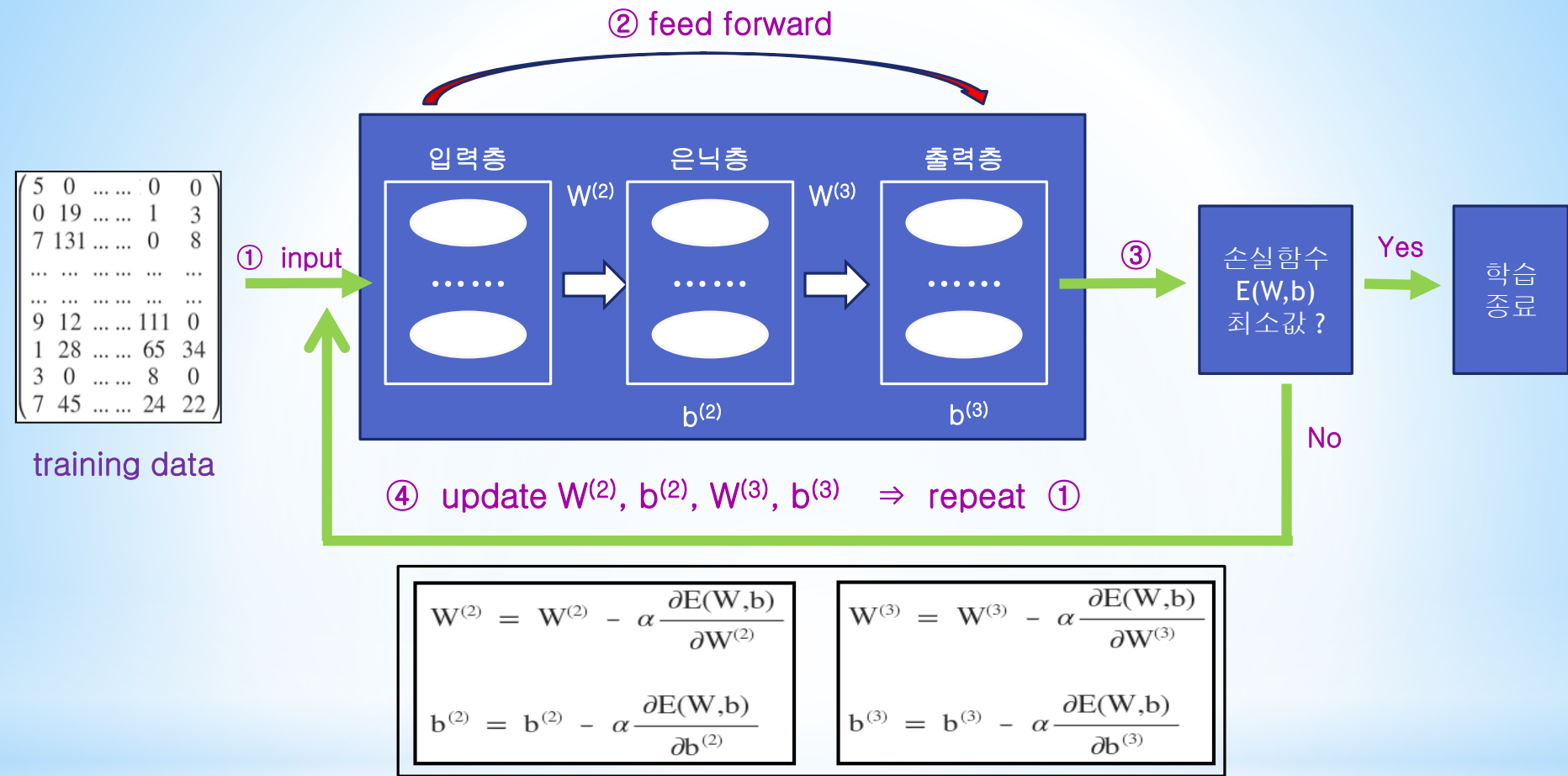


파이썬(Python)으로 구현하는

# 오차역전파 (Back Propagation)

- 개념 • 원리 -

# Review – 수치미분 문제점

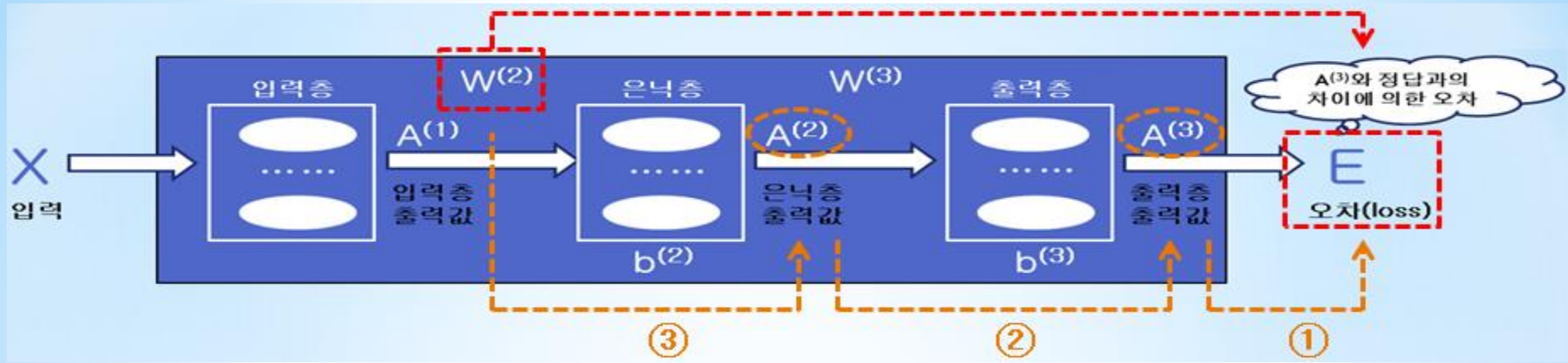


문제점

수치미분으로 가중치/바이어스 업데이트 시 많은 시간 소요 됨

[예] 784X300X10 딥러닝 아키텍처에서 60,000 만개의 데이터를 학습할 경우, 컴퓨터 환경에 따라서는 24시간 이상 소요 !!

# 오차역전파 (Back Propagation) – 개념 및 원리



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

체인 룰에 의해 ① ② ③ 으로 분리

①

②

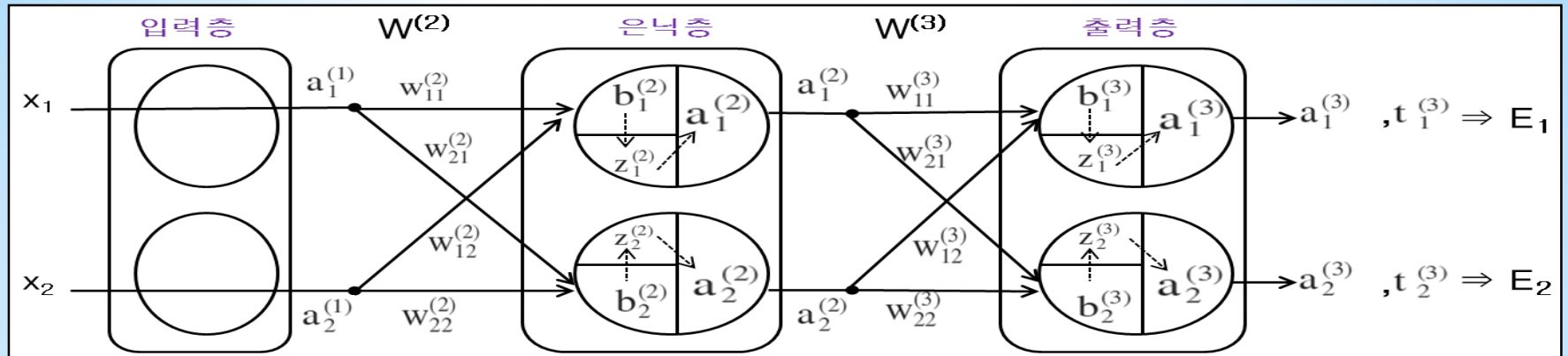
③

## ➤ 오차역전파 (Back Propagation) 개념 및 동작원리

- 가중치나 바이어스가 변할 때 최종오차가 얼마나 변하는지를 나타내는  $\frac{\partial E}{\partial W^{(2)}}$  같은 편미분을,
- 체인 룰(chain rule)을 이용하여 위 식의 ①, ②, ③ 처럼 국소(local) 미분으로 분리 한 후에,
- 이러한 국소 미분을 수학 공식으로 나타내서 계산하는 방법을 오차역전파 라고 함

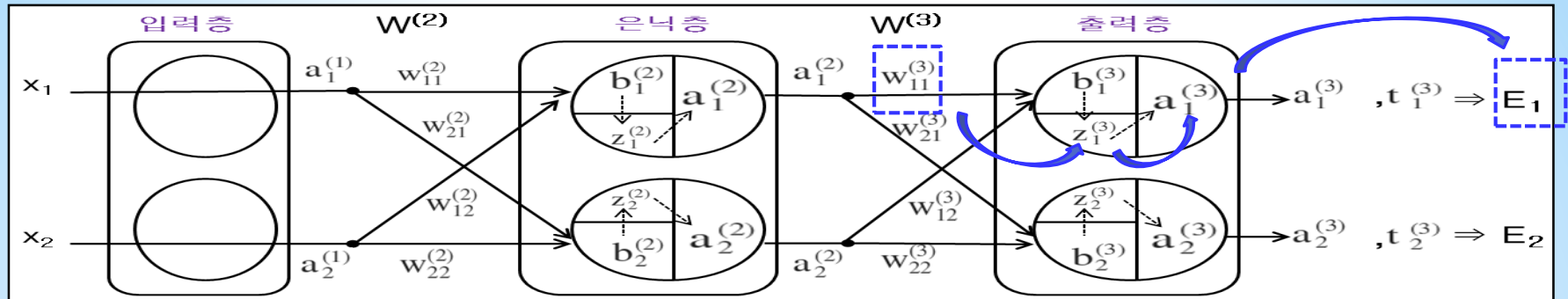
※ 오차역전파는 수치미분을 사용하지 않고 행렬로 표현되는 수학공식을 계산하기 때문에 빠른 계산이 가능함

# 오차역전파 (Back Propagation) – 각 층의 선형회귀 값 (z) / 각층의 출력 값 (a)



|     | 선형 회귀 값 (z)   | 출력 값 (a)  |
|-----|---|---|
| 입력층 | 입력 층에는 가중치가 없기 때문에 선형회귀 값은 적용하지 않음  | $a_1^{(1)} = x_1$ $a_2^{(1)} = x_2$   |
| 은닉층 | $z_1^{(2)} = a_1^{(1)}w_{11}^{(2)} + a_2^{(1)}w_{12}^{(2)} + b_1^{(2)}$ $z_2^{(2)} = a_1^{(1)}w_{21}^{(2)} + a_2^{(1)}w_{22}^{(2)} + b_2^{(2)}$ | $a_1^{(2)} = \text{sigmoid}(z_1^{(2)})$ $a_2^{(2)} = \text{sigmoid}(z_2^{(2)})$ |
| 출력층 | $z_1^{(3)} = a_1^{(2)}w_{11}^{(3)} + a_2^{(2)}w_{12}^{(3)} + b_1^{(3)}$ $z_2^{(3)} = a_1^{(2)}w_{21}^{(3)} + a_2^{(2)}w_{22}^{(3)} + b_2^{(3)}$ | $a_1^{(3)} = \text{sigmoid}(z_1^{(3)})$ $a_2^{(3)} = \text{sigmoid}(z_2^{(3)})$ |

# 오차역전파 (Back Propagation) – 시그모이드 함수 미분

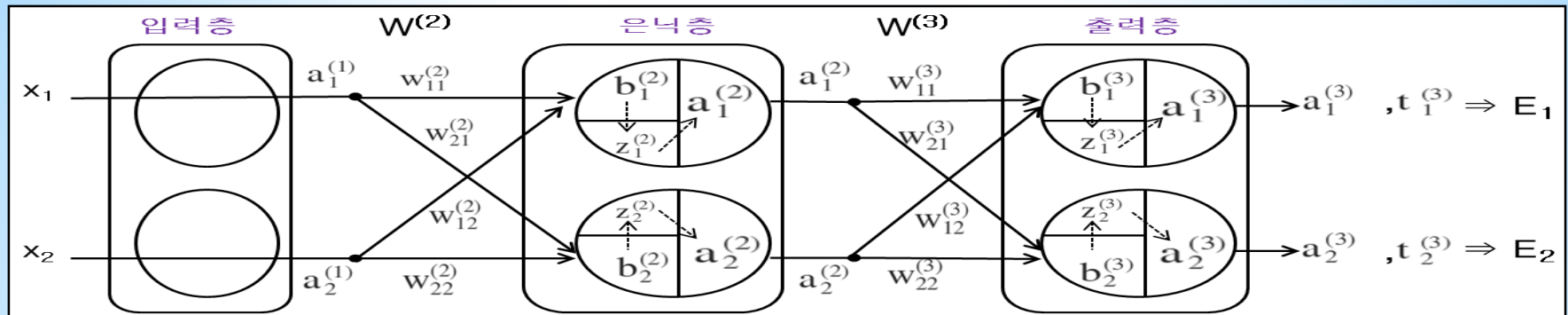


$$\begin{aligned}
 \frac{\partial \text{sigmoid}(z)}{\partial z} &= \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) \\
 &= \frac{\partial}{\partial z} (1 + e^{-z})^{-1} \quad \text{지수표현} \\
 &= \frac{e^{-z}}{(1 + e^{-z})^2} \quad \text{체인 룰 이용한 미분} \\
 &= \frac{1}{(1 + e^{-z})} \times \frac{e^{-z}}{(1 + e^{-z})} \quad \text{분모분리} \\
 &= \frac{1}{(1 + e^{-z})} \times \frac{(1 + e^{-z}) - 1}{(1 + e^{-z})} \quad \text{분자에 +1-1 해줌} \\
 &= \frac{1}{(1 + e^{-z})} \times \left( 1 - \frac{1}{(1 + e^{-z})} \right) \quad \text{약분 및 식 정리} \\
 &= \boxed{\text{sigmoid}(z) \times (1 - \text{sigmoid}(z))}
 \end{aligned}$$

적용  
사례  
→

$$\begin{aligned}
 \frac{\partial E_1}{\partial w_{11}^{(3)}} &= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}} \\
 &= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial \text{sigmoid}(z_1^{(3)})}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}} \quad \left( a_1^{(3)} = \text{sigmoid}(z_1^{(3)}) \right) \\
 &= \frac{\partial E_1}{\partial a_1^{(3)}} \times \boxed{\text{sigmoid}(z_1^{(3)}) \times (1 - \text{sigmoid}(z_1^{(3)}))} \times \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}} \\
 &= \frac{\partial E_1}{\partial a_1^{(3)}} \times \boxed{a_1^{(3)} \times (1 - a_1^{(3)})} \times \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}}
 \end{aligned}$$

# 오차역전파 (Back Propagation) - 최종 손실(에러) 값 E / 가중치 W / 바이어스 b



|                    |  |  |                    |   |
|--------------------|--|--|--------------------|---|
| 최종 손실값             | $E = \frac{1}{n} \sum_{i=1}^n (t_i^{(3)} - a_i^{(3)})^2 = \frac{1}{2} \{ (t_1^{(3)} - a_1^{(3)})^2 + (t_2^{(3)} - a_2^{(3)})^2 \} = E_1 + E_2 \rightarrow E_1 = \frac{1}{2} (t_1^{(3)} - a_1^{(3)})^2 \quad E_2 = \frac{1}{2} (t_2^{(3)} - a_2^{(3)})^2$ |  |                    |   |
| $W^{(2)}, W^{(3)}$ | $W^{(2)} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{pmatrix}$   | $W^{(3)} = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$ | $b^{(2)}, b^{(3)}$ | $b^{(2)} = \begin{pmatrix} b_1^{(2)} & b_2^{(2)} \end{pmatrix} \quad b^{(3)} = \begin{pmatrix} b_1^{(3)} & b_2^{(3)} \end{pmatrix}$ |

업데이트

업데이트

업데이트

업데이트

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

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

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

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

$$\frac{\partial E}{\partial w_{11}^{(2)}}, \frac{\partial E}{\partial w_{21}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \frac{\partial E}{\partial w_{22}^{(2)}}$$

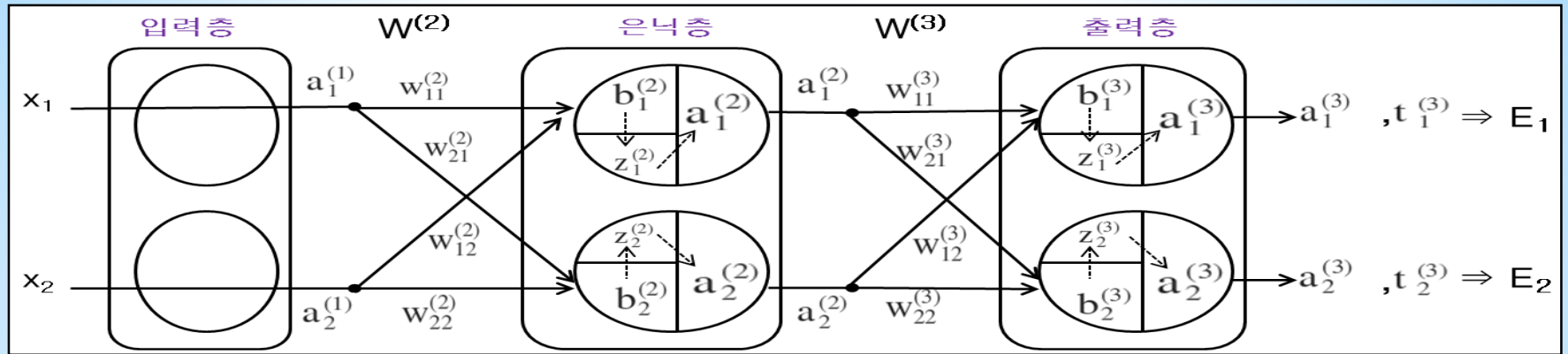
$$\frac{\partial E}{\partial w_{11}^{(3)}}, \frac{\partial E}{\partial w_{21}^{(3)}}, \frac{\partial E}{\partial w_{12}^{(3)}}, \frac{\partial E}{\partial w_{22}^{(3)}}$$

$$\frac{\partial E}{\partial b_1^{(2)}}, \frac{\partial E}{\partial b_2^{(2)}}$$

$$\frac{\partial E}{\partial b_1^{(3)}}, \frac{\partial E}{\partial b_2^{(3)}}$$



# 오차역전파 (Back Propagation) - 가중치 / 바이어스 업데이트



$$\frac{\partial E}{\partial w_{11}^{(2)}}, \frac{\partial E}{\partial w_{21}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \frac{\partial E}{\partial w_{22}^{(2)}}$$

$$\frac{\partial E}{\partial w_{11}^{(3)}}, \frac{\partial E}{\partial w_{21}^{(3)}}, \frac{\partial E}{\partial w_{12}^{(3)}}, \frac{\partial E}{\partial w_{22}^{(3)}}$$

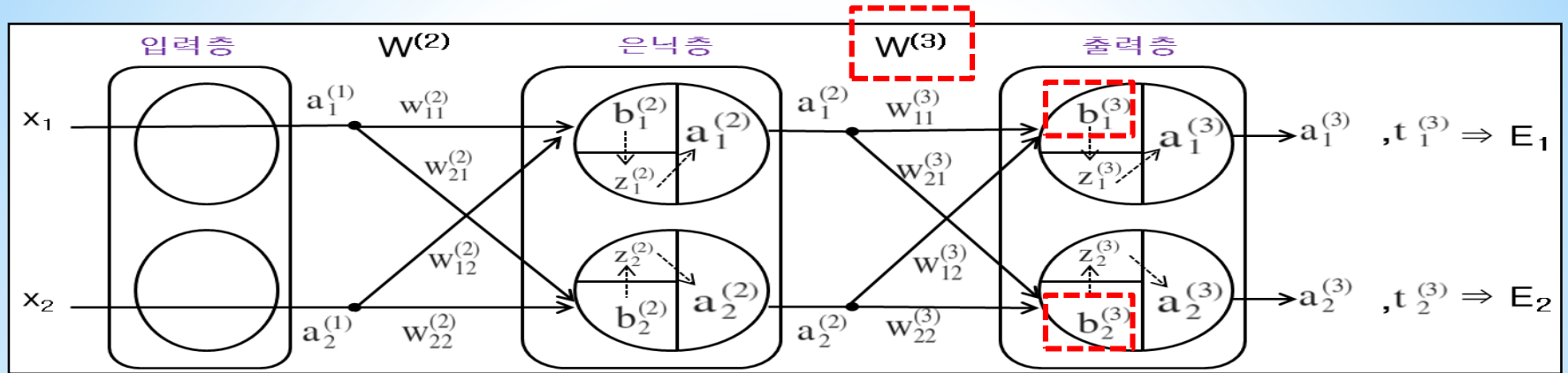
$$\frac{\partial E}{\partial b_1^{(2)}}, \frac{\partial E}{\partial b_2^{(2)}}$$

$$\frac{\partial E}{\partial b_1^{(3)}}, \frac{\partial E}{\partial b_2^{(3)}}$$



이러한 가중치  $W^{(2)}$  / 가중치  $W^{(3)}$  / 바이어스  $b^{(2)}$  / 바이어스  $b^{(3)}$  미분 값을 수치미분이 아닌 행렬 기반의 수학기공식을 이용하는 오차역전파 구현

# 출력층 오차역전파 공식 유도 - 출력층 가중치 $W^{(3)}$ / 출력층 바이어스 $b^{(3)}$



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

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

$$\frac{\partial E}{\partial w_{11}^{(3)}}, \frac{\partial E}{\partial w_{21}^{(3)}}, \frac{\partial E}{\partial w_{12}^{(3)}}, \frac{\partial E}{\partial w_{22}^{(3)}}$$

출력층 가중치  $W^{(3)}$

$$\frac{\partial E}{\partial b_1^{(3)}}, \frac{\partial E}{\partial b_2^{(3)}}$$

출력층 바이어스  $b^{(3)}$

$\partial E / \partial W^{(3)}, \partial E / \partial b^{(3)}$  오차역전파 공식 유도

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

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

$$\frac{\partial E}{\partial w_{11}^{(2)}}, \frac{\partial E}{\partial w_{21}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \frac{\partial E}{\partial w_{22}^{(2)}}$$

은닉층 가중치  $W^{(2)}$

$$\frac{\partial E}{\partial b_1^{(2)}}, \frac{\partial E}{\partial b_2^{(2)}}$$

은닉층 바이어스  $b^{(2)}$

$\partial E / \partial W^{(2)}, \partial E / \partial b^{(2)}$  오차역전파 공식 유도