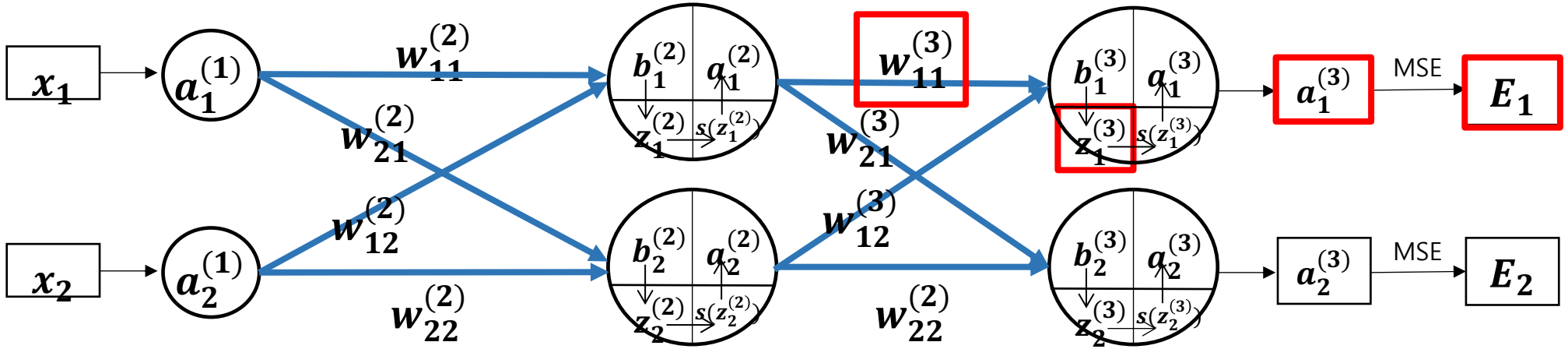
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}}, \quad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

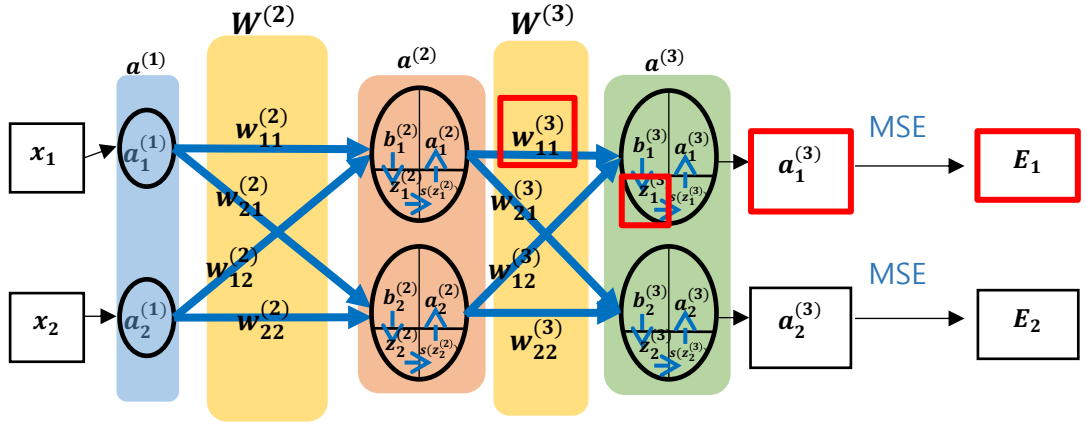
$$\frac{\partial E}{\partial w_{11}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $w_{11}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➤ $\frac{\partial E}{\partial W_{11}^{(3)}} = ?$

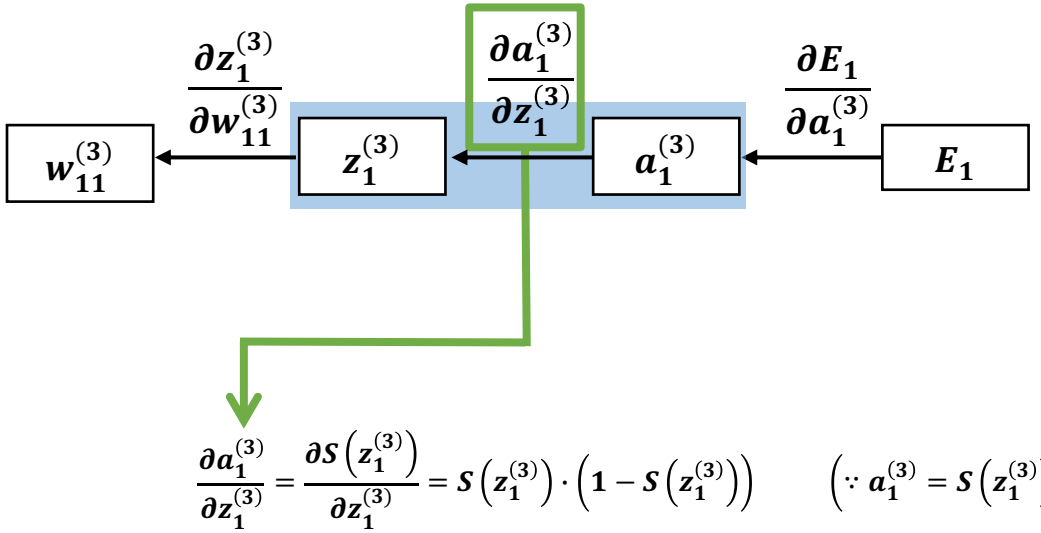
$$\frac{\partial E}{\partial w_{11}^{(3)}} = \frac{\partial E_1}{\partial w_{11}^{(3)}} + \frac{\partial E_2}{\partial w_{11}^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_1}{\partial w_{11}^{(3)}} \quad \left(\because \frac{\partial E_2}{\partial w_{11}^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}} \quad \cdots ③$$

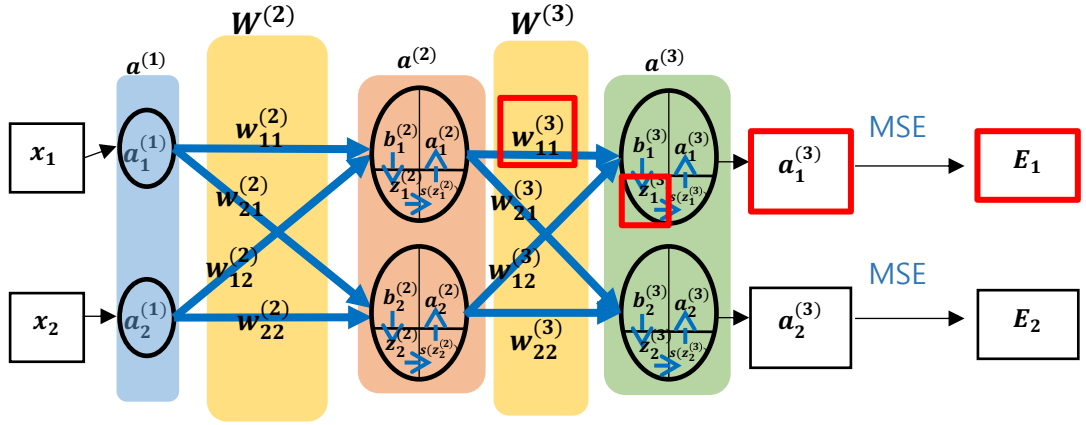
$$= \frac{\partial \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{11}^{(3)}} \quad \cdots ④$$

### Back Propagation



# 출력 층에서의 오차역전파

## $w_{11}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(3)}}$




➤ $\frac{\partial E}{\partial W_{11}^{(3)}} = ?$

$$= \frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{11}^{(3)}} \quad \cdots ④$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot \left(S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right)\right) \cdot a_1^{(2)} \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \cdot a_1^{(2)} \quad \cdots ⑥$$

$$\left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial W_{11}^{(3)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \cdot a_1^{(2)} \quad \cdots ⑦$$

- $\frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_1^{(3)} - t_1^{(3)}$
- $\frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{11}^{(3)}} = a_1^{(2)}$

# 출력 층에서의 오차역전파

## $w_{21}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{21}^{(3)}}$

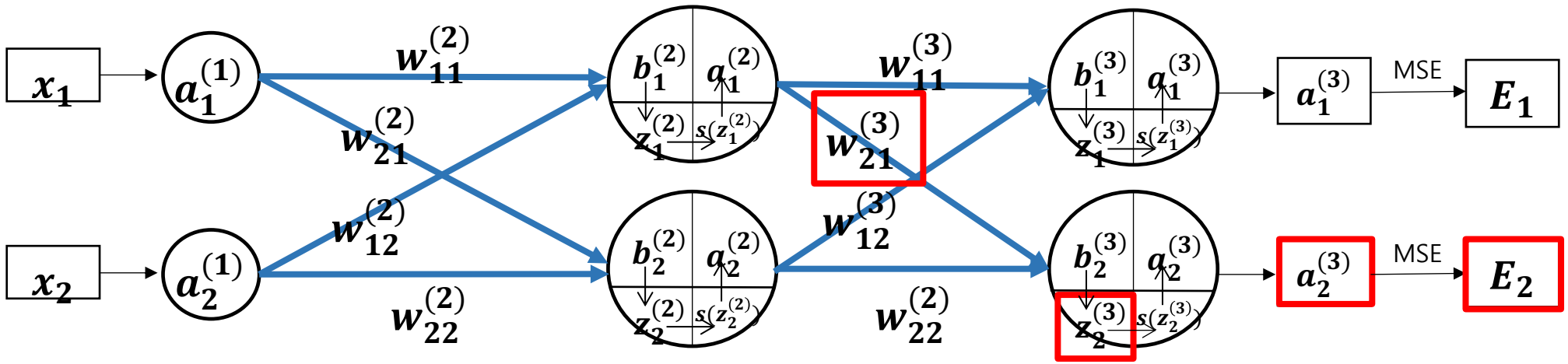
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}} \qquad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

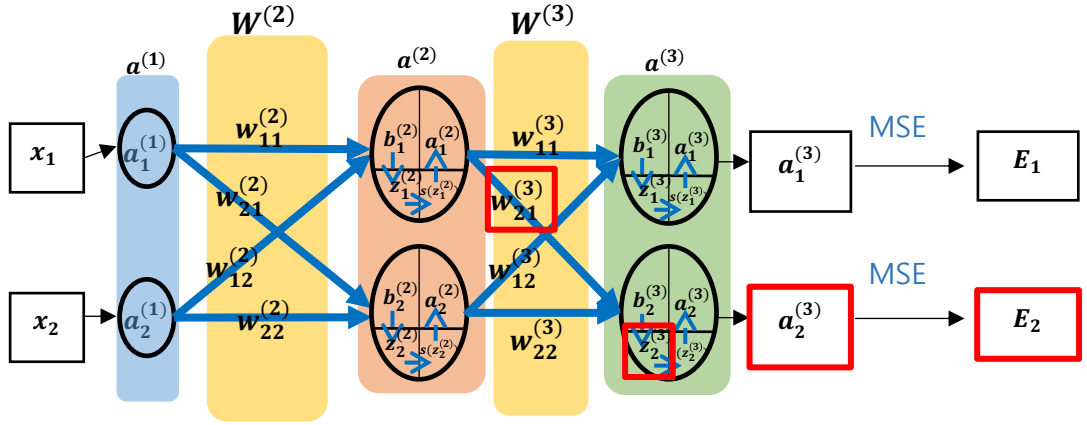
$$\frac{\partial E}{\partial w_{21}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $w_{21}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{21}^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➤ $\frac{\partial E}{\partial W_{21}^{(3)}} = ?$

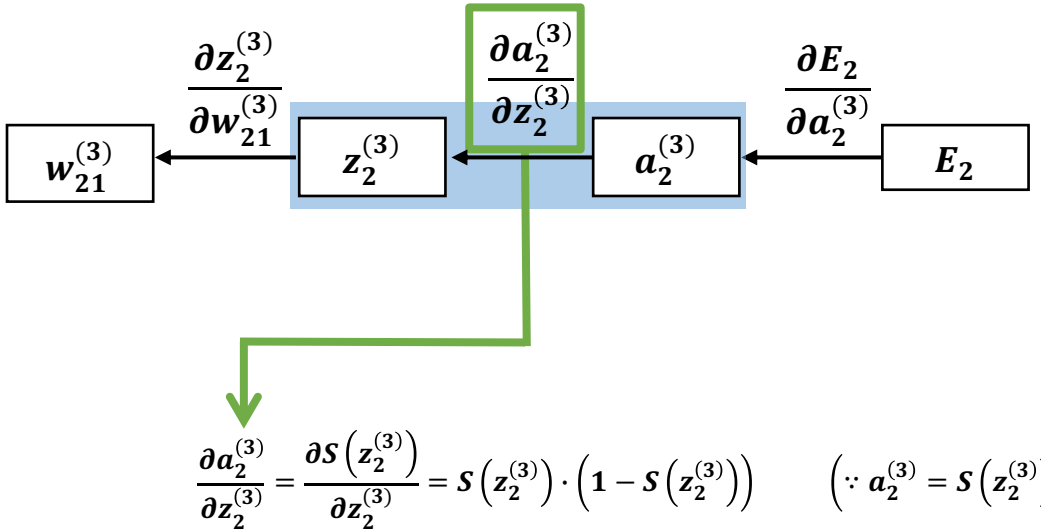
$$\frac{\partial E}{\partial w_{21}^{(3)}} = \frac{\partial E_1}{\partial w_{21}^{(3)}} + \frac{\partial E_2}{\partial w_{21}^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_2}{\partial w_{21}^{(3)}} \quad \left(\because \frac{\partial E_1}{\partial w_{21}^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial w_{21}^{(3)}} \quad \cdots ③$$

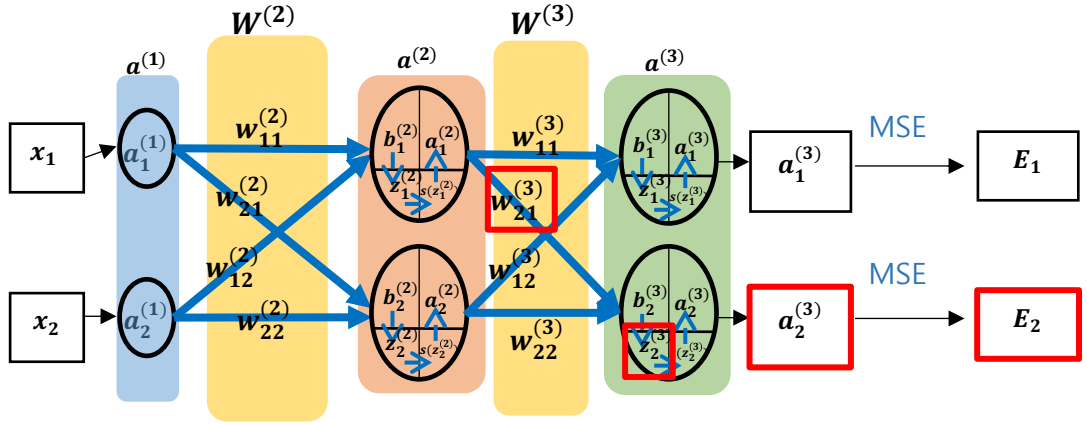
$$= \frac{\partial \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{21}^{(3)}} \quad \cdots ④$$

### Back Propagation



# 출력 층에서의 오차역전파

## $w_{21}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{21}^{(3)}}$




➤ $\frac{\partial E}{\partial W_{21}^{(3)}} = ?$

$$= \frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{21}^{(3)}} \quad \cdots ④$$

$$= \left(a_2^{(3)} - t_2^{(3)}\right) \cdot \left(S\left(z_2^{(3)}\right) \cdot \left(1 - S\left(z_2^{(3)}\right)\right)\right) \cdot a_1^{(2)} \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \cdot a_1^{(2)} \quad \cdots ⑥$$

$$\left(\because a_2^{(3)} = S\left(z_2^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial W_{21}^{(3)}} = \left(a_2^{(3)} - t_2^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \cdot a_1^{(2)} \quad \cdots ⑦$$

- $\frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_2^{(3)} - t_2^{(3)}$
- $\frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{21}^{(3)}} = a_1^{(2)}$

# 출력 층에서의 오차역전파

## $w_{12}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{12}^{(3)}}$

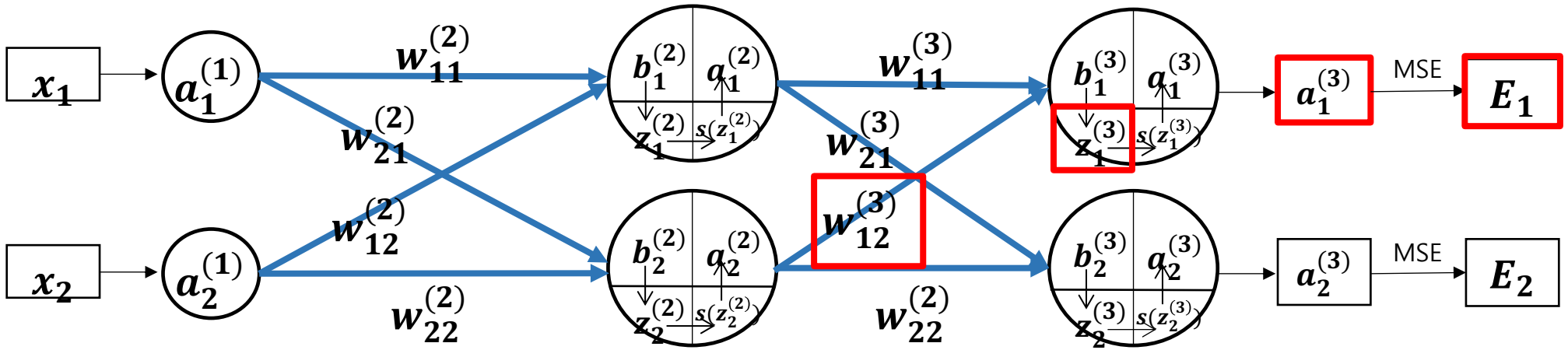
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}}, \quad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

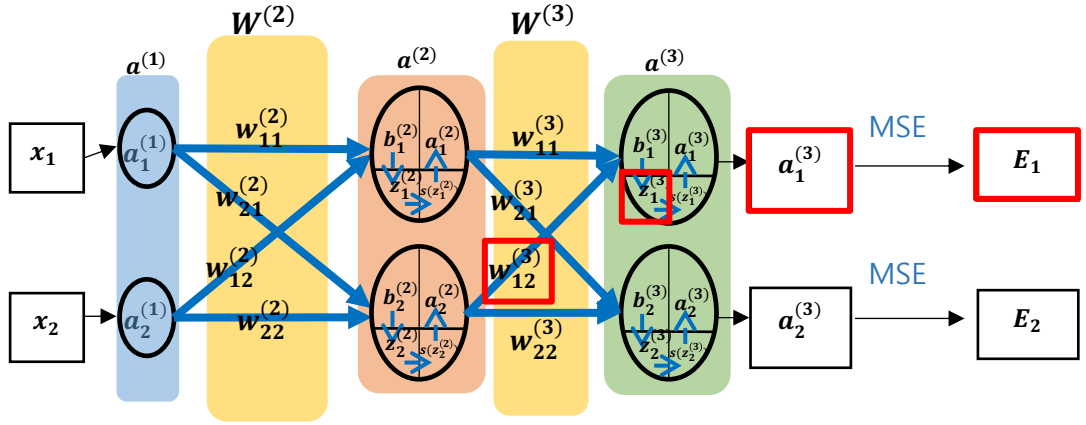
$$\frac{\partial E}{\partial w_{12}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $w_{12}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{12}^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➤ $\frac{\partial E}{\partial W_{12}^{(3)}} = ?$

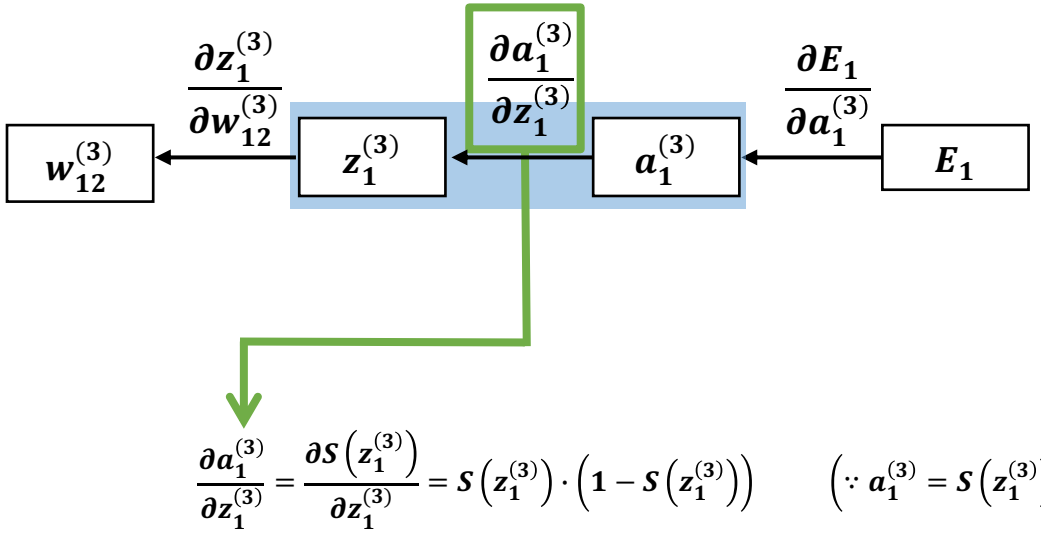
$$\frac{\partial E}{\partial w_{12}^{(3)}} = \frac{\partial E_1}{\partial w_{12}^{(3)}} + \frac{\partial E_2}{\partial w_{12}^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_1}{\partial w_{12}^{(3)}} \quad \left(\because \frac{\partial E_2}{\partial w_{12}^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial w_{12}^{(3)}} \quad \cdots ③$$

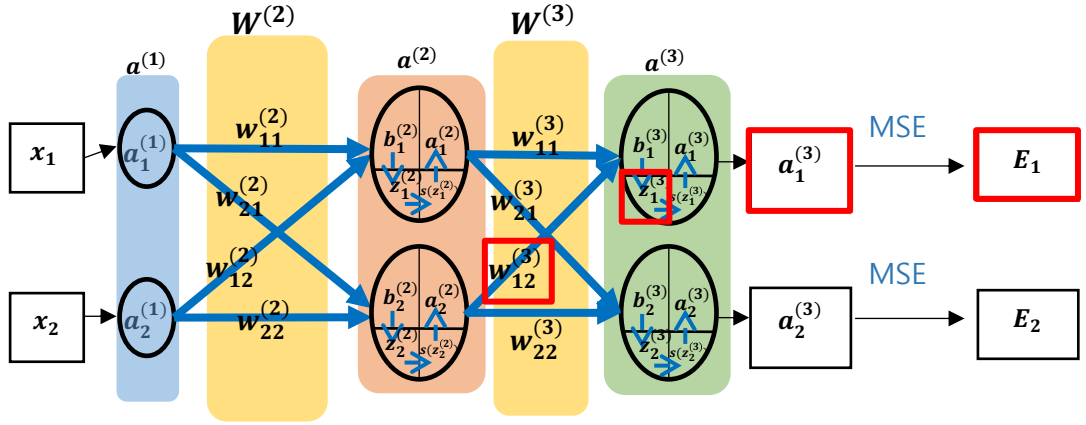
$$= \frac{\partial \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{12}^{(3)}} \quad \cdots ④$$

### Back Propagation



# 출력 층에서의 오차역전파

## $w_{12}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{12}^{(3)}}$




➤ $\frac{\partial E}{\partial W_{12}^{(3)}} = ?$

$$= \frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{12}^{(3)}} \quad \cdots ④$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot \left(S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right)\right) \cdot a_2^{(2)} \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \cdot a_2^{(2)} \quad \cdots ⑥$$

$$\left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial W_{12}^{(3)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \cdot a_2^{(2)} \quad \cdots ⑦$$

- $\frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_1^{(3)} - t_1^{(3)}$
- $\frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial w_{12}^{(3)}} = a_2^{(2)}$

# 출력 층에서의 오차역전파

## $w_{22}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{22}^{(3)}}$

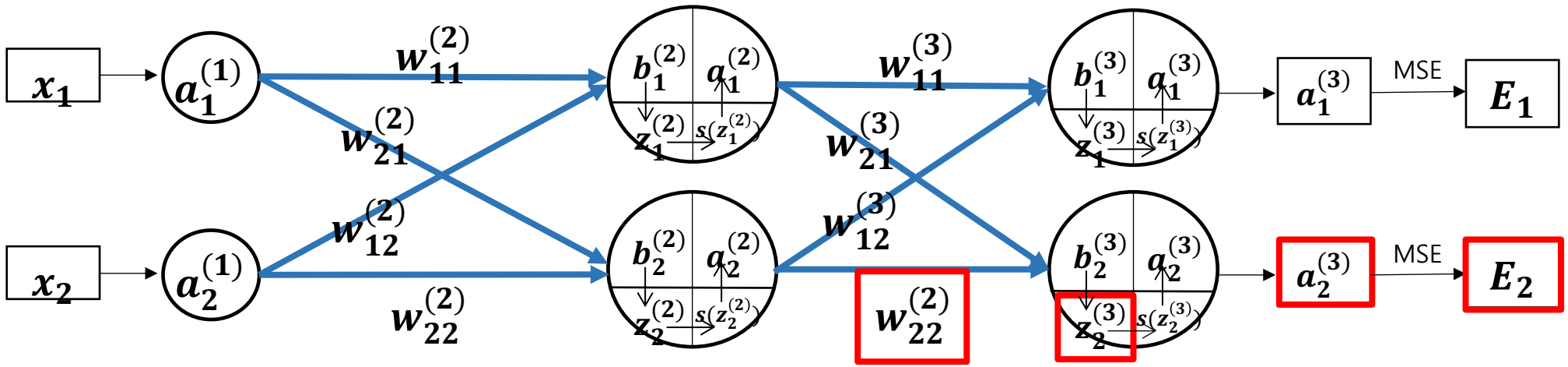
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}}$$
$$\frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

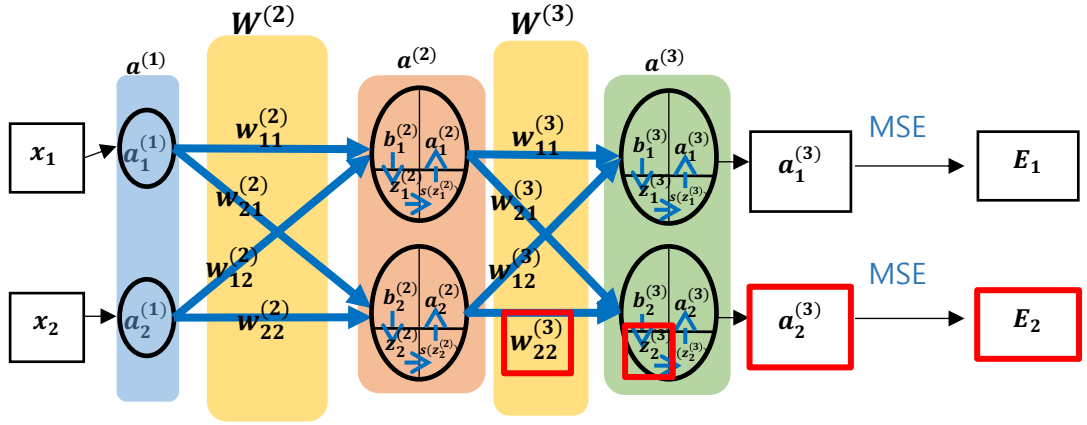
$$\frac{\partial E}{\partial w_{22}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $w_{22}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{22}^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➤ $\frac{\partial E}{\partial W_{22}^{(3)}} = ?$

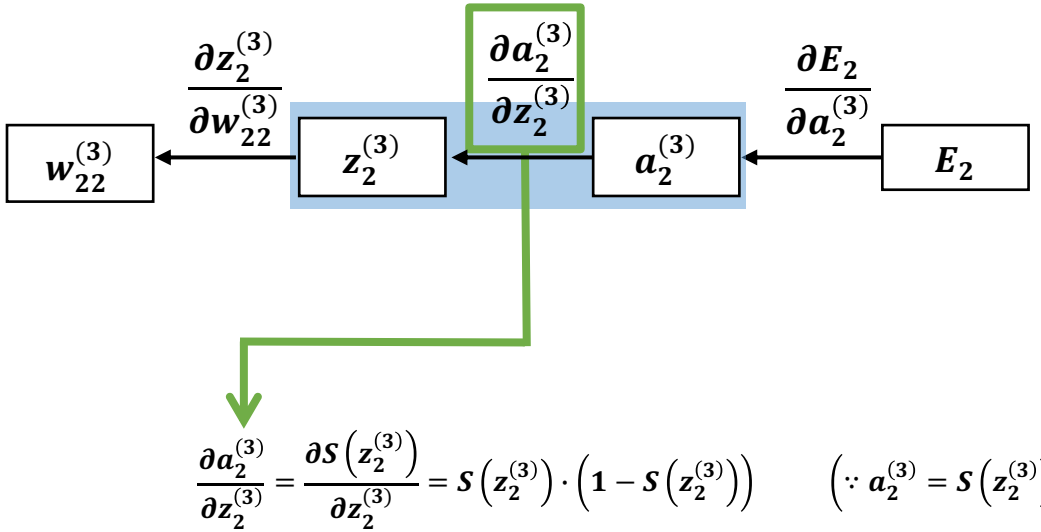
$$\frac{\partial E}{\partial w_{22}^{(3)}} = \frac{\partial E_1}{\partial w_{22}^{(3)}} + \frac{\partial E_2}{\partial w_{22}^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_2}{\partial w_{22}^{(3)}} \quad \left(\because \frac{\partial E_1}{\partial w_{22}^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial w_{22}^{(3)}} \quad \cdots ③$$

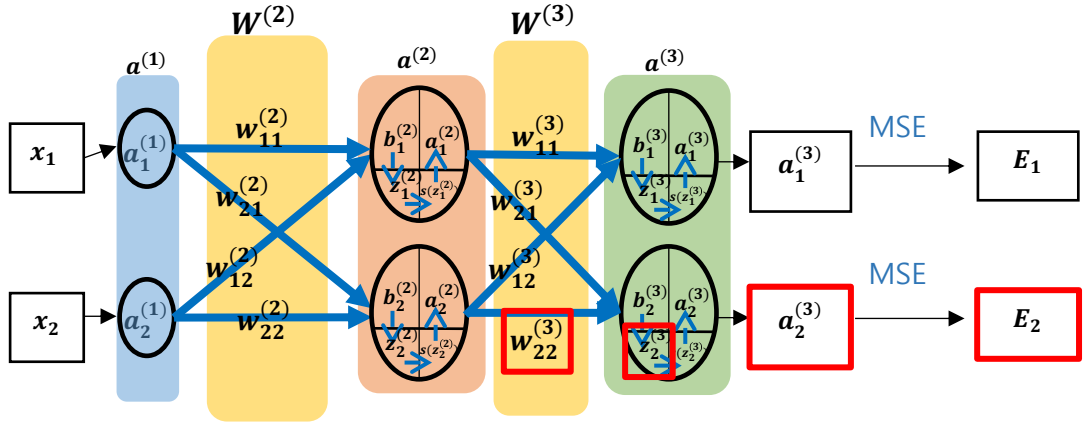
$$= \frac{\partial \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{22}^{(3)}} \quad \cdots ④$$

### Back Propagation



# 출력 층에서의 오차역전파

## $w_{22}^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{22}^{(3)}}$




➤ $\frac{\partial E}{\partial W_{22}^{(3)}} = ?$

$$= \frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{21}^{(3)}} \quad \cdots ④$$

$$= \left(a_2^{(3)} - t_2^{(3)}\right) \cdot \left(S\left(z_2^{(3)}\right) \cdot \left(1 - S\left(z_2^{(3)}\right)\right)\right) \cdot a_2^{(2)} \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \cdot a_2^{(2)} \quad \cdots ⑥$$

$$\left(\because a_2^{(3)} = S\left(z_2^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial W_{22}^{(3)}} = \left(a_2^{(3)} - t_2^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \cdot a_2^{(2)} \quad \cdots ⑦$$

- $\frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_2^{(3)} - t_2^{(3)}$
- $\frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial w_{22}^{(3)}} = a_2^{(2)}$

# 출력 층에서의 오차역전파

## $b_1^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_1^{(3)}}$

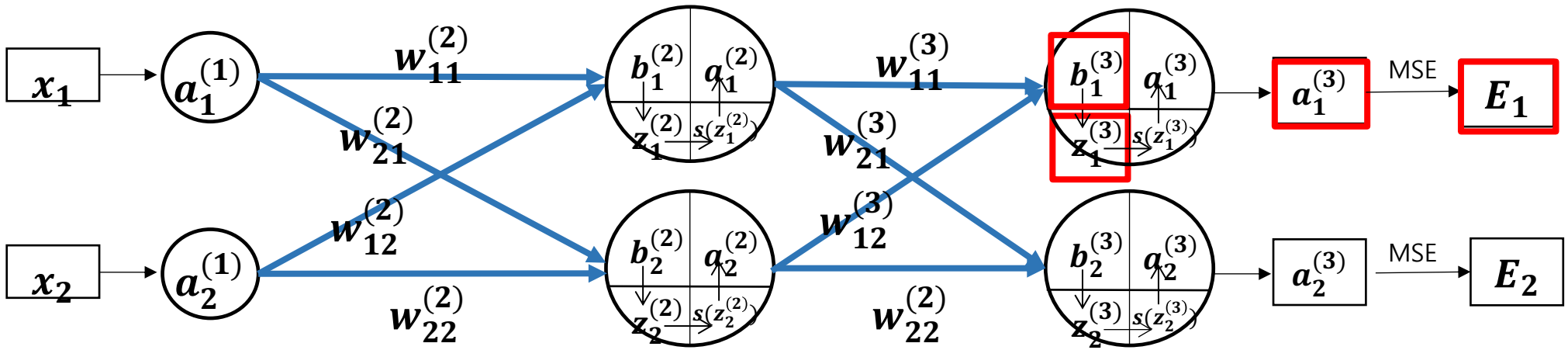
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}}, \quad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

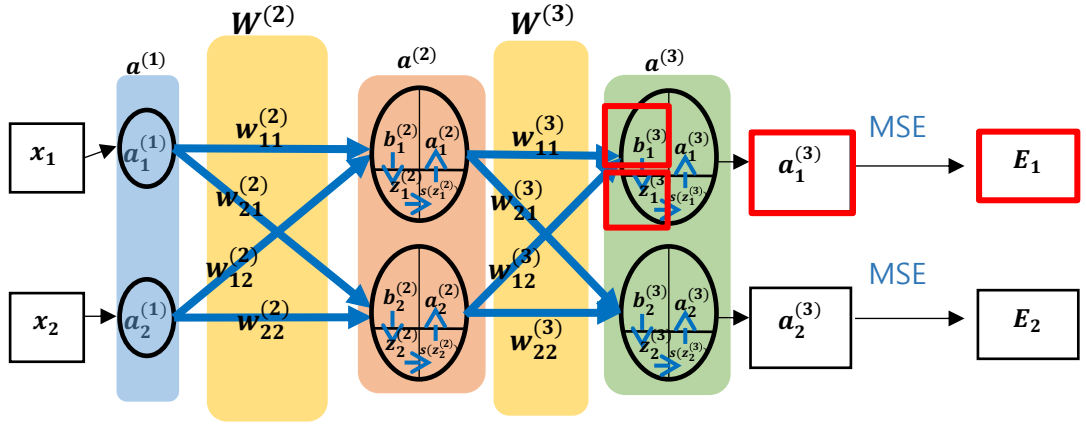
$$\frac{\partial E}{\partial w_{21}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $b_1^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_1^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➤ $\frac{\partial E}{\partial b_1^{(3)}} = ?$

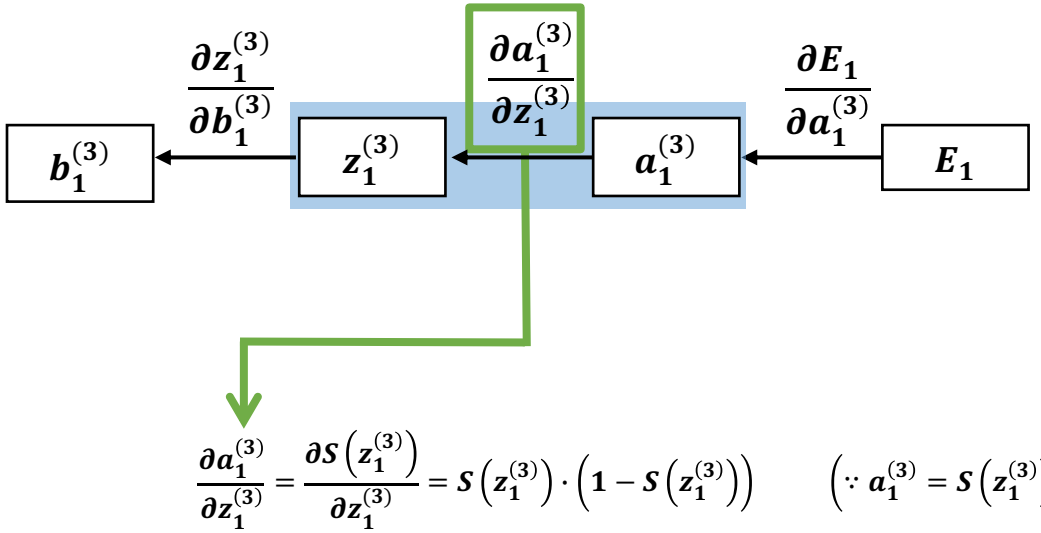
$$\frac{\partial E}{\partial b_1^{(3)}} = \frac{\partial E_1}{\partial b_1^{(3)}} + \frac{\partial E_2}{\partial b_1^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_1}{\partial b_1^{(3)}} \quad \left(\because \frac{\partial E_2}{\partial b_1^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial b_1^{(3)}} \quad \cdots ③$$

$$= \frac{\partial \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial b_1^{(3)}} \quad \cdots ④$$

### Back Propagation



# 출력 층에서의 오차역전파

## $b_1^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_1^{(3)}}$

➢ $\frac{\partial E}{\partial b_1^{(3)}} = ?$

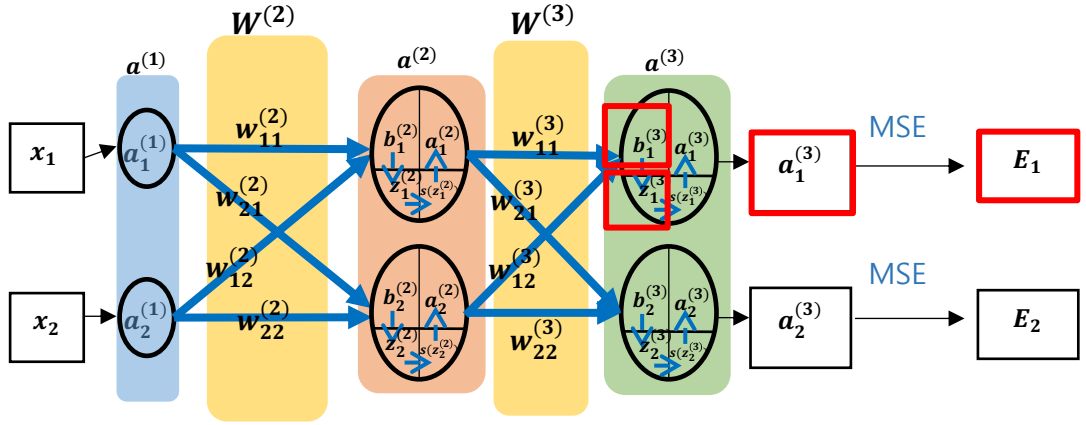
$$= \frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} \cdot \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial b_1^{(3)}} \quad \cdots ④$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot \left(S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right)\right) \cdot 1 \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \cdot 1 \quad \cdots ⑥$$

$$\left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial b_1^{(3)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)} \cdot \left(1 - a_1^{(3)}\right) \quad \cdots ⑦$$




- $\frac{\partial\left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right)}{\partial a_1^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_1^{(3)} - t_1^{(3)}$
- $\frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial b_1^{(3)}} = 1$

# 출력 층에서의 오차역전파

## $b_2^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_2^{(3)}}$

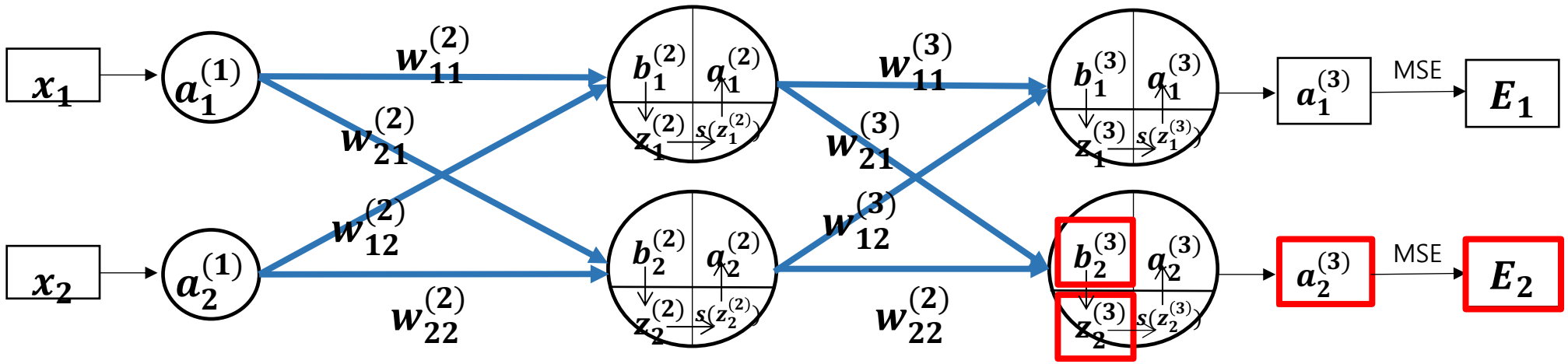
$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}}$$
$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}}$$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(3)}} \quad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$

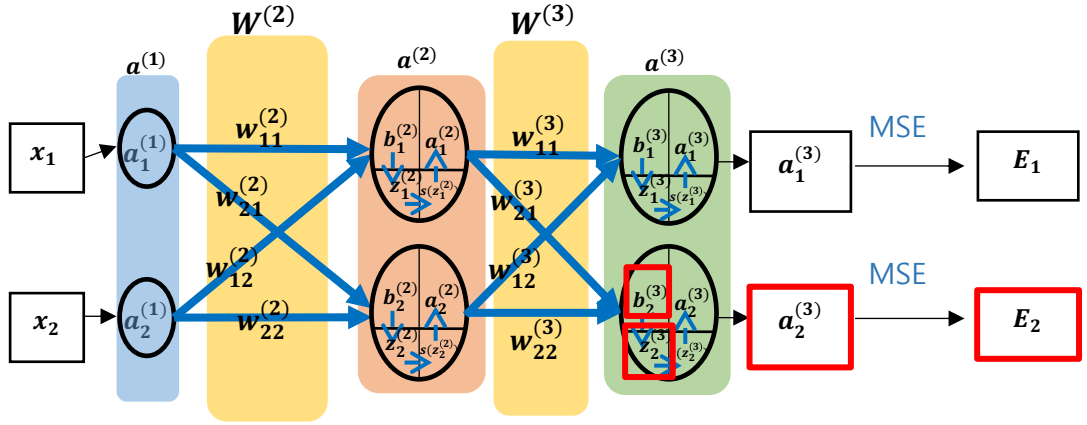
$$\frac{\partial E}{\partial w_{21}^{(3)}} = ?$$



# 출력 층에서의 오차역전파

## $b_1^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_1^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix}$$




➢ $\frac{\partial E}{\partial b_2^{(3)}} = ?$

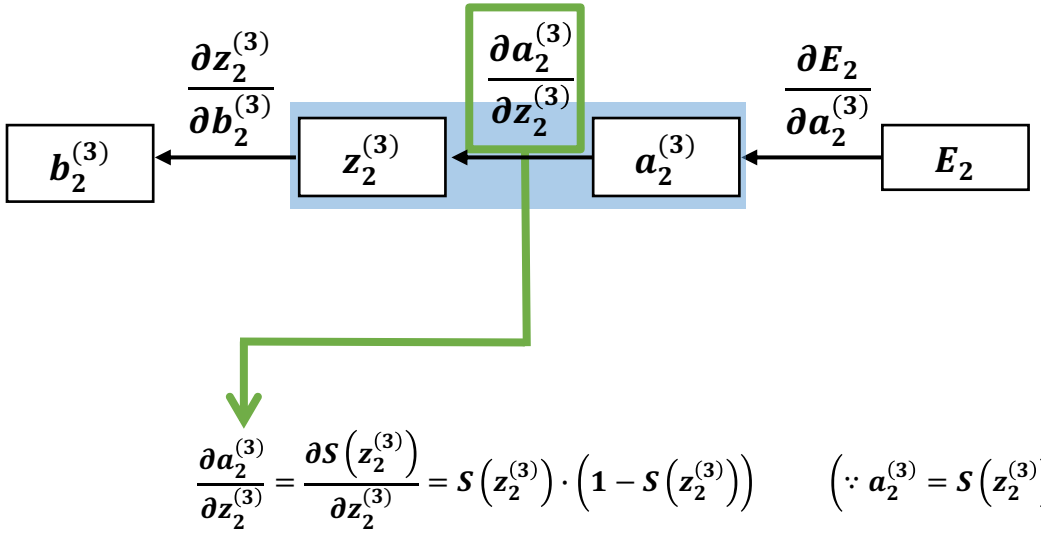
$$\frac{\partial E}{\partial b_2^{(3)}} = \frac{\partial E_1}{\partial b_1^{(3)}} + \frac{\partial E_2}{\partial b_1^{(3)}} \quad (\because E = E_1 + E_2) \quad \cdots ①$$

$$= \frac{\partial E_2}{\partial b_2^{(3)}} \quad \left(\because \frac{\partial E_1}{\partial b_2^{(3)}} = 0\right) \quad \cdots ②$$

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial b_2^{(3)}} \quad \cdots ③$$

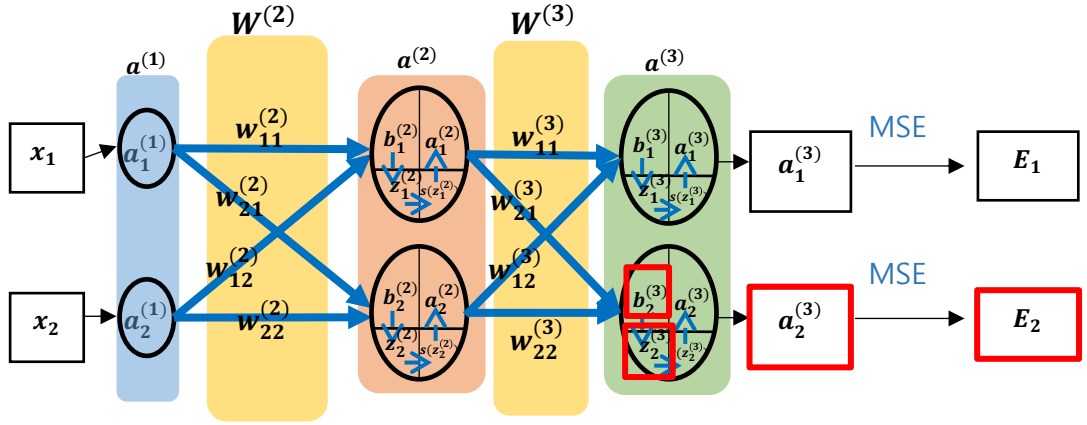
$$= \frac{\partial \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial \left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial b_2^{(3)}} \quad \cdots ④$$

Back Propagation



# 출력 층에서의 오차역전파

## $b_1^{(3)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial b_1^{(3)}}$




➤ $\frac{\partial E}{\partial b_2^{(3)}} = ?$

$$= \frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} \cdot \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} \cdot \frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial b_2^{(3)}} \quad \cdots ④$$

$$= \left(a_2^{(3)} - t_2^{(3)}\right) \cdot \left(S\left(z_2^{(3)}\right) \cdot \left(1 - S\left(z_2^{(3)}\right)\right)\right) \cdot 1 \quad \cdots ⑤$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \cdot 1 \quad \cdots ⑥$$

$$\left(\because a_2^{(3)} = S\left(z_2^{(3)}\right)\right)$$

$$\therefore \frac{\partial E}{\partial b_2^{(3)}} = \left(a_2^{(3)} - t_2^{(3)}\right) \cdot a_2^{(3)} \cdot \left(1 - a_2^{(3)}\right) \quad \cdots ⑦$$

- $\frac{\partial\left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right)}{\partial a_2^{(3)}} = 2 \times \left(\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)\right) \times \underbrace{(-1)}_{\text{속미분}} = a_2^{(3)} - t_2^{(3)}$
- $\frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right) \qquad \left(\because a_1^{(3)} = S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial\left(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}\right)}{\partial b_2^{(3)}} = 1$

# 출력 층에서의 오차역전파

## 출력 층에서의 오차역전파 - 정리

| | |
|---|---|
| 은닉층 출력 값 벡터 | $A2 = (a_1^{(2)} \quad a_2^{(2)})$ |
| 출력층 가상 손실 벡터 | $loss\_3 = ((a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) \quad (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}))$ |

- $$\frac{\partial E}{\partial W^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} = \begin{bmatrix} \left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) a_1^{(2)} & \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) a_1^{(2)} \\ \left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) a_2^{(2)} & \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) a_2^{(2)} \end{bmatrix}$$

$$= \begin{bmatrix} a_1^{(2)} \\ a_2^{(2)} \end{bmatrix} \begin{bmatrix} \left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) & \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) \end{bmatrix}$$

$$= A2^T \cdot loss\_3$$

- $$\frac{\partial E}{\partial b^{(3)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{bmatrix} = \begin{bmatrix} \left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) & \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) \end{bmatrix}$$

$$= loss\_3$$

# 출력 층에서의 오차역전파

## 출력 층에서의 오차역전파 - 정리

| | |
|---|---|
| 은닉층 출력 값 벡터 | $A2 = (a_1^{(2)} \quad a_2^{(2)})$ |
| 출력층 가상 손실 벡터 | $loss\_3 = ((a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) \quad (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}))$ |

- $\frac{\partial E}{\partial W^{(3)}} = A2^T \times loss\_3$
- $\frac{\partial E}{\partial b^{(3)}} = loss\_3$

➢ 출력층 가중치 계산 :

$$W^{(3)} := W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}} = W^{(3)} - \alpha(A2^T \cdot loss\_3)$$

➢ 출력층 편향 계산 :

$$b^{(3)} := b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}} = b^{(3)} - \alpha(loss\_3)$$

CONTENTS

# 은닉층에서의 오차역전파

## $w_{11}^{(2)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(2)}}$

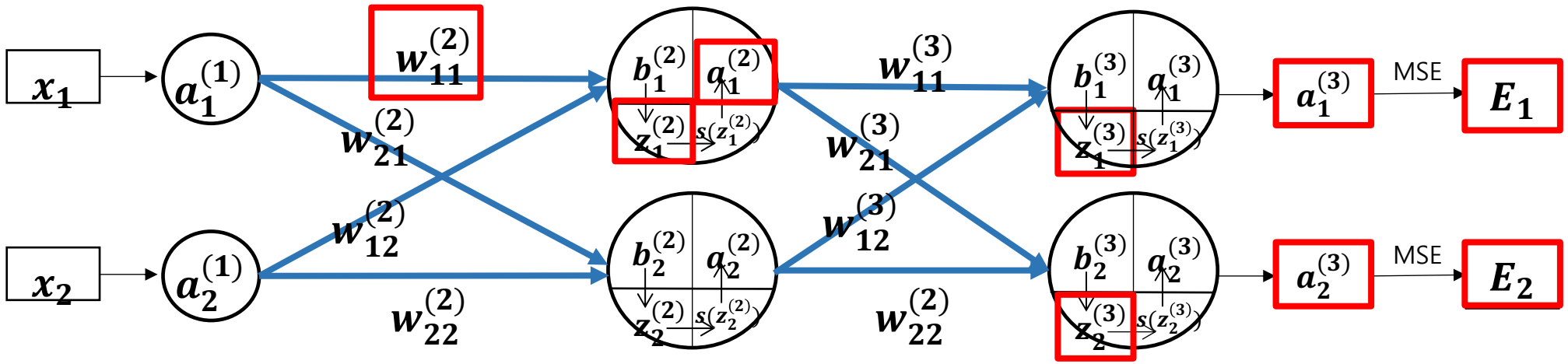
$$W^{(2)} := W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}}$$
$$b^{(2)} := b^{(2)} - \alpha \frac{\partial E}{\partial b^{(2)}}$$

$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{bmatrix} \qquad b^{(2)} = \begin{bmatrix} b_1^{(2)} & b_2^{(2)} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{(2)}}, \quad \frac{\partial E}{\partial b^{(3)}}$$

$$\frac{\partial E}{\partial W^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(2)}} & \frac{\partial E}{\partial w_{21}^{(2)}} \\ \frac{\partial E}{\partial w_{12}^{(2)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(2)}} & \frac{\partial E}{\partial b_2^{(2)}} \end{bmatrix}$$

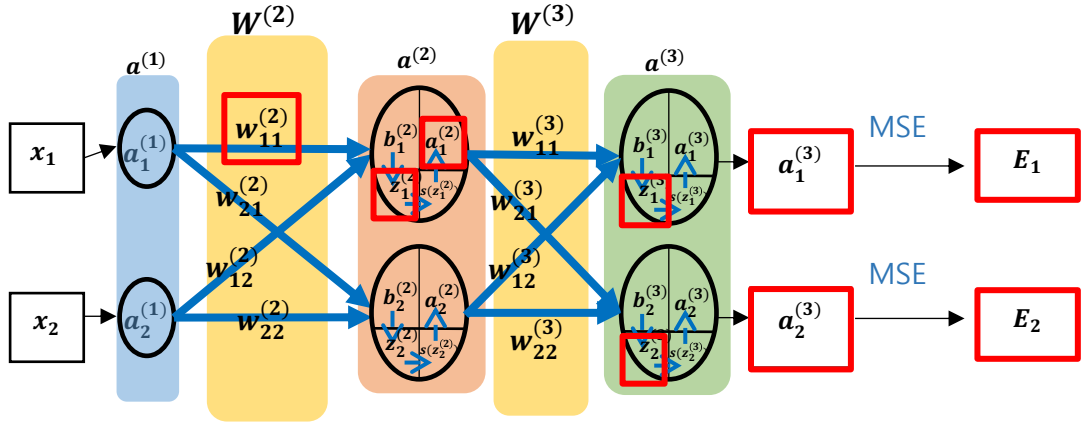
$$\frac{\partial E}{\partial w_{11}^{(2)}} = ?$$



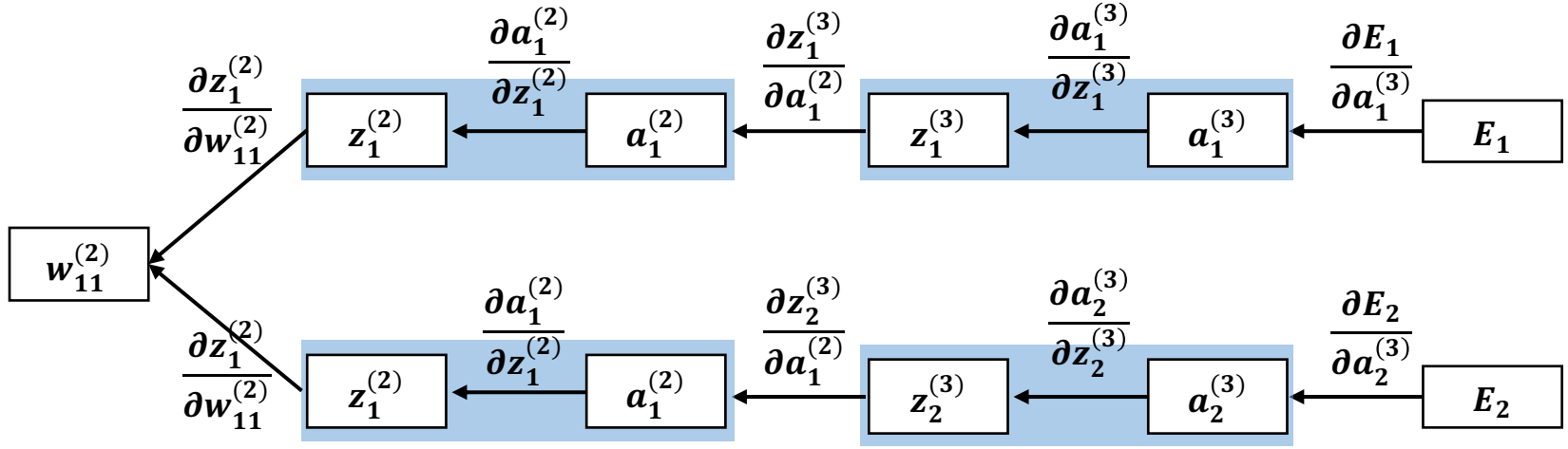
# 은닉층에서의 오차역전파

## $w_{11}^{(2)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(2)}}$

$$\frac{\partial E}{\partial W^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(2)}} & \frac{\partial E}{\partial w_{21}^{(2)}} \\ \frac{\partial E}{\partial w_{12}^{(2)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(2)}} & \frac{\partial E}{\partial b_2^{(2)}} \end{bmatrix}$$



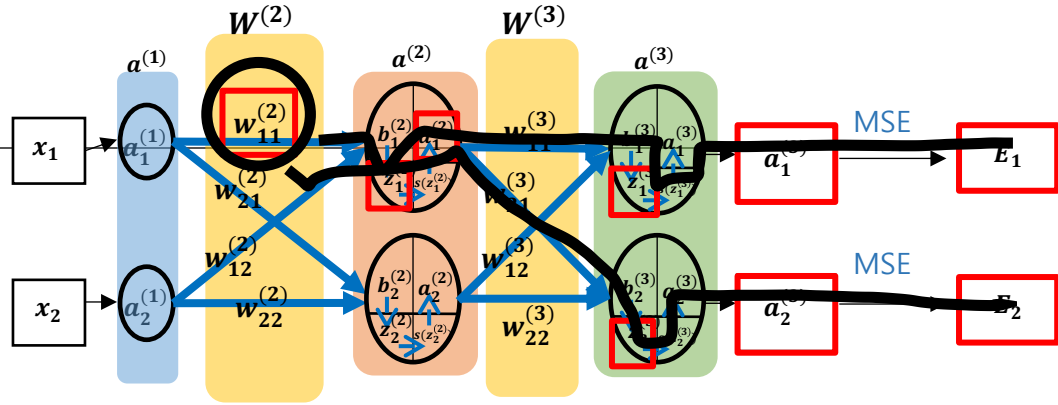

Back Propagation




$$\frac{\partial E}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}}$$

# 은닉층에서의 오차역전파

## $w_{11}^{(2)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(2)}}$




$$\frac{\partial E}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}}$$

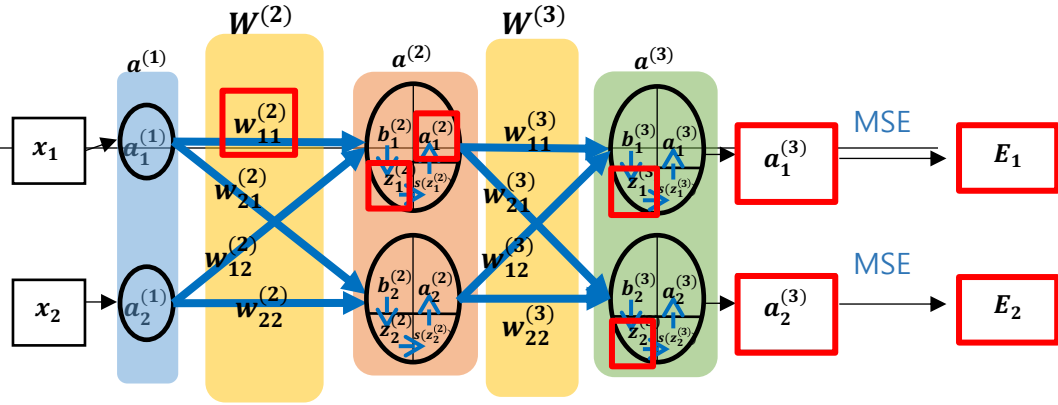
$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot S\left(z_1^{(3)}\right)\left(1 - S\left(z_1^{(3)}\right)\right) \cdot w_{11}^{(3)} \cdot S\left(z_1^{(2)}\right)\left(1 - S\left(z_1^{(2)}\right)\right) \cdot a_1^{(1)}$$
$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \cdot S\left(z_2^{(3)}\right)\left(1 - S\left(z_2^{(3)}\right)\right) \cdot w_{21}^{(3)} \cdot S\left(z_1^{(2)}\right)\left(1 - S\left(z_1^{(2)}\right)\right) \cdot a_1^{(1)}$$

- $\frac{\partial E_1}{\partial a_1^{(3)}} = \frac{\partial\left\{\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right\}}{\partial a_1^{(3)}} = \left(a_1^{(3)} - t_1^{(3)}\right)$
- $\frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} = \frac{\partial S\left(z_1^{(3)}\right)}{\partial z_1^{(3)}} = S\left(z_1^{(3)}\right) \cdot \left(1 - S\left(z_1^{(3)}\right)\right)$
- $\frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} = \frac{\partial\left(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}\right)}{\partial a_1^{(2)}} = w_{11}^{(3)}$
- $\frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = \frac{\partial S\left(z_1^{(2)}\right)}{\partial z_1^{(2)}} = S\left(z_1^{(2)}\right) \cdot \left(1 - S\left(z_1^{(2)}\right)\right)$
- $\frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} = \frac{\partial\left(a_1^{(1)} w_{11}^{(2)} + a_2^{(1)} w_{12}^{(2)}\right)}{\partial w_{11}^{(2)}} = a_1^{(1)}$

- $\frac{\partial E_2}{\partial a_2^{(3)}} = \frac{\partial\left\{\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right\}}{\partial a_2^{(3)}} = \left(a_2^{(3)} - t_2^{(3)}\right)$
- $\frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} = \frac{\partial S\left(z_2^{(3)}\right)}{\partial z_2^{(3)}} = S\left(z_2^{(3)}\right) \cdot \left(1 - S\left(z_2^{(3)}\right)\right)$
- $\frac{\partial z_2^{(3)}}{\partial a_1^{(2)}} = \frac{\partial\left(a_1^{(2)} w_{21} + a_2^{(2)} w_{22} + b_2\right)}{\partial a_1^{(2)}} = w_{21}$
- $\frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = \frac{\partial S\left(z_1^{(2)}\right)}{\partial z_1^{(2)}} = S\left(z_1^{(2)}\right) \cdot \left(1 - S\left(z_1^{(2)}\right)\right)$
- $\frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} = \frac{\partial\left(a_1^{(1)} w_{11}^{(2)} + a_2^{(1)} w_{12}^{(2)}\right)}{\partial w_{11}^{(2)}} = a_1^{(1)}$

# 은닉층에서의 오차역전파

## $w_{11}^{(2)}$에 대한 $E$의 변화율 - $\frac{\partial E}{\partial w_{11}^{(2)}}$




$$\frac{\partial E}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}}$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot S\left(z_1^{(3)}\right)\left(1 - S\left(z_1^{(3)}\right)\right) \cdot w_{11}^{(3)} \cdot S\left(z_1^{(2)}\right)\left(1 - S\left(z_1^{(2)}\right)\right) \cdot a_1^{(1)}$$
$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \cdot S\left(z_2^{(3)}\right)\left(1 - S\left(z_2^{(3)}\right)\right) \cdot w_{21}^{(3)} \cdot S\left(z_1^{(2)}\right)\left(1 - S\left(z_1^{(2)}\right)\right) \cdot a_1^{(1)}$$

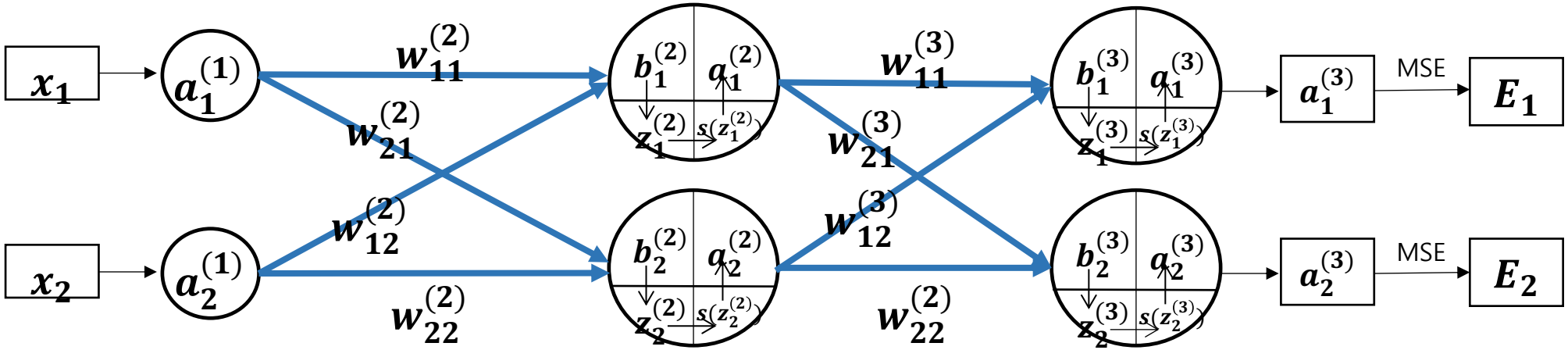
$$= \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)}\left(1 - a_1^{(3)}\right) \cdot w_{11}^{(3)} \cdot a_1^{(2)}\left(1 - a_1^{(2)}\right) \cdot a_1^{(1)}$$
$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \cdot a_2^{(3)}\left(1 - a_2^{(3)}\right) \cdot w_{21}^{(3)} \cdot a_1^{(2)}\left(1 - a_1^{(2)}\right) \cdot a_1^{(1)}$$

$$\therefore \frac{\partial E}{\partial w_{11}^{(2)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \cdot a_1^{(3)}\left(1 - a_1^{(3)}\right) \cdot w_{11}^{(3)} \cdot a_1^{(2)}\left(1 - a_1^{(2)}\right) \cdot a_1^{(1)} + \left(a_2^{(3)} - t_2^{(3)}\right) \cdot a_2^{(3)}\left(1 - a_2^{(3)}\right) \cdot w_{21}^{(3)} \cdot a_1^{(2)}\left(1 - a_1^{(2)}\right) \cdot a_1^{(1)}$$

# 은닉층에서의 오차역전파

## Do It Yourself $\frac{\partial E}{\partial w_{21}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \frac{\partial E}{\partial w_{22}^{(2)}}, \frac{\partial E}{\partial b_1^{(2)}}, \frac{\partial E}{\partial b_2^{(2)}}$

$$\frac{\partial E}{\partial W^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(2)}} & \frac{\partial E}{\partial w_{21}^{(2)}} \\ \frac{\partial E}{\partial w_{12}^{(2)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} \qquad \frac{\partial E}{\partial b^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(2)}} & \frac{\partial E}{\partial b_2^{(2)}} \end{bmatrix}$$



# 은닉층에서의 오차역전파

## 은닉층에서의 오차역전파 - 정리

| | |
|---|---|
| 입력층 출력 값 벡터 | $A1 = (a_1^{(1)} \quad a_2^{(1)})$ |
| 은닉층 출력 값 벡터 | $A2 = (a_1^{(2)} \quad a_2^{(2)})$ |
| 출력층 가상 손실 벡터 | $loss\_3 = ((a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) \quad (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}))$ |
| 출력층 가중치 | $W3 = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$ |
| 은닉층 가상 손실 벡터 | $loss\_2 = (loss\_3 \cdot W3^T) \times A2(1 - A2)$ |

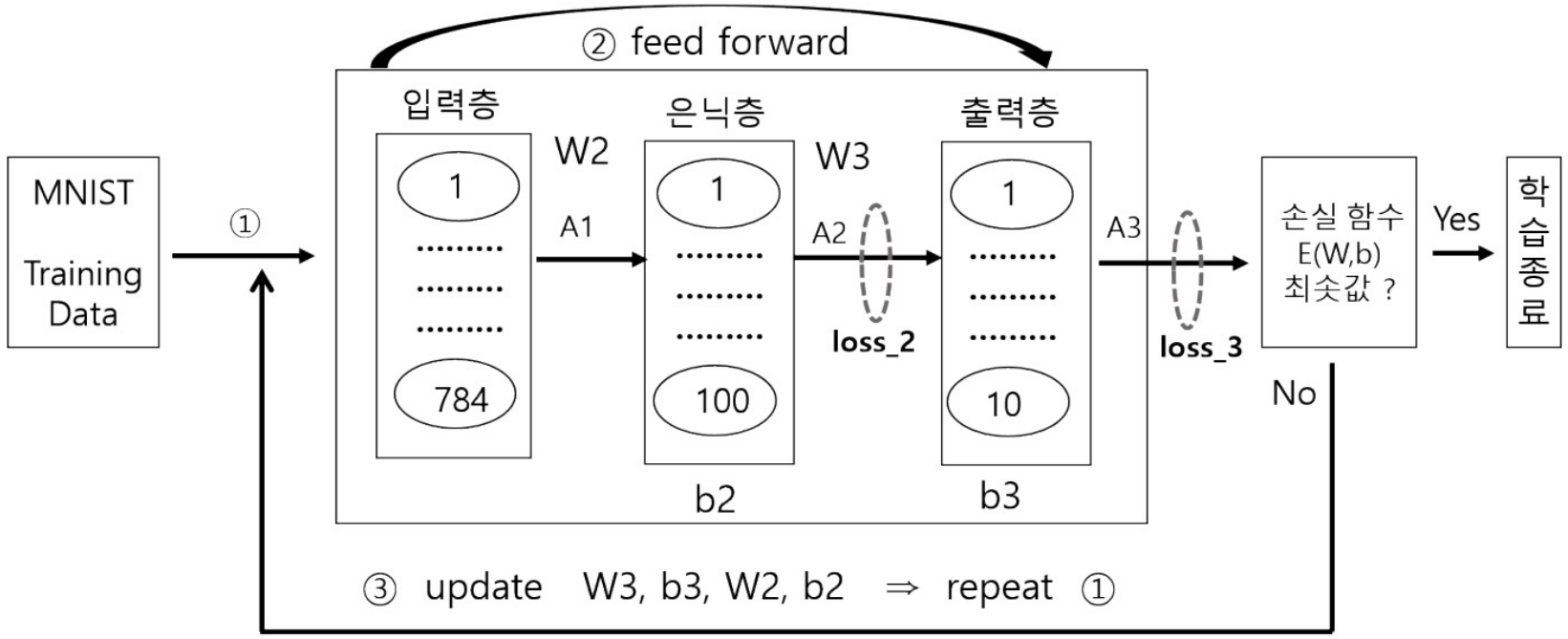
- $\frac{\partial E}{\partial W^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial w_{11}^{(2)}} & \frac{\partial E}{\partial w_{21}^{(2)}} \\ \frac{\partial E}{\partial w_{12}^{(2)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{bmatrix} = A1^T \cdot \left((loss\_3 \cdot W3^T) \times (A2 \times (1 - A2))\right) = A1^T \cdot loss\_2$

- $\frac{\partial E}{\partial b^{(2)}} = \begin{bmatrix} \frac{\partial E}{\partial b_1^{(2)}} & \frac{\partial E}{\partial b_2^{(2)}} \end{bmatrix} = \left((loss\_3 \cdot W3^T) \times (A2 \times (1 - A2))\right) = loss\_2$

# CONTENTS

# 오차역전파를 이용한 MNIST 분류

## 오차역전파를 이용한 딥러닝 architecture




- $W^{(3)} := W^{(3)} - \alpha\left(A^{(2)^T} \cdot loss\_3\right)$
- $W^{(2)} := W^{(2)} - \alpha\left(A^{(2)^T} \cdot loss\_2\right)$
- $b^{(3)} := b^{(3)} - \alpha \times loss\_3$
- $b^{(2)} := b^{(2)} - \alpha \times loss\_2$

- $A^{(1)} = \begin{bmatrix} a_1^{(1)} & a_2^{(1)} \end{bmatrix}$
- $A^{(2)} = \begin{bmatrix} a_1^{(2)} & a_2^{(2)} \end{bmatrix}$
- $loss\_3 = \begin{bmatrix} \left(a_1^{(3)} - t_1\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) & \left(a_2^{(3)} - t_2\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) \end{bmatrix}$
- $loss\_2 = \left(loss\_3 \cdot W^{(3)^T}\right) \times \left(A^{(2)}\left(1 - A^{(2)}\right)\right)$

# 오차역전파를 이용한 MNIST 분류

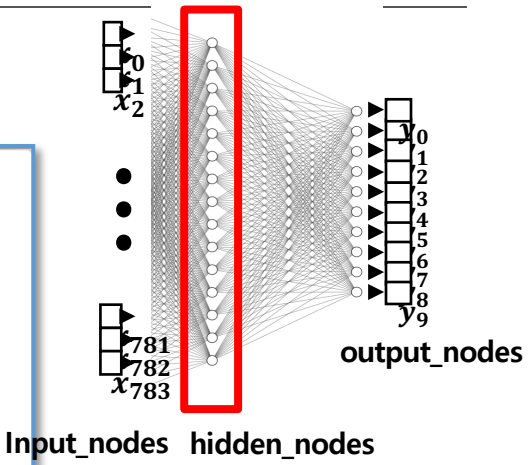
## CODE – NeuralNetwork class

```
class NeuralNetwork:
    def __init__(self, input_nodes, hidden_nodes, output_nodes, learning_rate):
    def feed_forward(self):                          #Feed Forward 수행
    def loss_val(self):                              #손실함수 값 계산
    def train(self):                                 #가중치, 편향 업데이트
    def predict(self, input_data):                   #미래 값 예측
    def accuracy(self, input_data, target_data):#정확도 측정
```

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – 생성자 __init__()

```python
class NeuralNetwork:

    def __init__(self, input_nodes, hidden_nodes, output_nodes, learning_rate):

        self.input_nodes = input_nodes
        self.hidden_nodes = hidden_nodes
        self.output_nodes = output_nodes

        # 은닉층 가중치  W2 = (784 X 100) Xavier 방법으로 self.W2 가중치 초기화
        self.W2 = np.random.randn(self.input_nodes, self.hidden_nodes) / np.sqrt(self.input_nodes)
        self.b2 = np.random.rand(self.hidden_nodes)

        # 출력층 가중치는 W3 = (100X10)  Xavier 방법으로 self.W3 가중치 초기화
        self.W3 = np.random.randn(self.hidden_nodes, self.output_nodes) / np.sqrt(self.hidden_nodes)
        self.b3 = np.random.rand(self.output_nodes)

        # 출력층 선형회귀 값 Z3, 출력값 A3 정의 (모두 행렬로 표시)
        self.Z3 = np.zeros([1,output_nodes])
        self.A3 = np.zeros([1,output_nodes])

        # 은닉층 선형회귀 값 Z2, 출력값 A2 정의 (모두 행렬로 표시)
        self.Z2 = np.zeros([1,hidden_nodes])
        self.A2 = np.zeros([1,hidden_nodes])

        # 입력층 선형회귀 값 Z1, 출력값 A1 정의 (모두 행렬로 표시)
        self.Z1 = np.zeros([1,input_nodes])
        self.A1 = np.zeros([1,input_nodes])

        # 학습률 learning rate 초기화
        self.learning_rate = learning_rate
```



# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – feed_forward

```
1     def feed_forward(self):
2
3         delta = 1e-7    # log 무한대 발산 방지
4
5         # 입력층 선형회귀 값 Z1, 출력값 A1 계산
6         self.Z1 = self.input_data
7         self.A1 = self.input_data
8
9         # 은닉층 선형회귀 값 Z2, 출력값 A2 계산
10        self.Z2 = np.dot(self.A1, self.W2) + self.b2
11        self.A2 = sigmoid(self.Z2)
12
13        # 출력층 선형회귀 값 Z3, 출력값 A3 계산
14        self.Z3 = np.dot(self.A2, self.W3) + self.b3
15        self.A3 = sigmoid(self.Z3)
16
17        return  -np.sum(self.target_data*np.log(self.A3 + delta) + (1-self.target_data)*np.log((1 - self.A3)+delta))
```

**CrossEntropy**

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – loss_val

```
    def loss_val(self):

        delta = 1e-7    # log 무한대 발산 방지

        # 입력층 선형회귀 값 Z1, 출력값 A1 계산
        self.Z1 = self.input_data
        self.A1 = self.input_data

        # 은닉층 선형회귀 값 Z2, 출력값 A2 계산
        self.Z2 = np.dot(self.A1, self.W2) + self.b2
        self.A2 = sigmoid(self.Z2)

        # 출력층 선형회귀 값 Z3, 출력값 A3 계산
        self.Z3 = np.dot(self.A2, self.W3) + self.b3
        self.A3 = sigmoid(self.Z3)

        return  -np.sum(self.target_data*np.log(self.A3 + delta) + (1-self.target_data)*np.log((1-self.A3)+delta))
```

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – train

```python
def train(self, input_data, target_data):   # input_data : 784 개, target_data : 10개

    self.target_data = target_data
    self.input_data = input_data

    # 먼저 feed forward 를 통해서 최종 출력값과 이를 바탕으로 현재의 에러 값 계산
    loss_val = self.feed_forward()

    # 출력층 loss 인 loss_3 구함
    loss_3 = (self.A3-self.target_data) * self.A3 * (1-self.A3)

    # 출력층 가중치 W3, 출력층 바이어스 b3 업데이트
    self.W3 = self.W3 - self.learning_rate * np.dot(self.A2.T, loss_3)

    self.b3 = self.b3 - self.learning_rate * loss_3

    # 은닉층 loss 인 loss_2 구함
    loss_2 = np.dot(loss_3, self.W3.T) * self.A2 * (1-self.A2)

    # 은닉층 가중치 W2, 은닉층 바이어스 b2 업데이트
    self.W2 = self.W2 - self.learning_rate * np.dot(self.A1.T, loss_2)

    self.b2 = self.b2 - self.learning_rate * loss_2
```

- $loss\_3 = \left[\left(a_1^{(3)} - t_1\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) \quad \left(a_2^{(3)} - t_2\right) a_2^{(3)} \left(1 - a_2^{(3)}\right)\right]$

- $W^{(3)} := W^{(3)} - \alpha \left(A^{(2)^T} \cdot loss\_3\right)$

- $b^{(3)} := b^{(3)} - \alpha \times loss\_3$

- $loss\_2 = \left(loss\_3 \cdot W^{(3)^T}\right) \times \left(A^{(2)}\left(1 - A^{(2)}\right)\right)$

- $W^{(2)} := W^{(2)} - \alpha \left(A^{(2)^T} \cdot loss\_2\right)$

- $b^{(2)} := b^{(2)} - \alpha \times loss\_2$

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – predict

```
    def predict(self, input_data):        # input_data 는 행렬로 입력됨 즉, (1, 784) shape 을 가짐

        Z2 = np.dot(input_data, self.W2) + self.b2
        A2 = sigmoid(Z2)

        Z3 = np.dot(A2, self.W3) + self.b3
        A3 = sigmoid(Z3)

        predicted_num = np.argmax(A3)

        return predicted_num
```

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – accuracy

```
    def accuracy(self, test_data):
        matched_list = []
        not_matched_list = []

        for index in range(len(test_data)):
            label = int(test_data[index, 0])
            # one-hot encoding을 위한 데이터 정규화 (data normalize)
            data = (test_data[index, 1:] / 255.0 * 0.99) + 0.01
            # predict 를 위해서 vector 을 matrix 로 변환하여 인수로 넘겨줌
            predicted_num = self.predict(np.array(data, ndmin=2))

            if label == predicted_num:
                matched_list.append(index)
            else:
                not_matched_list.append(index)

        print("Current Accuracy = ", 100*(len(matched_list)/(len(test_data))), " %")

        return matched_list, not_matched_list
```

# 오차역전파를 이용한 MNIST 분류

## CODE – NeuralNetwork class – 학습

```
# 0~9 숫자 이미지가 784개의 숫자 (28X28) 로 구성되어 있는 training data 읽어옴
training_data = np.loadtxt('./mnist_train.csv', delimiter=',', dtype=np.float32)

# 0~9 숫자 이미지가 784개의 숫자 (28X28) 로 구성되어 있는 test data 읽어옴
test_data = np.loadtxt('./mnist_test.csv', delimiter=',', dtype=np.float32)

input_nodes = 784       #input nodes 개수
hidden_nodes = 100      #hidden nodes 개수
output_nodes = 10       #outpuy nodes 개수
learning_rate = 0.3     #learning rate
epochs = 1              #반복 횟수

nn = NeuralNetwork(input_nodes, hidden_nodes, output_nodes, learning_rate)

start_time = datetime.now()

for i in range(epochs):

    for step in range(len(training_data)):  # train

        # input_data, target_data normalize
        target_data = np.zeros(output_nodes) + 0.01
        target_data[int(training_data[step, 0])] = 0.99

        input_data = ((training_data[step, 1:] / 255.0) * 0.99) + 0.01

        nn.train( np.array(input_data, ndmin=2), np.array(target_data, ndmin=2) )

        if step % 400 == 0:
            print("step = ", step,  ",  loss_val = ", nn.loss_val())

end_time = datetime.now()
print("\nelapsed time = ", end_time - start_time)
```

# 오차역전파를 이용한 MNIST 분류

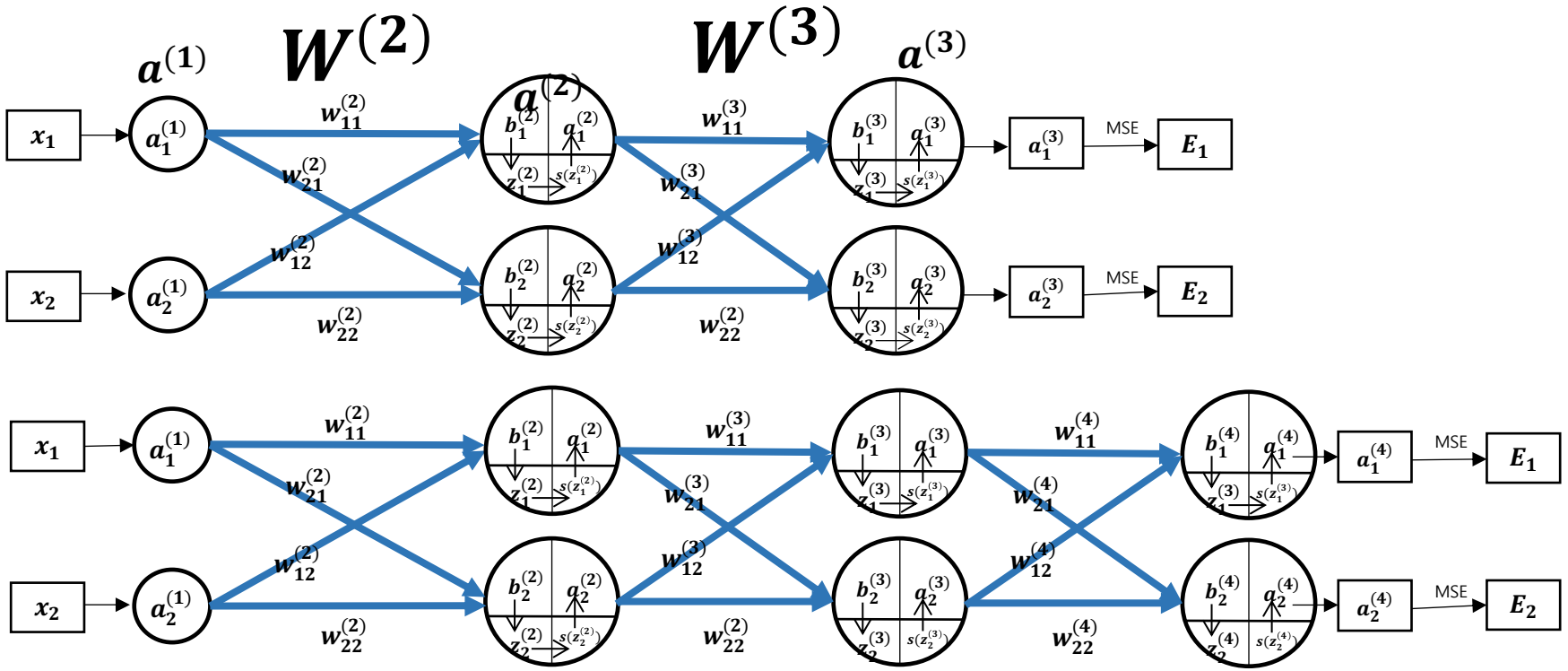
## 결과

| | |
|---|---|
| 학습 결과 | epochs = 0 , step = 0 , current loss_val = 3.596431774902841<br>epochs = 0 , step = 1000 , current loss_val = 0.8424336448981657<br>epochs = 0 , step = 2000 , current loss_val = 1.13223305094993<br>……<br>epochs = 0 , step = 57000 , current loss_val = 1.1467274015653924<br>epochs = 0 , step = 58000 , current loss_val = 0.9788280232573221<br>epochs = 0 , step = 59000 , current loss_val = 0.935433991132101 |
| 검증 코드 | test_data = np.loadtxt('./mnist_test.csv', delimiter=',', dtype=np.float32)<br><br>test_input_data = test_data[ :, 1: ]<br>test_target_data = test_data[ :, 0 ]<br>(true_list, false_list) = obj.accuracy(test_input_data, test_target_data) ⑤ |
| 검증 결과 | Current Accuracy = 0.9432 |

# CONTENTS

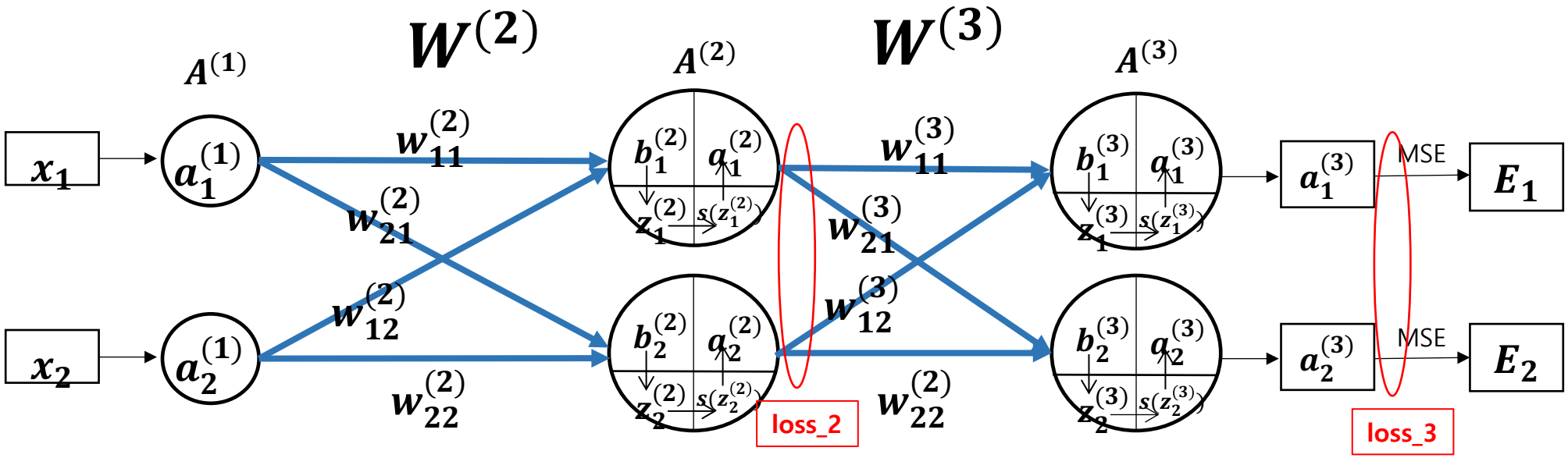
# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 한 개 vs 두 개 인 FFNN구조 비교



# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 한 개 인 FFNN의 일반화: 이전층 | 현재층 | 다음층 의 개념




- $A^{(1)} = \begin{bmatrix} a_1^{(1)} & a_2^{(1)} \end{bmatrix}$
- $A^{(2)} = \begin{bmatrix} a_1^{(2)} & a_2^{(2)} \end{bmatrix}$
- $A^{(3)} = \begin{bmatrix} a_1^{(3)} & a_2^{(3)} \end{bmatrix}$

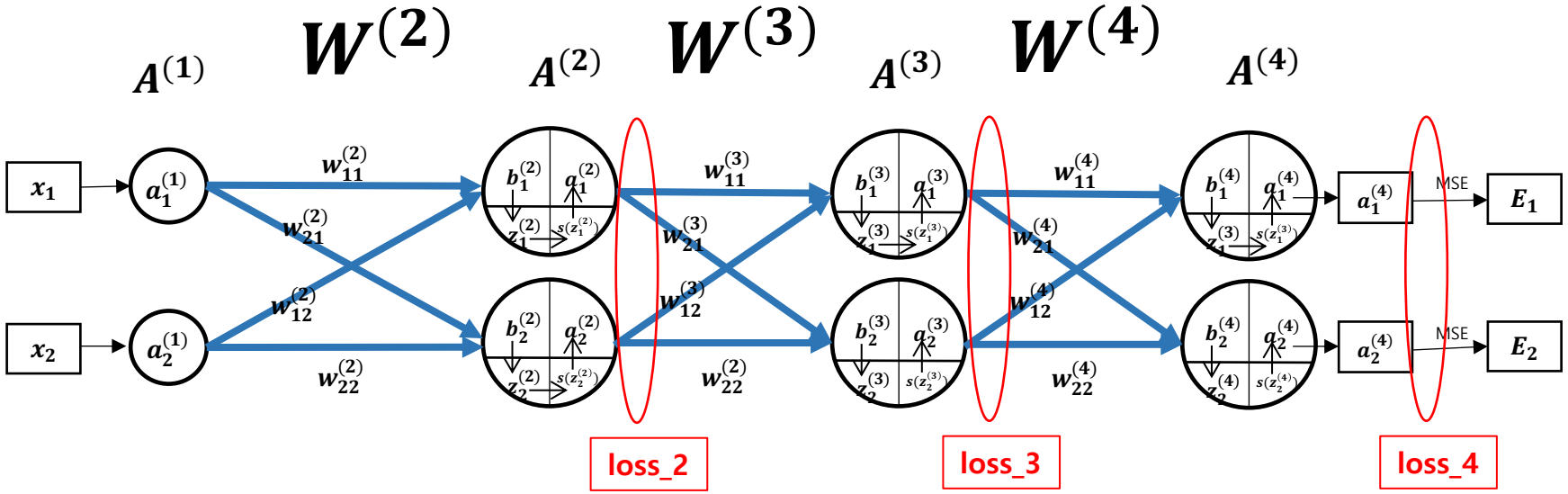
- $\boldsymbol{loss\_2} = \left(loss\_3 \cdot W^{(3)^T}\right) \times A^{(2)}\left(1 - A^{(2)}\right)$
- $\boldsymbol{loss\_3} = \begin{bmatrix} \left(a_1^{(3)} - t_1\right) a_1^{(3)}\left(1 - a_1^{(3)}\right) & \left(a_2^{(3)} - t_2\right) a_2^{(3)}\left(1 - a_2^{(3)}\right) \end{bmatrix}$

- $W^{(2)} := W^{(2)} - \alpha\left(A^{(2)^T} \cdot loss\_2\right)$
- $b^{(2)} := b^{(2)} - \alpha \times loss\_2$

- $W^{(3)} := W^{(3)} - \alpha\left(A^{(2)^T} \cdot loss\_3\right)$
- $b^{(3)} := b^{(3)} - \alpha \times loss\_3$

# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 두 개 인 FFNN의 일반화: 이전층 | 현재층 | 다음층 의 개념

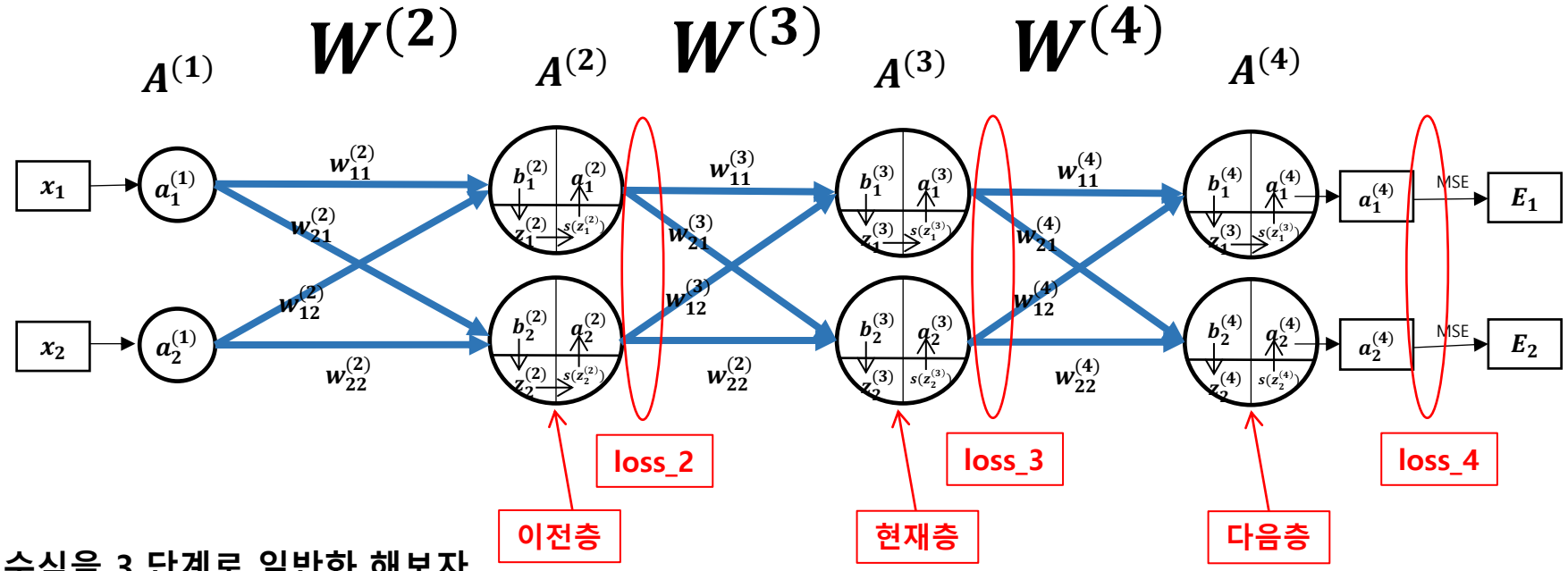



**오차 역전파의 수식을 3 단계로 일반화 해보자**

- 출력층의 손실계산 일반화: 출력층 손실=(출력층출력-정답) X 출력층출력(1-출력층출력)
- 은닉층의 손실계산 일반화: 은닉층의 현재손실 = (다음층손실 · 다음층적용 가중치 $W^T$) * 현재층출력(1-현재층 출력)
- 현재층의 바이어스 변화율 $\frac{\partial E}{\partial b^{(현재층)}}$ = 현재층 손실, $\frac{\partial E}{\partial W^{(현재층)}}$ 현재층에 적용되는 가중치 변화율 = (이전층 출력)^t · 현재층 손실

# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 두 개 인 FFNN의 일반화: 이전층 | 현재층 | 다음층 의 개념

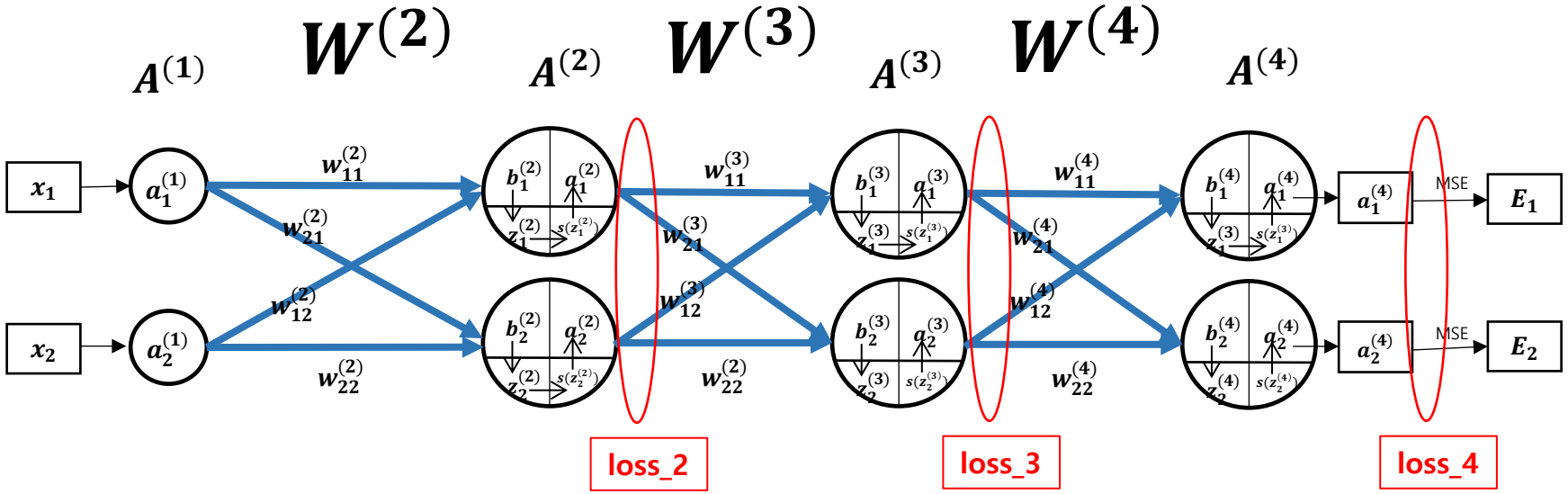



**오차 역전파의 수식을 3 단계로 일반화 해보자**

- 출력층의 손실계산 일반화: 출력층 손실=(출력층출력-정답) X 출력층출력(1-출력층출력) • $loss\_4 = (A^{(4)} - Target) \times A^{(4)}(1 - A^{(4)})$
- 은닉층의 손실계산 일반화: 은닉층의 현재손실 = (다음층손실 · 다음층적용 가중치 $W^T$) * 현재층출력(1-현재층 출력)
  - $loss\_2 = \left[\left(loss\_3 \cdot W^{(3)^T}\right) \times \ A^{(2)}(1 - A^{(2)})\right]$ • $loss\_3 = \left[\left(loss\_4 \cdot W^{(4)^T}\right) \times \ A^{(3)}(1 - A^{(3)})\right]$
- 현재층의 바이어스 변화율 $\frac{\partial E}{\partial b^{(현재층)}}$ = 현재층 손실, $\frac{\partial E}{\partial W^{(현재층)}}$ 현재층에 적용되는 가중치 변화율 = (이전층 출력)^t · 현재층 손실
  - $b^{(2)} := b^{(2)} - \alpha \times loss\_2$
  - $W^{(2)} := W^{(2)} - \alpha \left(A^{(1)^T} \cdot loss\_2\right)$

# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 두 개 인 FFNN의 일반화: 이전층 | 현재층 | 다음층 의 개념




**정리**

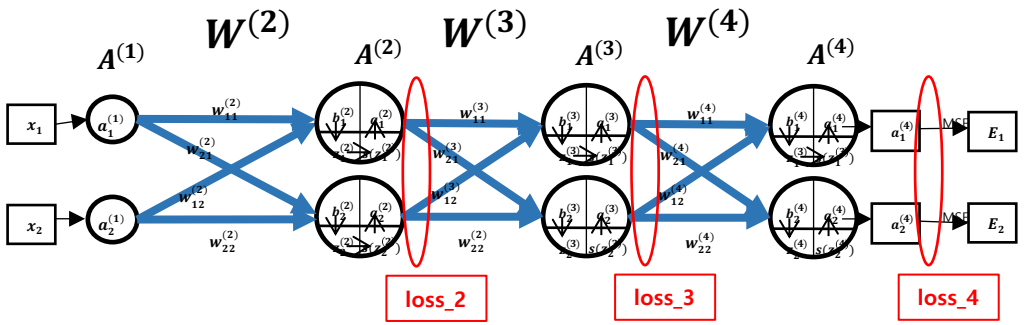
- $W^{(4)} := W^{(4)} - \alpha\left(A^{(3)^T} \cdot loss\_4\right)$
- $W^{(3)} := W^{(3)} - \alpha\left(A^{(2)^T} \cdot loss\_3\right)$
- $W^{(2)} := W^{(2)} - \alpha\left(A^{(1)^T} \cdot loss\_2\right)$

- $b^{(4)} := b^{(4)} - \alpha \times loss\_4$
- $b^{(3)} := b^{(3)} - \alpha \times loss\_3$
- $b^{(2)} := b^{(2)} - \alpha \times loss\_2$

- $\boldsymbol{loss\_4} = \left(A^{(4)} - Target\right) \times A^{(4)}\left(1 - A^{(4)}\right)$
- $\boldsymbol{loss\_3} = \left[\left(loss\_4 \cdot W^{(4)^T}\right) \times \quad A^{(3)}\left(1 - A^{(3)}\right)\right]$
- $\boldsymbol{loss\_2} = \left[\left(loss\_3 \cdot W^{(3)^T}\right) \times \quad A^{(2)}\left(1 - A^{(2)}\right)\right]$

# 은닉층이 여러 개일 경우의 오차 역전파

## 은닉층이 두 개 인 FFNN의 일반화 구현




정리

- $W^{(4)} := W^{(4)} - \alpha\left(A^{(3)^T} \cdot loss\_4\right)$
- $W^{(3)} := W^{(3)} - \alpha\left(A^{(2)^T} \cdot loss\_3\right)$
- $W^{(2)} := W^{(2)} - \alpha\left(A^{(1)^T} \cdot loss\_2\right)$
- $b^{(4)} := b^{(4)} - \alpha \times loss\_4$
- $b^{(3)} := b^{(3)} - \alpha \times loss\_3$
- $b^{(2)} := b^{(2)} - \alpha \times loss\_2$
- $\boldsymbol{loss\_4} = \left(A^{(4)} - Target\right) \times A^{(4)}\left(1 - A^{(4)}\right)$
- $\boldsymbol{loss\_3} = \left[\left(loss\_4 \cdot W^{(4)^T}\right) \times \quad A^{(3)}\left(1 - A^{(3)}\right)\right]$
- $\boldsymbol{loss\_2} = \left[\left(loss\_3 \cdot W^{(3)^T}\right) \times \quad A^{(2)}\left(1 - A^{(2)}\right)\right]$

어떻게 구현할까 고민해 봅시다!


감사합니다.