2021 인공지능 소수전공

46~50차시: 오차역전파

2021.08.03 17:30~22:15

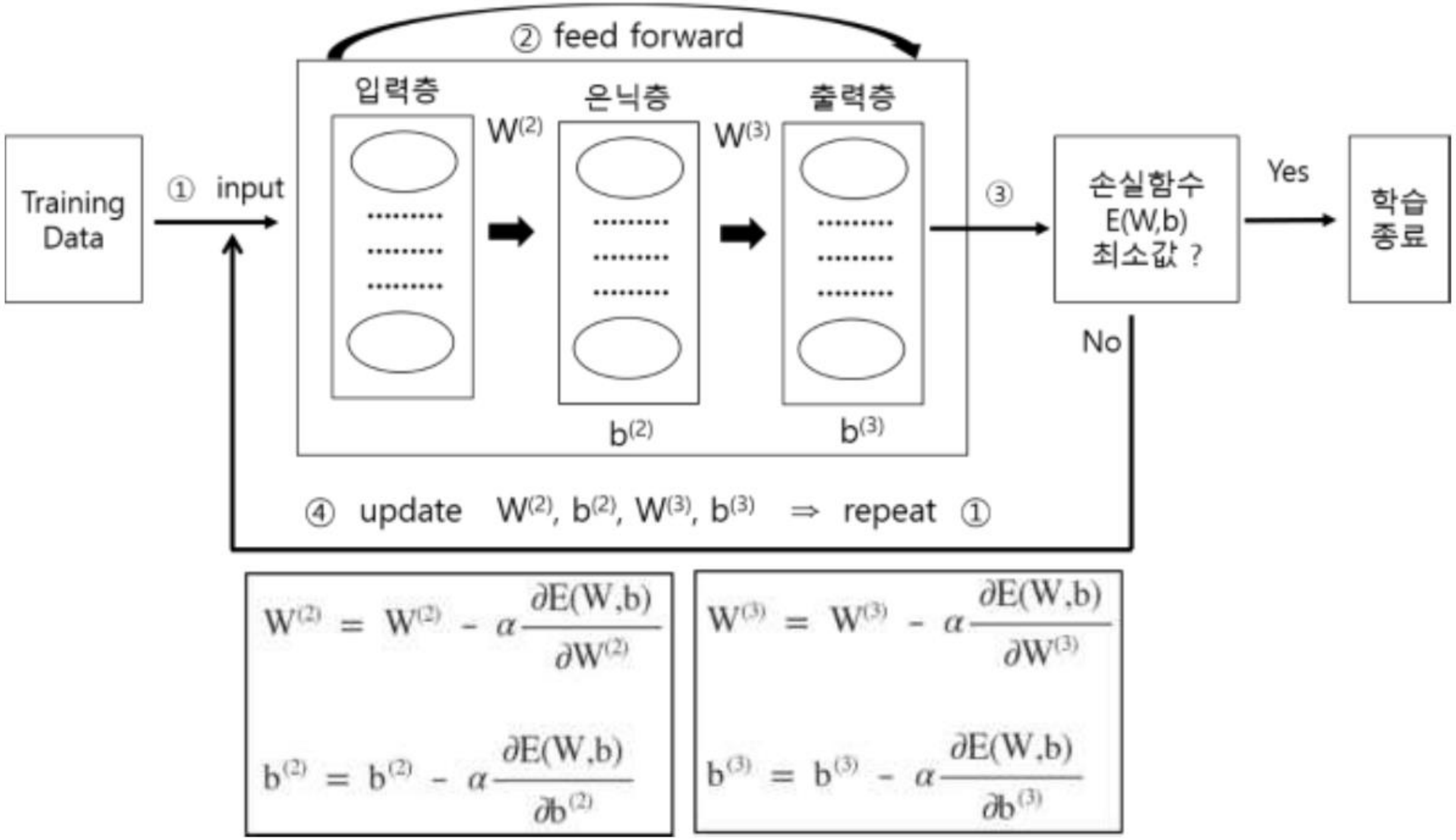
Seokhwan Yang

오차 역전파(Back Propagation)

- 내용
  - 수치미분 MNIST
  - 오차 역전파(Back Propagation) 개념과 동작원리
  - 오차 역전파와 관련된 수학 공식 유도
  - 오차 역전파 기반 딥러닝 아키텍처
  - 오차 역전파 MNIST

수치 미분 기반 MNIST

- **수치 미분을 이용한 딥러닝 아키텍처**



알고리즘

① 트레이닝 데이터에 대하여

② 피드 포워드 함수 수행 후

③ 손실 함수 계산

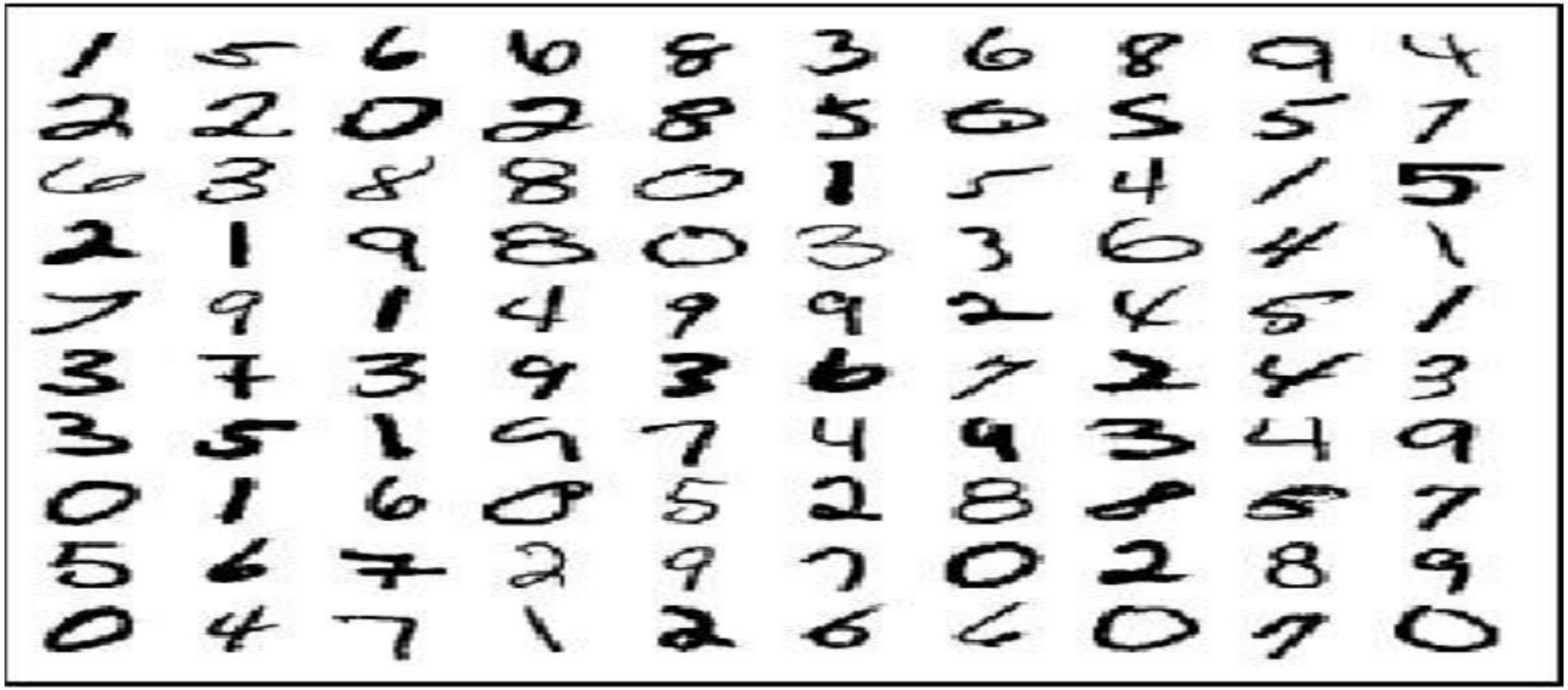
④ 손실 함수가 최소가 아니면

⑤ 수치 미분을 통해서 가중치와 편향치를 갱신하고

⑥ 다시 피드 포워드 반복

수치 미분 기반 MNIST

- **MNIST (필기체 손글씨)**



수치 미분 기반 MNIST

- MNIST 다운로드 사이트
  - Training Data (csv format)
    - http://www.pjreddie.com/media/files/mnist_train.csv
  - Test Data (csv format)
    - http://www.pjreddie.com/media/files/mnist_test.csv

- 강의 Github의 data 폴더에 있음

수치 미분 기반 MNIST

- **MNIST 데이터 구조**
  - **mnist_train.csv**
    - 총 6만개의 레이블링 된 데이터
    - 각 글자는 1 + 28x28 (1 + 784 개 숫자 데이터, ‘,’로 구분되어 있음)
    - 처음 1개는 레이블링 값, 나머지 784개는 문자의 픽셀 데이터
  - **mnist_test.csv**
    - 총 1만개의 레이블링 된 데이터
    - 나머지는 mnist_train.csv 파일과 동일

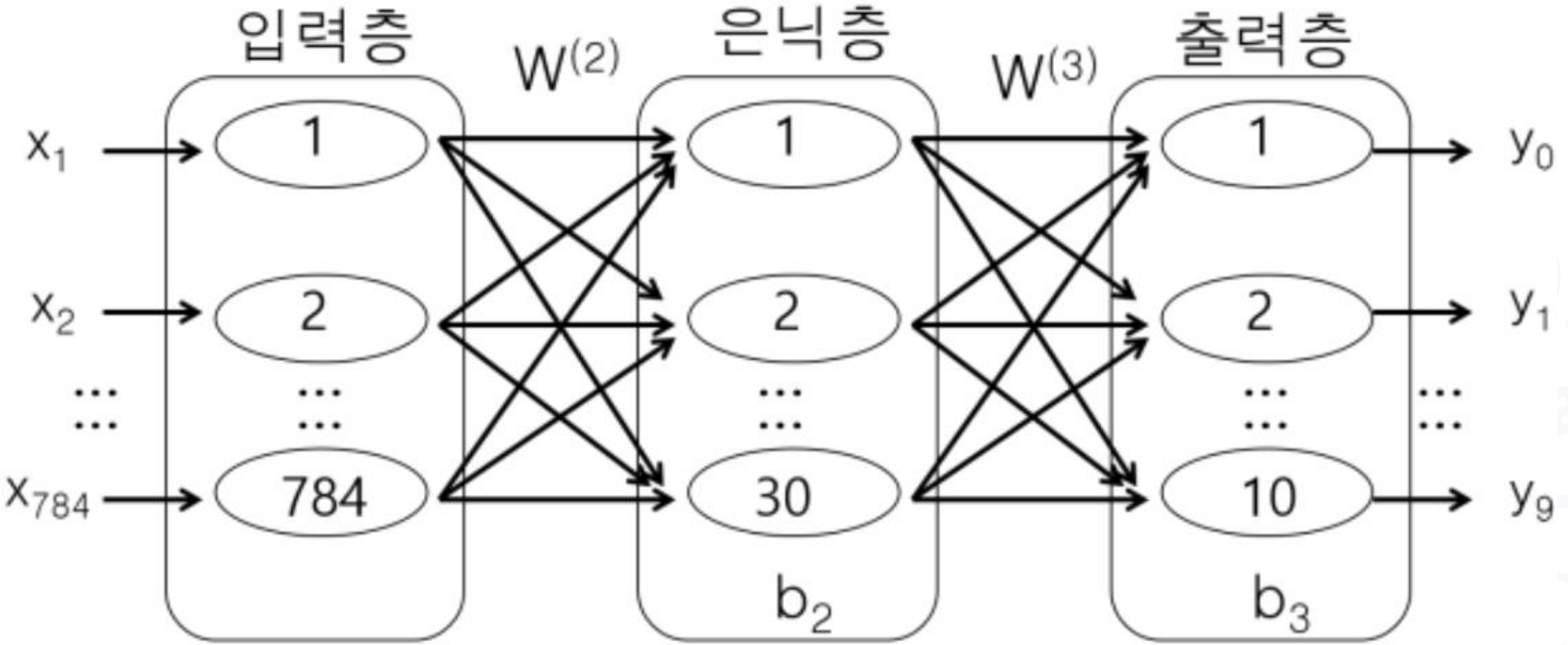
수치 미분 기반 MNIST

- **mnist_train.csv 구조 예시 (5)**

```
5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,18,18,18,126,136,175,26,166,255,247,127,0,0
,0,0,0,0,0,0,0,0,0,0,30,36,94,154,170,253,253,253,253,253,225,172,253,242,195,64,0,0,0,0,0,0,0,0,0,0,0,49,238,253,253
,253,253,253,253,253,253,251,93,82,82,56,39,0,0,0,0,0,0,0,0,0,0,0,0,18,219,253,253,253,253,253,198,182,247,241,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,80,156,107,253,253,205,11,0,43,154,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,14,1,154,253,9
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,139,253,190,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,11,
190,253,70,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,35,241,225,160,108,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,81,240,253,253,119,25,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,45,186,253,253,150,27,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,16,93,252,253,187,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,249,253,249,64,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,46,130,183,253,253,207,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,39,148,229,253,2
53,253,250,182,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,24,114,221,253,253,253,253,201,78,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,23,66,213,253,253,253,253,198,81,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,18,171,219,253,253,253,253,195,80,9,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,55,172,226,253,253,253,253,244,133,11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,136,253,253,253,212,135
,132,16,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

수치 미분 기반 MNIST

- **MNIST 인식을 위한 딥러닝 아키텍처**



수치 미분 기반 MNIST

- **One-Hot Encoding 예시**

인덱스	0	1	2	3	4	5	6	7	8	9
값	0.01	0.01	0.01	0.01	0.01	0.99	0.01	0.01	0.01	0.01
노드번호	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9

수치 미분 기반 MNIST

- **수치 미분 기반 딥러닝 모델의 단점**
  - **가중치와 편향치의 학습에 많은 시간 소요**
    - 예: MNIST의 경우
      - 784개(28x28)입력 층 노드, 30개의 은닉 층 노드, 10개의 출력 층 노드 보유
      - 학습 시 평균 20시간 이상 소요(1 CPU 환경)
      - 입력 데이터의 개수가 증가하거나 은닉 층 노드 개수 증가 시 더 많은 시간 소요가 예상됨

수치 미분 기반 MNIST

- [문제 1] 입력 데이터 정규화 및 정답 데이터 One-Hot Encoding
  - 1) MNIST 입력데이터를 0.01 ~ 1.00 값으로 정규화 하는 코드를 구현하시오
  - 2) MNIST 정답데이터를 One-Hot Encoding 방식으로 표현하시오

수치 미분 기반 MNIST

- [문제 2] 수치미분을 이용한 MNIST_Test class 검증
  - 1) MNIST_Test 클래스에서 np.random.rand(…) 가중치 초기화
  - 2) 은닉층 노드 1 개 (h_nodes) 설정
  - 3) obj = MNIST_Test(i_nodes, h_nodes, o_nodes, learning_rate)
  - obj.train(input_data, target_data)
  - obj.accuracy(test_input_data, test_target_data)

수치 미분 기반 MNIST

- [문제 3] 수치미분을 이용한 MNIST_Test class 검증
  - 1) MNIST_Test 클래스에서 Xavier / He 방식으로 가중치 초기화
  - 2) 은닉층 노드 1 개 (h_nodes) 설정
  - 3) obj = MNIST_Test(i_nodes, h_nodes, o_nodes, learning_rate)
  - obj.train(input_data, target_data)
  - obj.accuracy(test_input_data, test_target_data)

수치 미분 기반 MNIST

- [문제 4] MNIST_Test 클래스에서 Xavier / He 방식으로 가중치 초기화하고, 은닉층
  - 노드수를 2 개로 하여 소요시간 및 정확도를 측정하시오
  - obj = MNIST_Test(i_nodes, h_nodes, o_nodes, learning_rate)
  - obj.train(input_data, target_data)
  - obj.accuracy(test_input_data, test_target_data)

수치 미분 기반 MNIST

- [문제 5] MNIST_Test 클래스에서 Xavier / He 방식으로 가중치 초기화하고, 은닉층
  - 노드수를 8 개로 하여 소요시간 및 정확도를 측정하시오
  - obj = MNIST_Test(i_nodes, h_nodes, o_nodes, learning_rate)
  - obj.train(input_data, target_data)
  - obj.accuracy(test_input_data, test_target_data)

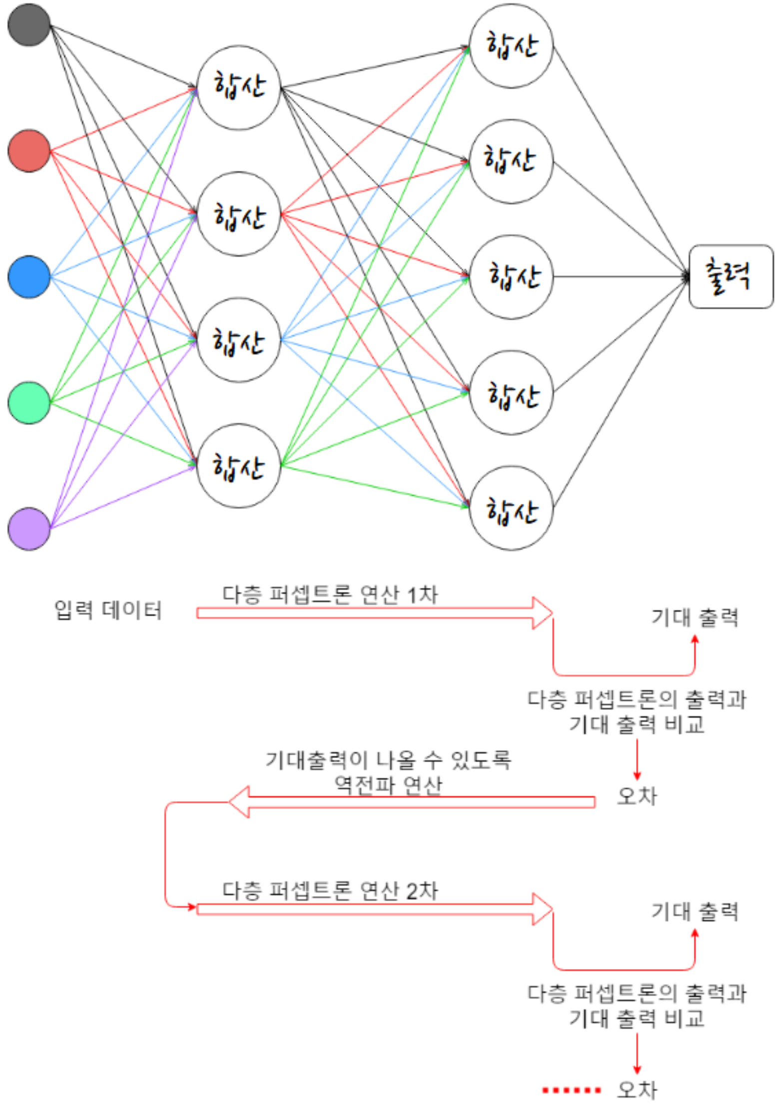
수치 미분 기반 MNIST

- [문제 6] MNIST_Test 클래스에서 Xavier / He 방식으로 가중치 초기화하고, 은닉층
  - 노드수를 30 개로 하여 소요시간 및 정확도를 측정하시오
  - obj = MNIST_Test(i_nodes, h_nodes, o_nodes, learning_rate)
  - obj.train(input_data, target_data)
  - obj.accuracy(test_input_data, test_target_data)

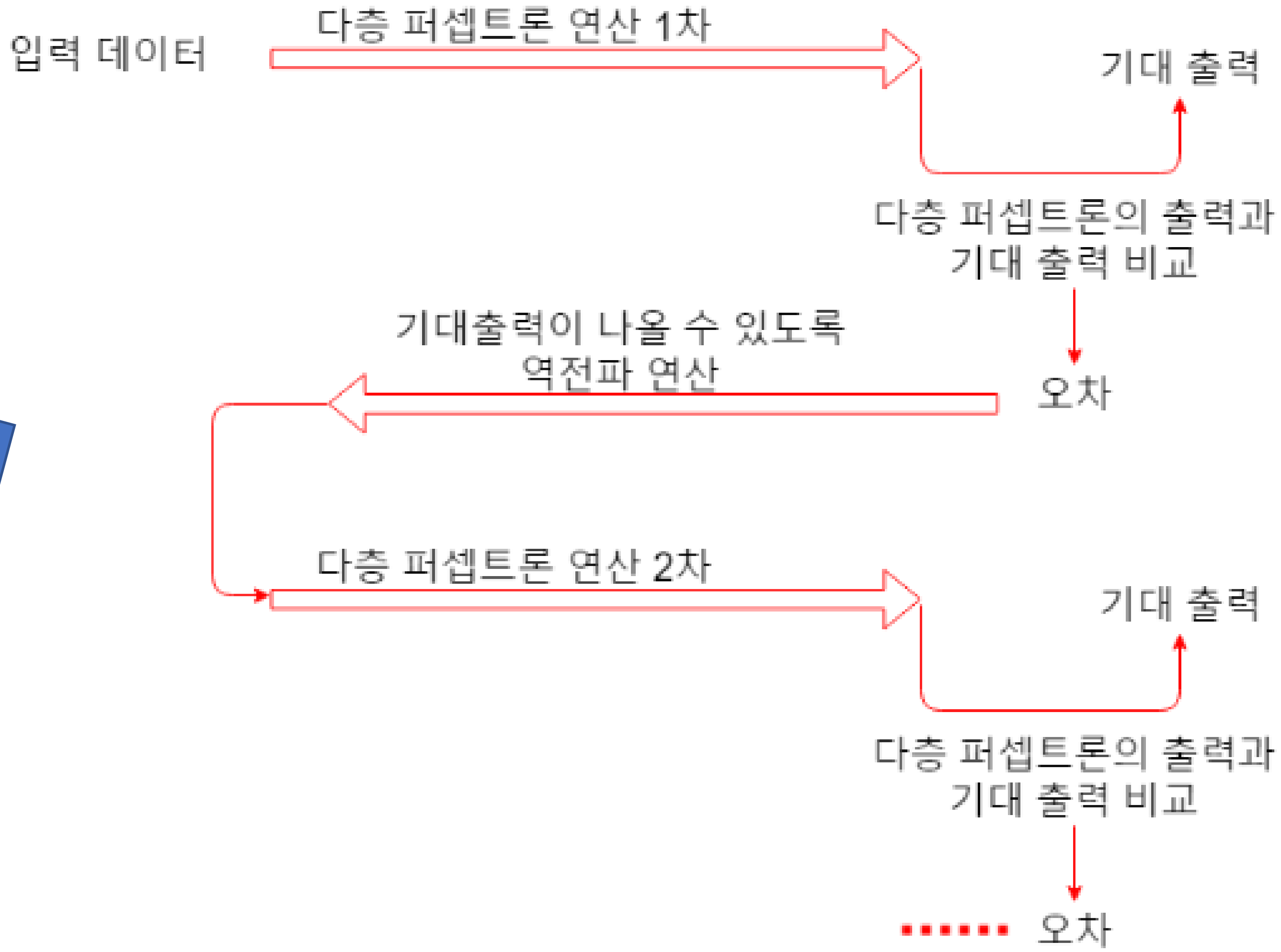
오차 역전파의 개념

- 오차 역전파란?
  - 딥러닝 아키텍처의 은닉층에 포함된 각 노드의 가중치, 편향치 갱신을 위한
  - 편미분(partial derivative) 식을 그대로 계산하는 것이 아니라
  - 체인 룰(chain rule)을 이용하여 국소(local) 미분으로 분리한 후에
  - 분리된 국소(local) 미분을 계산하기 쉬운 형태의 수학공식으로 나타내는 것
  - 수치 미분을 사용하지 않고 행렬(matrix)로 표현되는 수학공식으로 계산되기 때문에 빠른 계산이 가능하다

오차 역전파의 개념



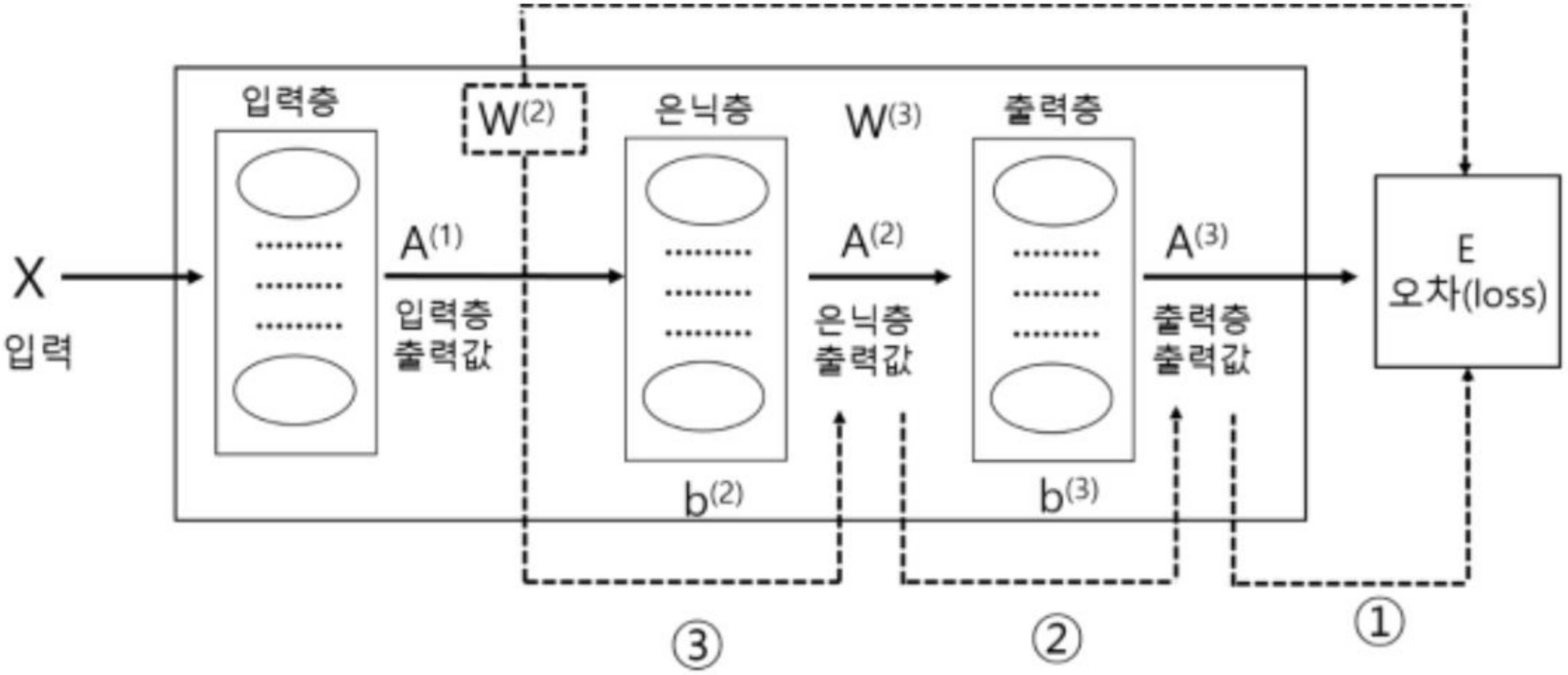
합산
출력
입력 데이터
다층 퍼셉트론 연산 1차
기대 출력
다층 퍼셉트론의 출력과 기대 출력 비교
기대출력이 나올 수 있도록 역전파 연산
오차
다층 퍼셉트론 연산 2차



입력 데이터
다층 퍼셉트론 연산 1차
기대 출력
다층 퍼셉트론의 출력과
기대 출력 비교
기대출력이 나올 수 있도록
역전파 연산
오차
다층 퍼셉트론 연산 2차
기대 출력
다층 퍼셉트론의 출력과
기대 출력 비교
오차

오차 역전파의 동작원리

- **동작원리**



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

체인 룰 (chain rule) 적용 → ① ② ③

오차 역전파의 동작원리

- 표현
  - 입력 데이터: X
  - 입력 층의 출력 값: A(1) 행렬
  - 은닉 층의 출력 값: A(2) 행렬, A(1)→A(2) 가중치: W(2), 편향치: b(2)
  - 출력 층의 출력 값: A(3) 행렬, A(2)→A(3) 가중치: W(3), 편향치: b(3)
  - 출력 층의 출력 값 A(3)과 정답 사이의 오차: E

오차 역전파의 동작원리

- 기존 신경망의 가중치 업데이트 방식
  - $W^{(3)} = W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}}$ 과 같은 편미분이 포함된 방정식 이용

- 오차 역전파에서의 체인 룰에 의한 편미분 방정식 분해

$$W^{(2)} = W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}} = W^{(2)} - \alpha \left(\underbrace{\frac{\partial E}{\partial A^{(3)}}}_{①} \cdot \underbrace{\frac{\partial A^{(3)}}{\partial A^{(2)}}}_{②} \cdot \underbrace{\frac{\partial A^{(2)}}{\partial W^{(2)}}}_{③} \right)$$

체인 룰 (chain rule) 적용

  - ①②③과 같은 국소 미분의 곱 형태로 표현

오차 역전파의 동작원리

- 동작 원리

① 가중치 $W^{(2)}$ / 편향치 $b^{(2)}$ 등이 변할 때 최종 오차 E가 얼마나 변하는지 나타내는 $\frac{\partial E}{\partial W^{(2)}}$ 또는 $\frac{\partial E}{\partial b^{(2)}}$ 와 같은 편미분 식을

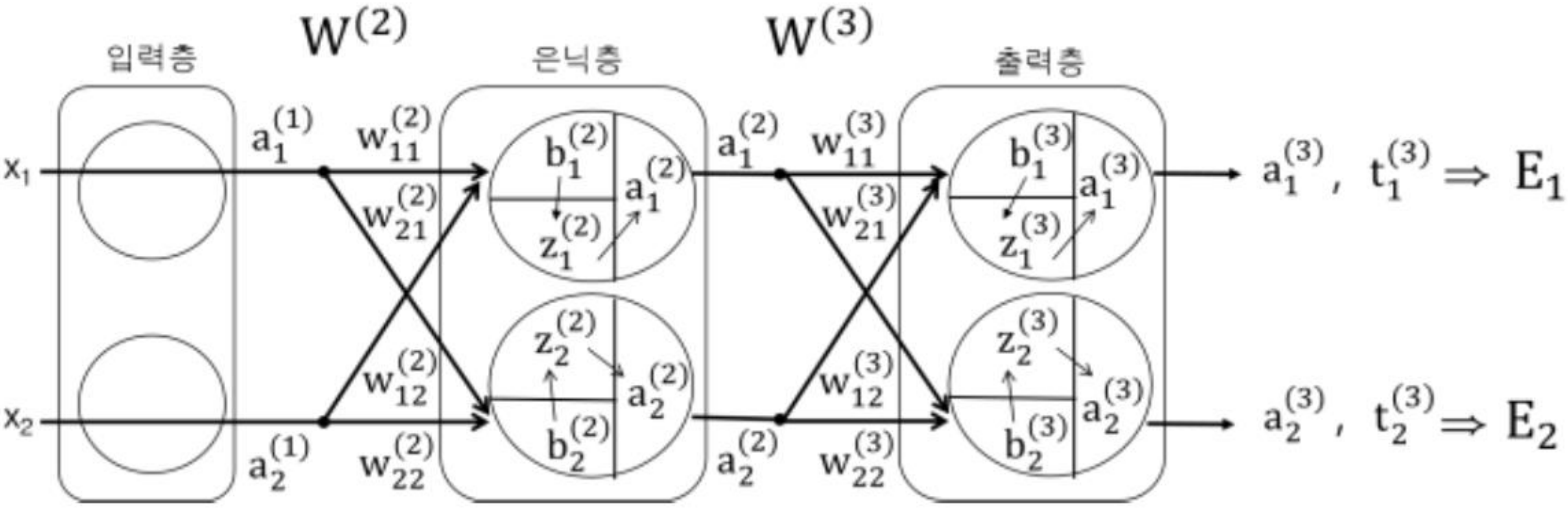
② 체인 룰을 이용하여 국소 미분으로 분리한 후에

③ 국소 미분을 수학공식으로 나타내서

④ 최종적으로는 수치 미분이 아닌 곱하기 형태의 산술식으로 계산함

오차 역전파의 동작원리

- 오차 역전파를 위한 기본 신경망 구조



오차 역전파의 동작원리

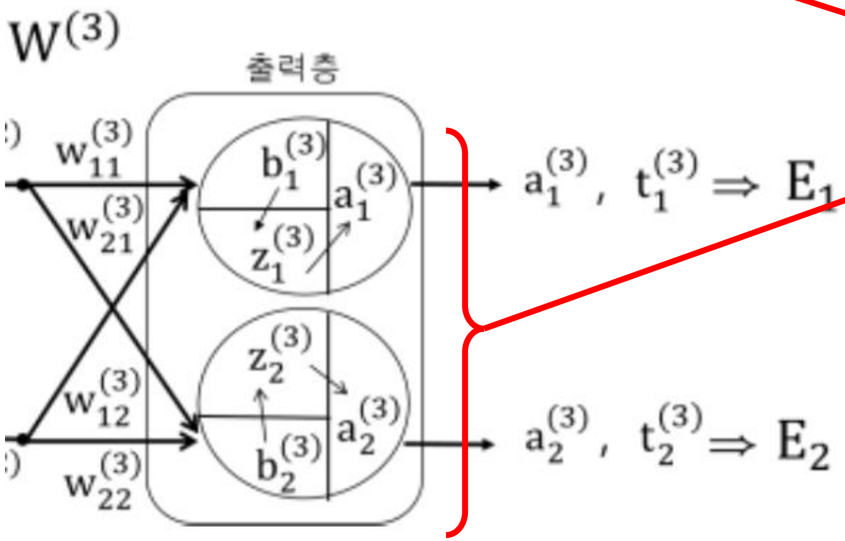
- 각 층의 가중치 W, 편향치 b, 최종 오차 E

가중치	$W^{(2)} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{pmatrix}$	$W^{(3)} = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$
바이어스	$b^{(2)} = \begin{pmatrix} b_1^{(2)} & b_2^{(2)} \end{pmatrix}$	$b^{(3)} = \begin{pmatrix} b_1^{(3)} & b_2^{(3)} \end{pmatrix}$
오차 E	$E = \frac{1}{2}\sum_{i=1}^{2}(t_i^{(3)} - a_i^{(3)})^2 = \frac{1}{2}\left\{\left(t_1^{(3)} - a_1^{(3)}\right)^2 + \left(t_2^{(3)} - a_2^{(3)}\right)^2\right\} = E_1 + E_2$	

- 오차 E는 정답 t와 출력 층에서의 출력 값 a의 사이의 차이를 제곱하는 평균 제곱 오차(MSE)를 적용

- **각 층의 가중치 W, 편향치 b, 최종 오차 E**

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



- 출력 층 노드가 2개이므로 오차를 ½로 계산
- 총 오차 E는 $E_1 + E_2$와 같이 오차의 덧셈으로 표현 가능
- 각각의 오차들은 서로에게 영향을 주지 않는 독립변수이므로 선형적인 덧셈으로 연결 가능

오차 역전파의 동작원리

- **각 층의 선형 회귀 값(z), 출력 값(a)**

	선형 회귀 값 (z)	출력 값 (a)
입력 층	입력 층에는 가중치가 없기 때문에 선형 회귀 값은 적용하지 않습니다.	$a_1^{(1)} = x_1$
입력 층		$a_2^{(1)} = x_2$
은닉 층	$z_1^{(2)} = a_1^{(1)} w_{11}^{(2)} + a_2^{(1)} w_{12}^{(2)} + b_1^{(2)}$	$a_1^{(2)} = \mathrm{sigmoid}(z_1^{(2)})$
은닉 층	$z_2^{(2)} = a_1^{(1)} w_{21}^{(2)} + a_2^{(1)} w_{22}^{(2)} + b_2^{(2)}$	$a_2^{(2)} = \mathrm{sigmoid}(z_2^{(2)})$
출력 층	$z_1^{(3)} = a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}$	$a_1^{(3)} = \mathrm{sigmoid}(z_1^{(3)})$
출력 층	$z_2^{(3)} = a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}$	$a_2^{(3)} = \mathrm{sigmoid}(z_2^{(3)})$

- **시그모이드 함수 미분**

$$\frac{\partial \mathrm{sigmoid}(z)}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) \quad ①$$

$$= \frac{e^{-z}}{(1+e^{-z})^2} \quad ②$$

$$= \frac{1}{1+e^{-z}} \times \frac{e^{-z}}{1+e^{-z}} \quad ③$$

$$= \frac{1}{1+e^{-z}} \times \frac{(1+e^{-z})-1}{1+e^{-z}} \quad ④$$

분자에 1을 더하고 다시 빼 주어도
분자식은 변하지 않음

$$= \frac{1}{1+e^{-z}} \times \left(1 - \frac{1}{1+e^{-z}}\right) \quad ⑤$$

$$= \mathrm{sigmoid}(z) \times (1 - \mathrm{sigmoid}(z)) \quad ⑥$$

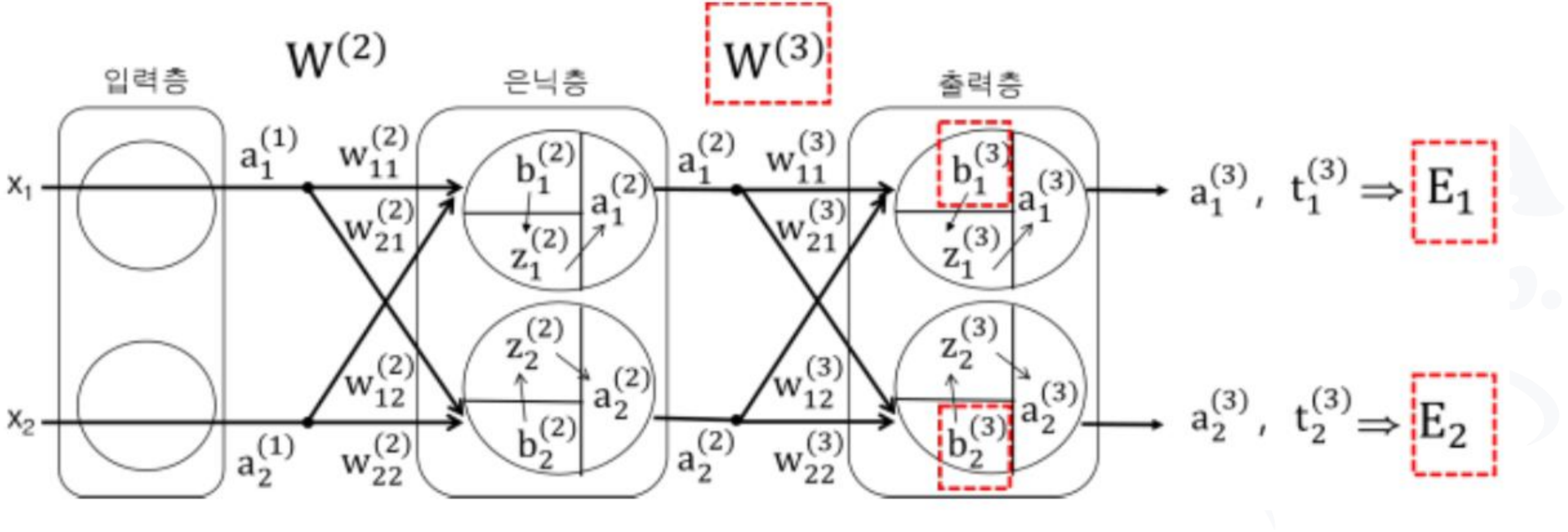
오차 역전파와 관련된 수학 공식 유도

- 시그모이드 함수 미분

$$\frac{\partial \mathrm{sigmoid(z)}}{\partial \mathrm{z}} = \mathrm{sigmoid(z)} \times (1 - \mathrm{sigmoid(z)})$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 오차 역전파를 위한 기본 신경망 구조(출력 층)



오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력층의 오차역전파(Back Propagation)를 구한다는 것은
  - 출력층에 적용되는 가중치 W (3) / 바이어스 b (3) 가 변할 때
  - 최종 오차 E (E=E 1 +E 2) 는 얼마나 변하는지 알 수 있는 공식을 구하는 것

  - 출력층 가중치 W (3) / 바이어스 b (3) 업데이트

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

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 그런데 출력 층에서의 가중치 $\boldsymbol{W}^{(3)}$ / 편향치 $\boldsymbol{b}^{(3)}$ 는
  - 가중치 $\boldsymbol{W}^{(3)}$ 는 2x2 크기의 행렬이며, 편향치 $\boldsymbol{b}^{(3)}$는 크기가 2인 벡터이므로

$$\mathrm{W}^{(3)} = \begin{pmatrix} \mathrm{w}_{11}^{(3)} & \mathrm{w}_{21}^{(3)} \\ \mathrm{w}_{12}^{(3)} & \mathrm{w}_{22}^{(3)} \end{pmatrix} \quad , \quad \mathrm{b}^{(3)} = \begin{pmatrix} \mathrm{b}_1^{(3)} & \mathrm{b}_2^{(3)} \end{pmatrix}$$

출력층 가중치 $\boldsymbol{W}^{(3)}$ / 편향치 $\boldsymbol{b}^{(3)}$

오차 역전파와 관련된 수학 공식 유도

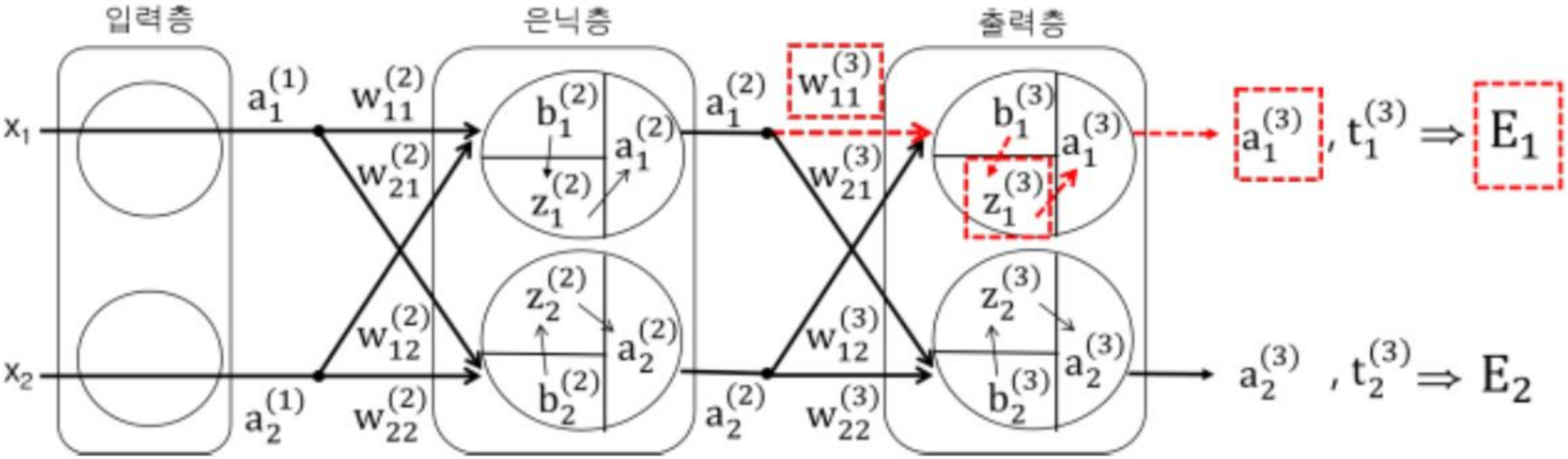
- 출력 층의 오차 역전파 공식
  - 위와 같이 출력 층의 가중치 $W^{(3)}$는 4 개의 요소 값을 가지며, 편향치 $b^{(3)}$는 2 개의 요소 값을 가짐
  - 따라서 편미분으로 나타나는 2 개의 항 $\frac{\partial E}{\partial W^{(3)}}$, $\frac{\partial E}{\partial b^{(3)}}$은 총 6 개의 편미분 식으로 분해됨

출력층 가중치 편미분 $\frac{\partial E}{\partial W^{(3)}}$, 바이어스 편미분 $\frac{\partial E}{\partial b^{(3)}}$ 분해

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

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{11}^{(3)}}$



$w_{11}^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- **출력 층의 오차 역전파 공식**
  - **출력 층 $\frac{\partial E}{\partial w_{11}^{(3)}}$ 유도 과정**

$w_{11}^{(3)}$ 이 변할 때,
오차 E가 얼마나
변할 것인가?

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

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial w_{11}^{(3)}} \quad ②$$

$$= \frac{\partial\left\{\frac{1}{2}\left(t_1^{(3)} - a_1^{(3)}\right)^2\right\}}{\partial a_1^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_1^{(3)})}{\partial z_1^{(3)}} \times \frac{\partial(a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)})}{\partial w_{11}^{(3)}} \quad ③$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \times \mathrm{sigmoid}\left(z_1^{(3)}\right) \times \left(1 - \mathrm{sigmoid}\left(z_1^{(3)}\right)\right) \times a_1^{(2)} \quad ④$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \times a_1^{(3)} \times \left(1 - a_1^{(3)}\right) \times a_1^{(2)} \quad ⑤$$

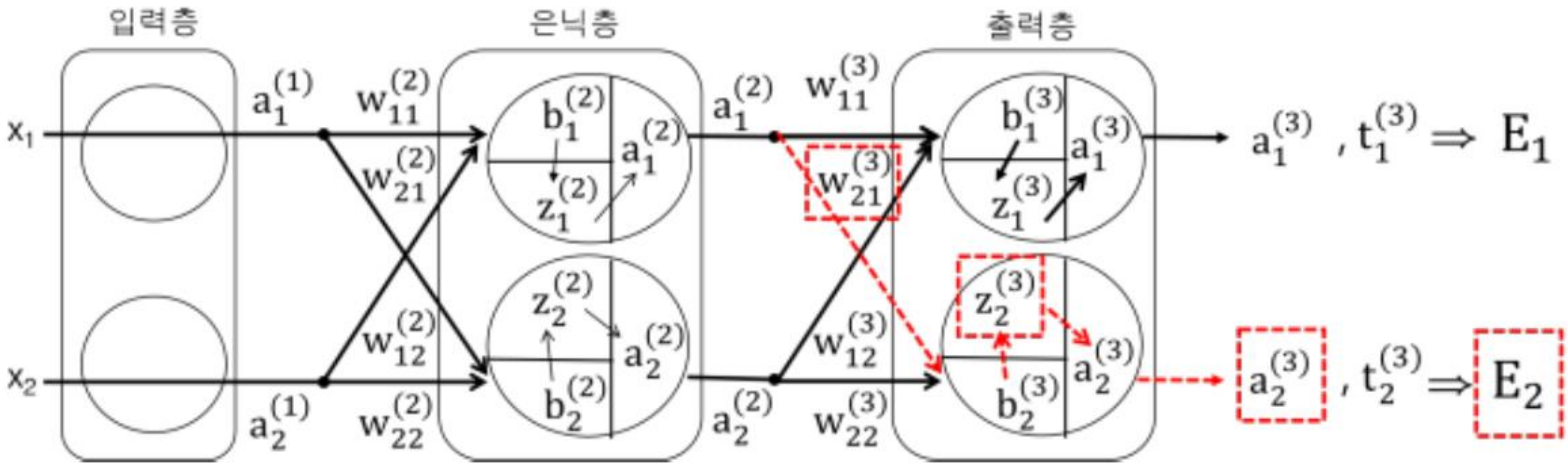
오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{11}^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial w_{11}^{(3)}} = (a_1^{(3)} - t_1^{(3)}) \times a_1^{(3)} \times (1 - a_1^{(3)}) \times a_1^{(2)}$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{21}^{(3)}}$



$w_{21}^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{21}^{(3)}}$ 유도 과정

$w_{21}^{(3)}$ 이 변할 때, 오차 E가 얼마나 변할 것인가?

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

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial w_{21}^{(3)}} \quad ②$$

$$= \frac{\partial \left\{ \frac{1}{2} \left(t_2^{(3)} - a_2^{(3)} \right)^2 \right\}}{\partial a_2^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_2^{(3)})}{\partial z_2^{(3)}} \times \frac{\partial (a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)})}{\partial w_{21}^{(3)}} \quad ③$$

$$= \left(a_2^{(3)} - t_2^{(3)} \right) \times \mathrm{sigmoid}\left(z_2^{(3)} \right) \times \left(1 - \mathrm{sigmoid}\left(z_2^{(3)} \right) \right) \times a_1^{(2)} \quad ④$$

$$= (a_2^{(3)} - t_2^{(3)}) \times a_2^{(3)} \times (1 - a_2^{(3)}) \times a_1^{(2)} \quad ⑤$$

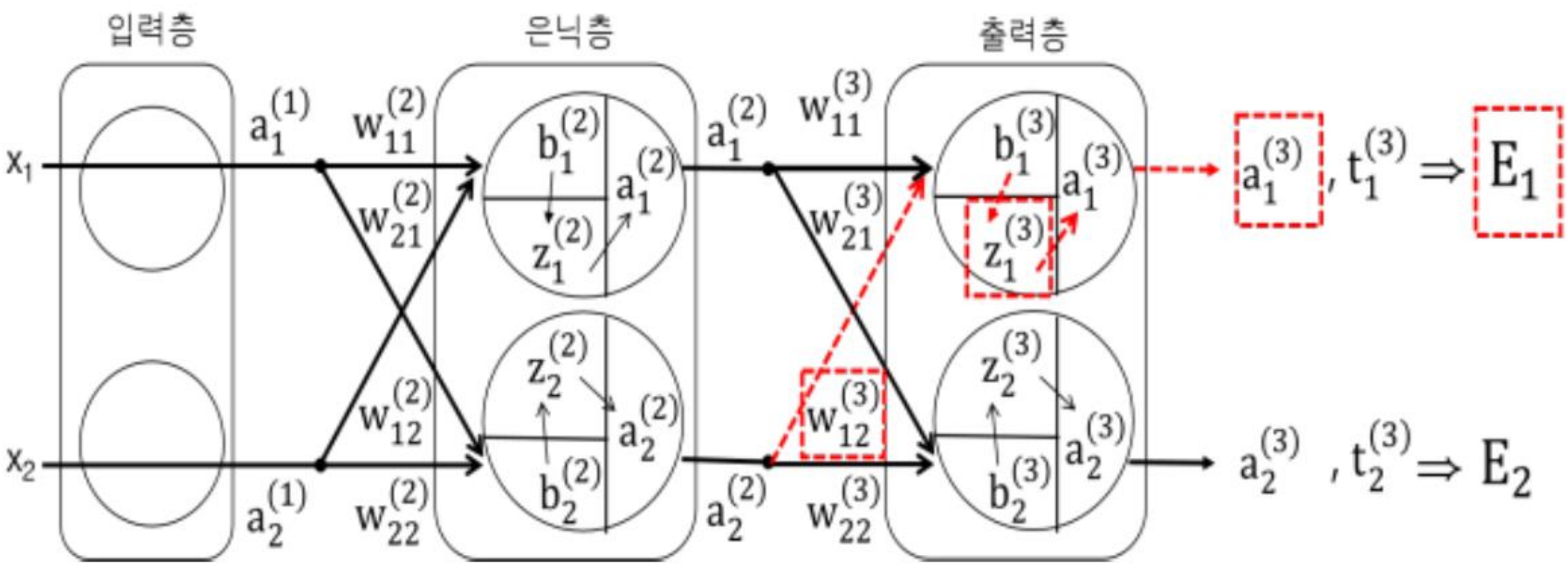
오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{21}^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial w_{21}^{(3)}} = (a_2^{(3)} - t_2^{(3)}) \times a_2^{(3)} \times (1 - a_2^{(3)}) \times a_1^{(2)}$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{12}^{(3)}}$



$w_{12}^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- **출력 층의 오차 역전파 공식**
  - **출력 층 $\frac{\partial E}{\partial w_{12}^{(3)}}$ 유도 과정**

$w_{12}^{(3)}$ 이 변할 때, 오차 E가 얼마나 변할 것인가?

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

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial w_{12}^{(3)}} \quad ②$$

$$= \frac{\partial \left\{ \frac{1}{2} \left(t_1^{(3)} - a_1^{(3)} \right)^2 \right\}}{\partial a_1^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_1^{(3)})}{\partial z_1^{(3)}} \times \frac{\partial (a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)})}{\partial w_{12}^{(3)}} \quad ③$$

$$= \left(a_1^{(3)} - t_1^{(3)} \right) \times \mathrm{sigmoid}\left(z_1^{(3)} \right) \times \left(1 - \mathrm{sigmoid}\left(z_1^{(3)} \right) \right) \times a_2^{(2)} \quad ④$$

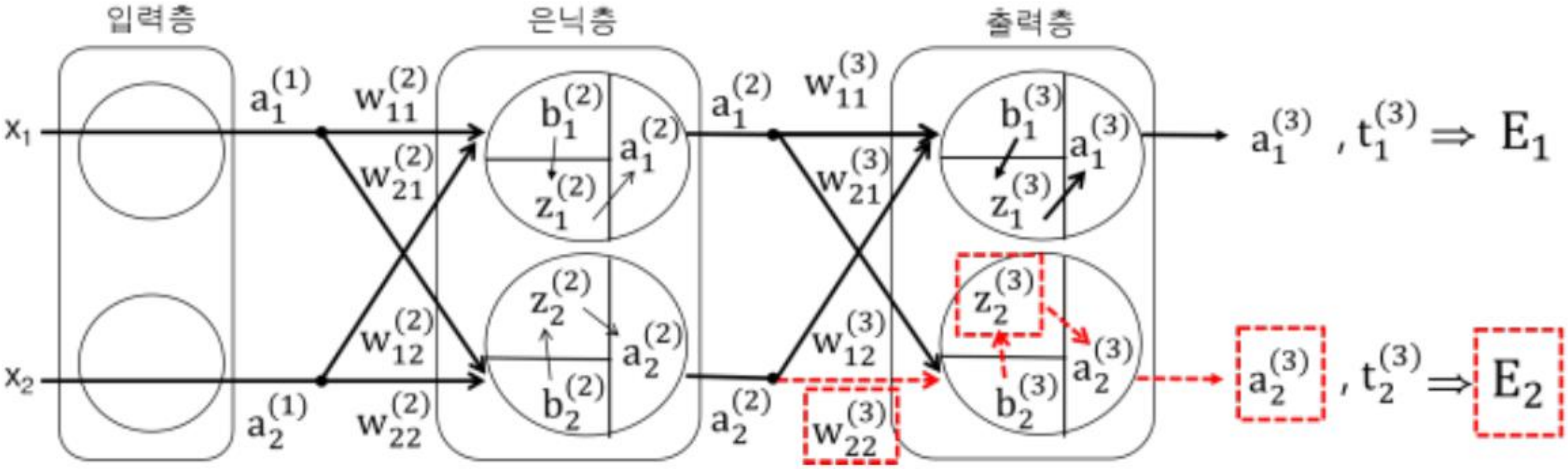
$$= \left(a_1^{(3)} - t_1^{(3)} \right) \times a_1^{(3)} \times \left(1 - a_1^{(3)} \right) \times a_2^{(2)} \quad ⑤$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{12}^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial w_{12}^{(3)}} = (a_1^{(3)} - t_1^{(3)}) \times a_1^{(3)} \times (1 - a_1^{(3)}) \times a_2^{(2)}$$

- 출력 층의 오차 역전파 공식
  - 출력 층 $\dfrac{\partial E}{\partial w_{22}^{(3)}}$



$w_{22}^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- **출력 층의 오차 역전파 공식**
  - **출력 층 $\frac{\partial E}{\partial w_{22}^{(3)}}$ 유도 과정**

$w_{22}^{(3)}$ 이 변할 때, 오차 E가 얼마나 변할 것인가?

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

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial w_{22}^{(3)}} \quad ②$$

$$= \frac{\partial\left\{\frac{1}{2}\left(t_2^{(3)} - a_2^{(3)}\right)^2\right\}}{\partial a_2^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_2^{(3)})}{\partial z_2^{(3)}} \times \frac{\partial(a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)})}{\partial w_{22}^{(3)}} \quad ③$$

$$= \left(a_2^{(3)} - t_2^{(3)}\right) \times \mathrm{sigmoid}\left(z_2^{(3)}\right) \times \left(1 - \mathrm{sigmoid}\left(z_2^{(3)}\right)\right) \times a_2^{(2)} \quad ④$$

$$= \left(a_2^{(3)} - t_2^{(3)}\right) \times a_2^{(3)} \times \left(1 - a_2^{(3)}\right) \times a_2^{(2)} \quad ⑤$$

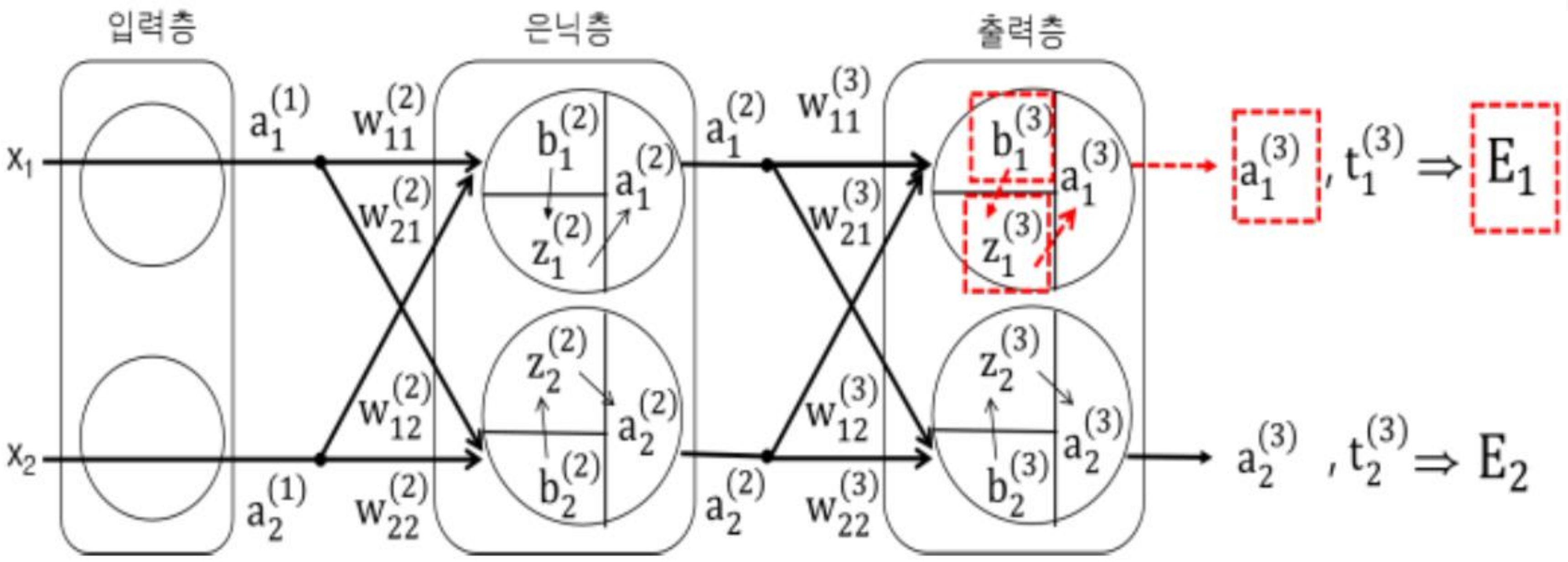
오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial w_{22}^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial w_{22}^{(3)}} = (a_2^{(3)} - t_2^{(3)}) \times a_2^{(3)} \times (1 - a_2^{(3)}) \times a_2^{(2)}$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial b_1^{(3)}}$



$b_1^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- **출력 층의 오차 역전파 공식**
  - **출력 층 $\frac{\partial E}{\partial b_1^{(3)}}$ 유도 과정**

$b_1^{(3)}$ 이 변할 때, 오차 E가 얼마나 변할 것인가?

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

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial b_1^{(3)}} \quad ②$$

$$= \frac{\partial \left\{ \frac{1}{2} \left(t_1^{(3)} - a_1^{(3)} \right)^2 \right\}}{\partial a_1^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_1^{(3)})}{\partial z_1^{(3)}} \times \frac{\partial (a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)})}{\partial b_1^{(3)}} \quad ③$$

$$= \left(a_1^{(3)} - t_1^{(3)} \right) \times \mathrm{sigmoid}\left(z_1^{(3)} \right) \times \left(1 - \mathrm{sigmoid}\left(z_1^{(3)} \right) \right) \times 1 \quad ④$$

$$= (a_1^{(3)} - t_1^{(3)}) \times a_1^{(3)} \times (1 - a_1^{(3)}) \times 1 \quad ⑤$$

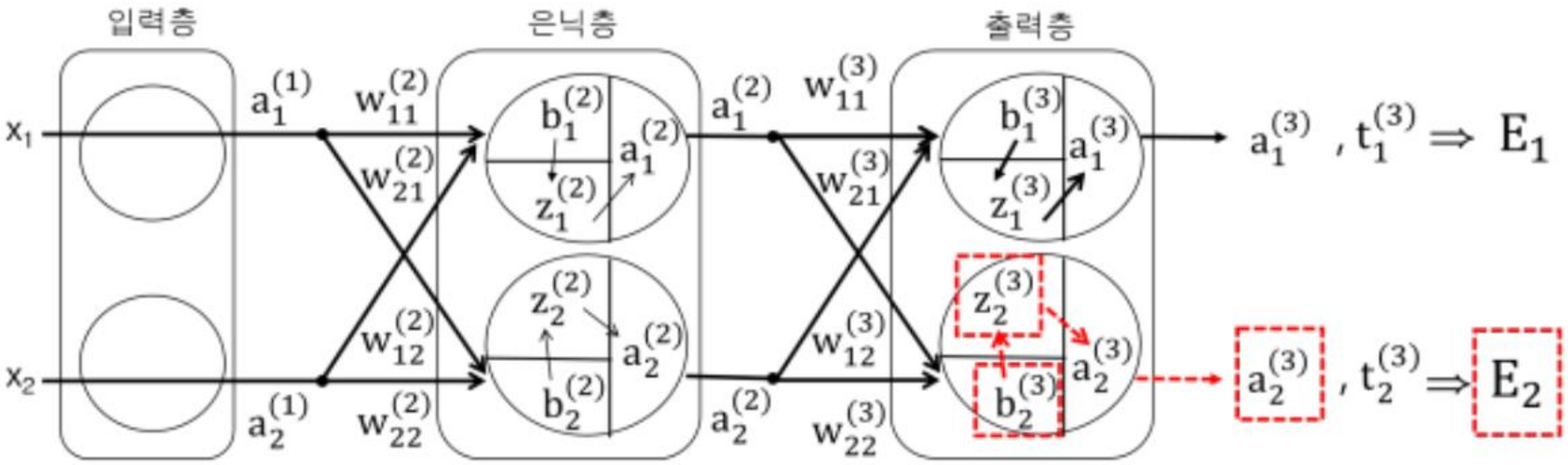
오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial b_1^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial b_1^{(3)}} = (a_1^{(3)} - t_1^{(3)}) \times a_1^{(3)} \times (1 - a_1^{(3)}) \times 1$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial b_2^{(3)}}$



$b_2^{(3)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial b_2^{(3)}}$ 유도 과정

$b_2^{(3)}$ 이 변할 때, 오차 E가 얼마나 변할 것인가?

$$\frac{\partial E}{\partial b_2^{(3)}} = \frac{\partial E_1}{\partial b_2^{(3)}} + \frac{\partial E_2}{\partial b_2^{(3)}} \qquad ①$$

$$= \frac{\partial E_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial b_2^{(3)}} \qquad ②$$

$$= \frac{\partial \left\{ \frac{1}{2} \left(t_2^{(3)} - a_2^{(3)} \right)^2 \right\}}{\partial a_2^{(3)}} \times \frac{\partial \mathrm{sigmoid}(z_2^{(3)})}{\partial z_2^{(3)}} \times \frac{\partial (a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)})}{\partial b_2^{(3)}} \qquad ③$$

$$= \left(a_2^{(3)} - t_2^{(3)} \right) \times \mathrm{sigmoid}\left(z_2^{(3)} \right) \times \left(1 - \mathrm{sigmoid}\left(z_2^{(3)} \right) \right) \times 1 \qquad ④$$

$$= \left(a_2^{(3)} - t_2^{(3)} \right) \times a_2^{(3)} \times \left(1 - a_2^{(3)} \right) \times 1 \qquad ⑤$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층의 오차 역전파 공식
  - 출력 층 $\frac{\partial E}{\partial b_2^{(3)}}$ 값은 아래와 같이 곱셈 형태의 수식으로 바뀜

$$\frac{\partial E}{\partial b_2^{(3)}} = (a_2^{(3)} - t_2^{(3)}) \times a_2^{(3)} \times (1 - a_2^{(3)}) \times 1$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층 오차역전파 일반공식
  - 지금까지 유도한 가중치 4개 ($w_{11}^{(3)}$ $w_{21}^{(3)}$ $w_{12}^{(3)}$ $w_{22}^{(3)}$), 편향치 2개 ($b_1^{(3)}$ $b_2^{(3)}$)에 대한 편미분 식을
  - 출력 층의 가상의 손실(loss) 개념을 이용해서
  - 행렬(matrix) 식으로 나타낼 수 있는
  - 출력 층의 최종적인 오차역전파 공식을 유도

오차 역전파와 관련된 수학 공식 유도

- 출력 층 오차역전파 일반공식
  - 먼저 은닉 층 출력 값 벡터 A2 와 출력 층에서의 가상의 손실을 나타내는 벡터 loss_3 정의

은닉 층 출력 값 벡터	$A2 = \left(a_1^{(2)} \quad a_2^{(2)} \right)$
출력 층 가상 손실 벡터	$loss\_3 = \left(\left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) \quad \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) \right)$

  - A2, loss_3 가 정의 되었다면, 우리가 지금까지 유도했던 출력층의 오차 역전파 공식 6 개 ($\frac{\partial E}{\partial w_{11}^{(3)}}$, $\frac{\partial E}{\partial w_{21}^{(3)}}$, $\frac{\partial E}{\partial w_{12}^{(3)}}$, $\frac{\partial E}{\partial w_{22}^{(3)}}$, $\frac{\partial E}{\partial b_1^{(3)}}$, $\frac{\partial E}{\partial b_2^{(3)}}$) 를 이용하여 $\frac{\partial E}{\partial W^{(3)}}$, $\frac{\partial E}{\partial b^{(3)}}$ 를 유도

- 출력 층 오차역전파 일반공식
  - 출력 층 가중치 변화율에 따른 오차 변화율 $\frac{\partial E}{\partial W^{(3)}}$

$$\frac{\partial E}{\partial W^{(3)}} = \begin{pmatrix} \frac{\partial E}{\partial w_{11}^{(3)}} & \frac{\partial E}{\partial w_{21}^{(3)}} \\ \frac{\partial E}{\partial w_{12}^{(3)}} & \frac{\partial E}{\partial w_{22}^{(3)}} \end{pmatrix} \quad ①$$

$$= \begin{pmatrix} (a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)})a_1^{(2)} & (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)})a_1^{(2)} \\ (a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)})a_2^{(2)} & (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)})\, a_2^{(2)} \end{pmatrix} \quad ②$$

$$= \begin{pmatrix} a_1^{(2)}(a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) & a_1^{(2)}(a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}) \\ a_2^{(2)}(a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) & a_2^{(2)}(a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}) \end{pmatrix} \quad ③$$

$$= \begin{pmatrix} a_1^{(2)} \\ a_2^{(2)} \end{pmatrix} \cdot \begin{pmatrix} (a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) & (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}) \end{pmatrix} \quad ④$$

$$= \ A2^{T} \ \cdot \ \text{loss\_3} \quad ⑤$$

오차 역전파와 관련된 수학 공식 유도

- 출력 층 오차역전파 일반공식
  - 출력 층 편향치 변화율에 따른 오차 변화율 $\frac{\partial E}{\partial b^{(3)}}$

$$\frac{\partial E}{\partial b^{(3)}} = \left(\frac{\partial E}{\partial b_1^{(3)}} \quad \frac{\partial E}{\partial b_2^{(3)}} \right) \quad ①$$

$$= \left((a_1^{(3)} - t_1^{(3)})a_1^{(3)}(1 - a_1^{(3)}) \quad (a_2^{(3)} - t_2^{(3)})a_2^{(3)}(1 - a_2^{(3)}) \right) \quad ②$$

$$= \text{loss\_3} \quad ③$$

  - 출력 층에서의 가중치 변화율 $\frac{\partial E}{\partial W^{(3)}}$는 은닉 층의 출력 값에 대한 전치 행렬 $(A2^T)$과 출력 층의 가상 손실(loss_3) 와의 행렬 곱으로 계산되며
  - 출력 층에서의 편향치 변화율 $\frac{\partial E}{\partial b^{(3)}}$는 출력 층의 가상 손실(loss_3)로 나타남

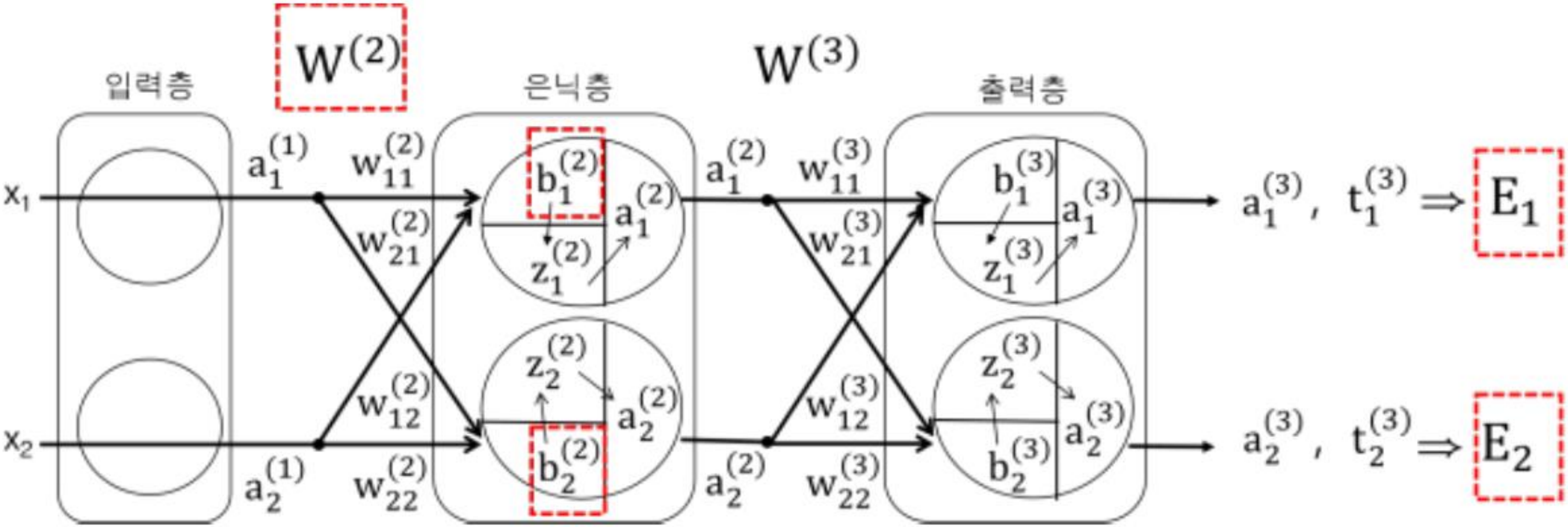
오차 역전파와 관련된 수학 공식 유도

- 출력 층 오차역전파 일반공식
  - 오차역전파를 이용한 출력 층의 가중치 $\mathbf{W}^{(3)}$, 편향치 $\mathbf{b}^{(3)}$ 계산

출력 층 가중치 $W^{(3)}$ 계산	$W^{(3)} = W^{(3)} - \alpha \frac{\partial E}{\partial W^{(3)}} = W^{(3)} - \alpha \times (A2^{T} \cdot loss\_3)$
출력 층 바이어스 $b^{(3)}$ 계산	$b^{(3)} = b^{(3)} - \alpha \frac{\partial E}{\partial b^{(3)}} = b^{(3)} - \alpha \times loss\_3$

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식



오차역전파를 위한 기본 신경망 구조 (은닉층)

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식

은닉층 가중치 $W^{(2)}$ / 바이어스 $b^{(2)}$ 업데이트

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

은닉층 가중치 $W^{(2)}$ / 바이어스 $b^{(2)}$

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

오차 역전파와 관련된 수학 공식 유도

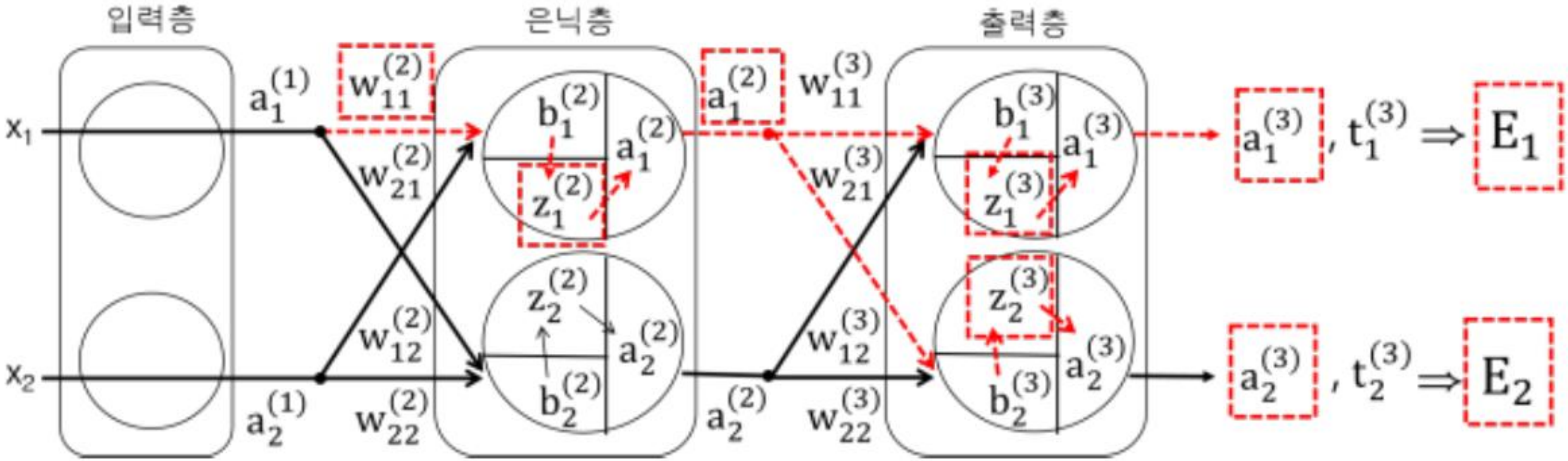
- 은닉 층 오차역전파 공식

은닉층 가중치 편미분 $\frac{\partial E}{\partial W^{(2)}}$, 바이어스 편미분 $\frac{\partial E}{\partial b^{(2)}}$ 분해

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

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{11}^{(2)}}$



$w_{11}^{(2)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{11}^{(2)}}$

은닉층 $\frac{\partial E}{\partial w_{11}^{(2)}}$ 유도과정

$$\frac{\partial E}{\partial w_{11}^{(2)}} = \frac{\partial E_1}{\partial w_{11}^{(2)}} + \frac{\partial E_2}{\partial w_{11}^{(2)}} \quad ①$$

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

$$= (a_1^{(3)} - t_1^{(3)}) \times \mathrm{sigmoid}(z_1^{(3)})\left(1 - \mathrm{sigmoid}(z_1^{(3)})\right) \times w_{11}^{(3)} \times \mathrm{sigmoid}(z_1^{(2)})\left(1 - \mathrm{sigmoid}(z_1^{(2)})\right) \times a_1^{(1)} \quad ③$$

$$+ (a_2^{(3)} - t_2^{(3)}) \times \mathrm{sigmoid}(z_2^{(3)})\left(1 - \mathrm{sigmoid}(z_2^{(3)})\right) \times w_{21}^{(3)} \times \mathrm{sigmoid}(z_1^{(2)})\left(1 - \mathrm{sigmoid}(z_1^{(2)})\right) \times a_1^{(1)}$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \times a_1^{(3)}\left(1 - a_1^{(3)}\right) \times w_{11}^{(2)} \times a_1^{(2)}\left(1 - a_1^{(2)}\right) \times a_1^{(1)} \quad ④$$

$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \times a_2^{(3)}\left(1 - a_2^{(3)}\right) \times w_{21}^{(2)} \times a_1^{(2)}\left(1 - a_1^{(2)}\right) \times a_1^{(1)}$$

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{11}^{(2)}}$

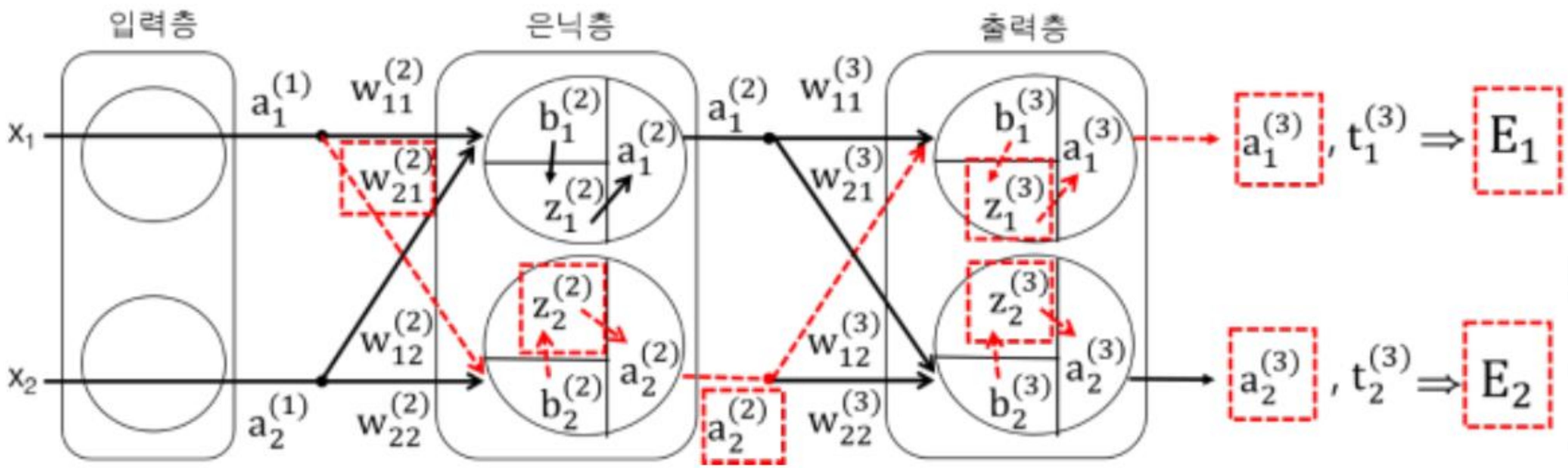
은닉층 $\frac{\partial E}{\partial w_{11}^{(2)}}$ 오차역전파 공식

$$\frac{\partial E}{\partial w_{11}^{(2)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \times a_1^{(3)}\left(1 - a_1^{(3)}\right) \times w_{11}^{(2)} \times a_1^{(2)}\left(1 - a_1^{(2)}\right) \times a_1^{(1)}$$

$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \times a_2^{(3)}\left(1 - a_2^{(3)}\right) \times w_{21}^{(2)} \times a_1^{(2)}\left(1 - a_1^{(2)}\right) \times a_1^{(1)}$$

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{21}^{(2)}}$



$w_{21}^{(2)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{21}^{(2)}}$

은닉층 $\frac{\partial E}{\partial w_{21}^{(2)}}$ 유도과정

$$\frac{\partial E}{\partial w_{21}^{(2)}} = \frac{\partial E_1}{\partial w_{21}^{(2)}} + \frac{\partial E_2}{\partial w_{21}^{(2)}} \qquad ①$$

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial a_2^{(2)}} \times \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \times \frac{\partial z_2^{(2)}}{\partial w_{21}^{(2)}} + \frac{\partial E_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial a_2^{(2)}} \times \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \times \frac{\partial z_2^{(2)}}{\partial w_{21}^{(2)}} \qquad ②$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \times \mathrm{sigmoid}\left(z_1^{(3)}\right)\left(1 - \mathrm{sigmoid}\left(z_1^{(3)}\right)\right) \times w_{12}^{(3)} \times \mathrm{sigmoid}\left(z_2^{(2)}\right)\left(1 - \mathrm{sigmoid}\left(z_2^{(2)}\right)\right) \times a_1^{(1)} \qquad ③$$
$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \times \mathrm{sigmoid}\left(z_2^{(3)}\right)\left(1 - \mathrm{sigmoid}\left(z_2^{(3)}\right)\right) \times w_{22}^{(3)} \times \mathrm{sigmoid}\left(z_2^{(2)}\right)\left(1 - \mathrm{sigmoid}\left(z_2^{(2)}\right)\right) \times a_1^{(1)}$$

$$= \left(a_1^{(3)} - t_1^{(3)}\right) \times a_1^{(3)}\left(1 - a_1^{(3)}\right) \times w_{12}^{(3)} \times a_2^{(2)}\left(1 - a_2^{(2)}\right) \times a_1^{(1)} \qquad ④$$
$$+ \left(a_2^{(3)} - t_2^{(3)}\right) \times a_2^{(3)}\left(1 - a_2^{(3)}\right) \times w_{22}^{(3)} \times a_2^{(2)}\left(1 - a_2^{(2)}\right) \times a_1^{(1)}$$

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{21}^{(2)}}$

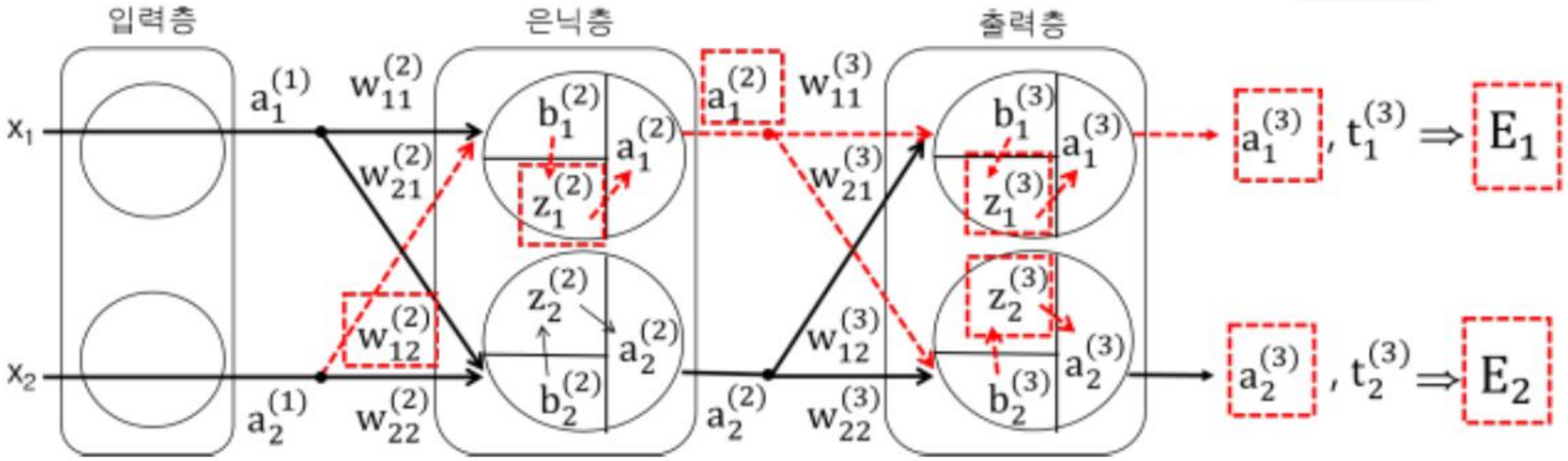
은닉층 $\frac{\partial E}{\partial w_{21}^{(2)}}$ 오차역전파 공식

$$\frac{\partial E}{\partial w_{21}^{(2)}} = \left(a_1^{(3)} - t_1^{(3)}\right) \times a_1^{(3)}\left(1 - a_1^{(3)}\right) \times w_{12}^{(3)} \times a_2^{(2)}\left(1 - a_2^{(2)}\right) \times a_1^{(1)}$$

$$+\left(a_2^{(3)} - t_2^{(3)}\right) \times a_2^{(3)}\left(1 - a_2^{(3)}\right) \times w_{22}^{(3)} \times a_2^{(2)}\left(1 - a_2^{(2)}\right) \times a_1^{(1)}$$

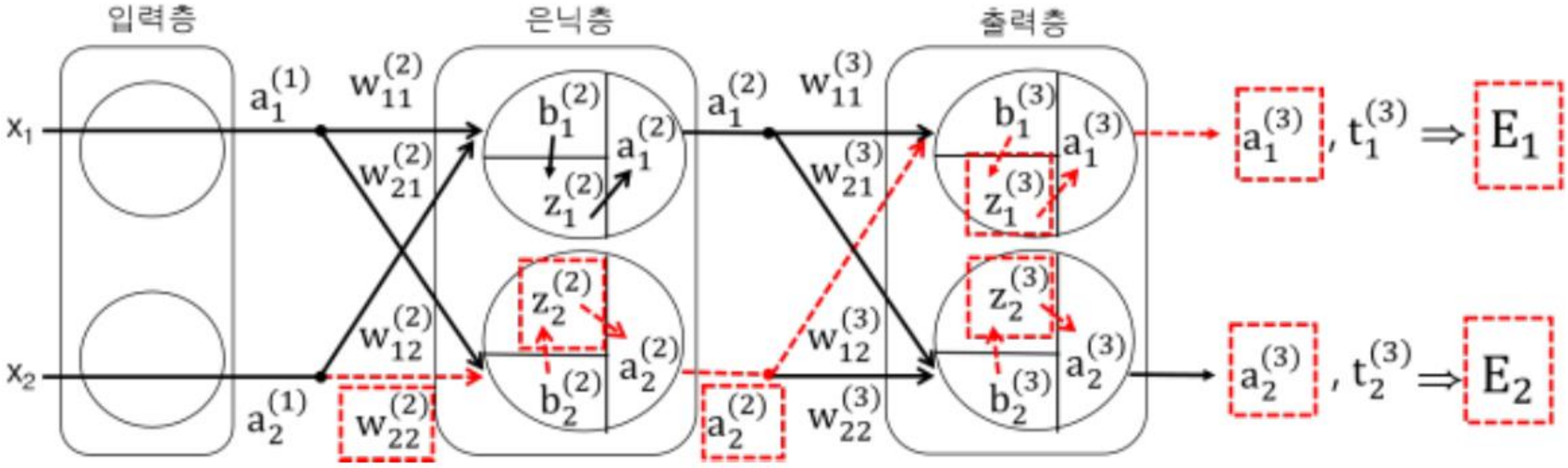
오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{12}^{(2)}}$



$w_{12}^{(2)}$ 변화에 따른 오차 E 변화

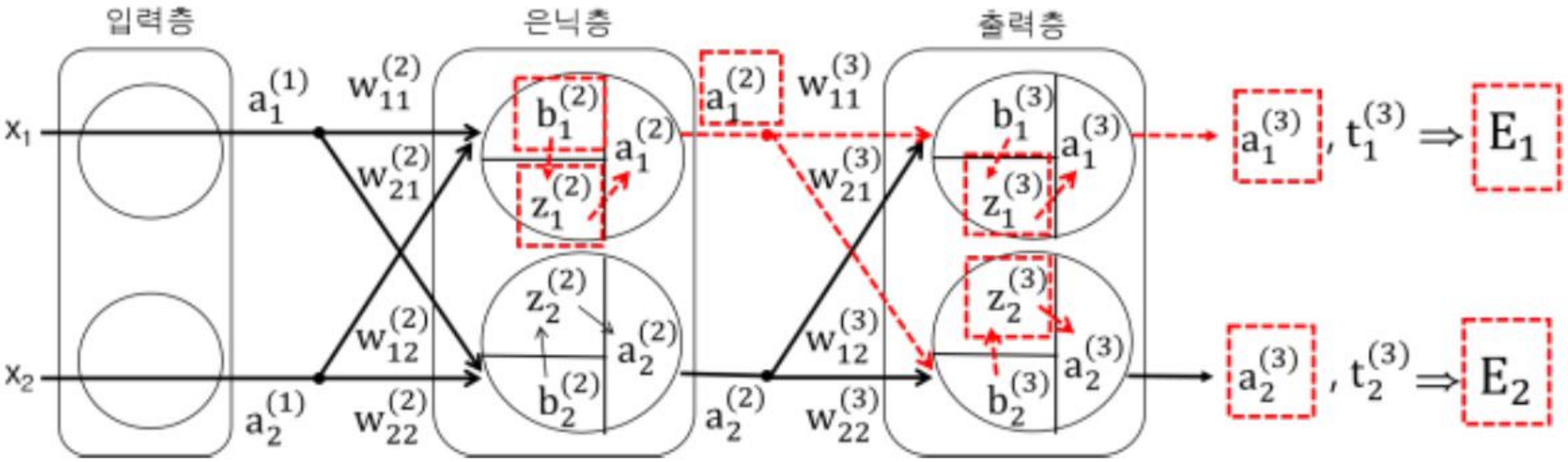
- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial w_{22}^{(2)}}$



$w_{22}^{(2)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

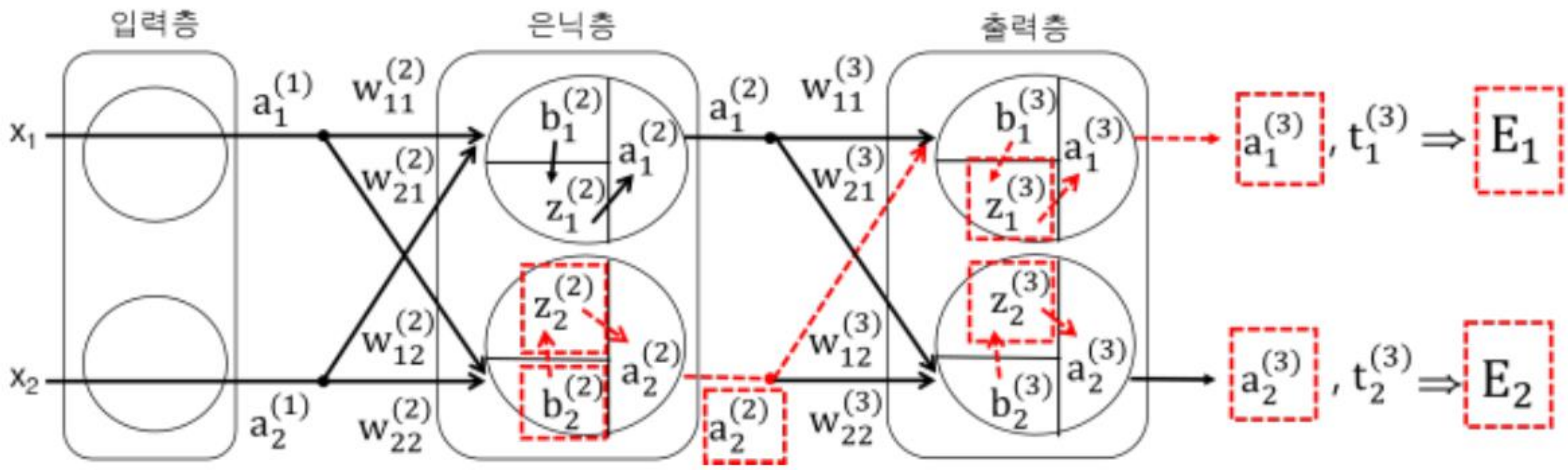
- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial b_1^{(2)}}$



$b_1^{(2)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- 은닉 층 오차역전파 공식
  - 은닉 층 $\frac{\partial E}{\partial b_2^{(2)}}$



$b_2^{(2)}$ 변화에 따른 오차 E 변화

오차 역전파와 관련된 수학 공식 유도

- **은닉 층 오차 역전파 최종 공식**

입력 층 출력 A1 / 은닉 층 출력 A2

입력 층 출력 값 벡터	$A1 = \left(a_1^{(1)} \quad a_2^{(1)} \right)$
은닉 층 출력 값 벡터	$A2 = \left(a_1^{(2)} \quad a_2^{(2)} \right)$

은닉 층 가상 손실 loss_2 / 출력 층 가상 손실 loss_3

출력 층 가상 손실 벡터	$loss\_3 = \left(\left(a_1^{(3)} - t_1^{(3)}\right) a_1^{(3)} \left(1 - a_1^{(3)}\right) \quad \left(a_2^{(3)} - t_2^{(3)}\right) a_2^{(3)} \left(1 - a_2^{(3)}\right) \right)$
출력 층 가중치	$W3 = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$
은닉 층 가상 손실 벡터	$loss\_2 = \left(loss\_3 \cdot \ W3^T \right) \times A2(1 - A2)$

오차 역전파와 관련된 수학 공식 유도

- **은닉 층 오차 역전파 최종 공식**

은닉 층 가중치 변화율에 따른 오차 변화율 $\frac{\partial E}{\partial W^{(2)}}$

$$\frac{\partial E}{\partial W^{(2)}} = \begin{pmatrix} \frac{\partial E}{\partial w_{11}^{(2)}} & \frac{\partial E}{\partial w_{21}^{(2)}} \\ \frac{\partial E}{\partial w_{12}^{(2)}} & \frac{\partial E}{\partial w_{22}^{(2)}} \end{pmatrix} \quad ①$$

$$= A1^{T} \cdot \left(\left(\text{loss\_3} \cdot W3^{T} \right) \times A2(1 - A2) \right) \quad ②$$

$$= A1^{T} \cdot \text{loss\_2} \quad ③$$

오차 역전파와 관련된 수학 공식 유도

- **은닉 층 오차 역전파 최종 공식**

은닉 층 바이어스 변화율에 따른 오차 변화율 $\frac{\partial E}{\partial b^{(2)}}$

$$\frac{\partial E}{\partial b^{(2)}} = \left(\frac{\partial E}{\partial b_1^{(2)}} \quad \frac{\partial E}{\partial b_2^{(2)}} \right) \quad ①$$

$$= \left(\left(loss\_3 \cdot W3^T \right) \times A2(1 - A2) \right) \quad ②$$

$$= loss\_2 \quad ③$$

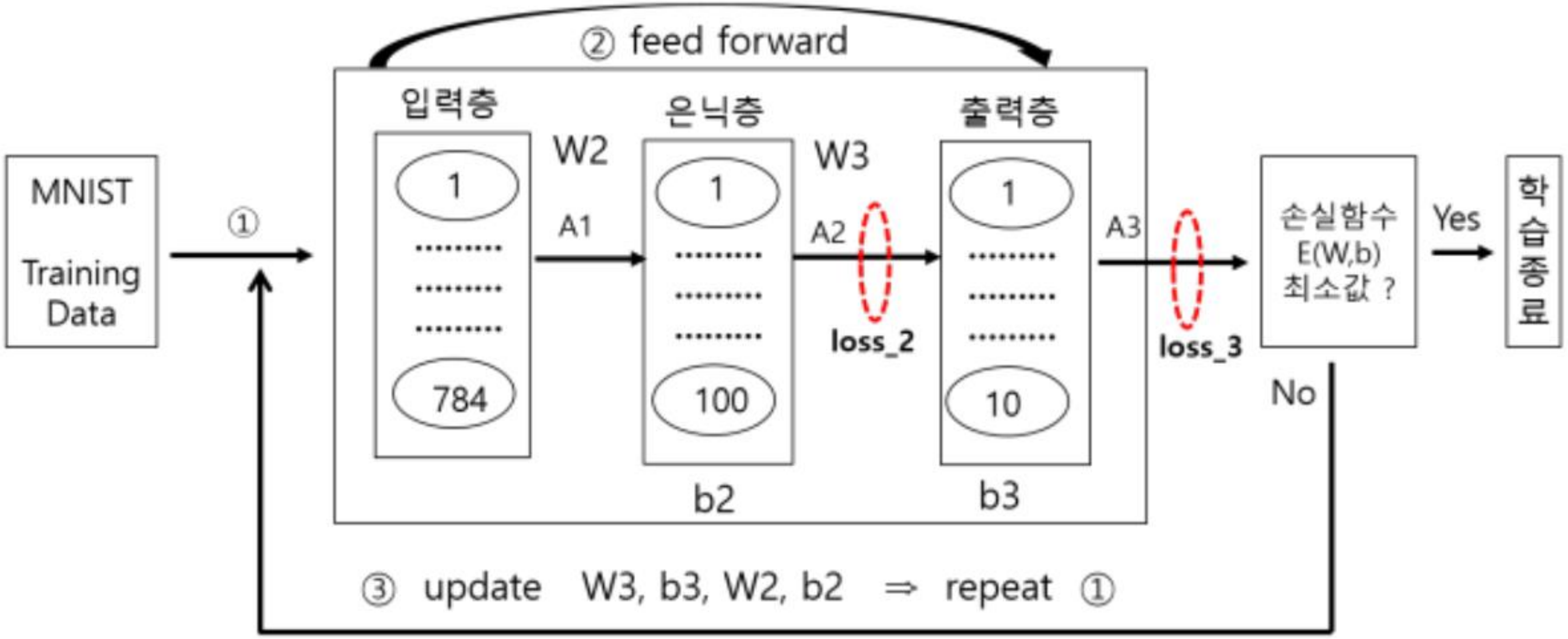
오차 역전파와 관련된 수학 공식 유도

- 오차역전파를 이용한 은닉 층의 가중치 $W^{(2)}$, 편향치 $b^{(2)}$ 계산

은닉 층 가중치 $W^{(2)}$ 계산	$W^{(2)} = W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}} = W^{(2)} - \alpha \times (A1^{T} \cdot loss\_2)$
은닉 층 바이어스 $b^{(2)}$ 계산	$b^{(2)} = b^{(2)} - \alpha \frac{\partial E}{\partial b^{(2)}} = b^{(2)} - \alpha \times loss\_2$

오차 역전파 기반 딥러닝 아키텍처

- 오차 역전파를 이용한 신경망 아키텍처



$W3 = W3 - \alpha \times (A2^T \cdot loss\_3)$

$b3 = b3 - \alpha \times loss\_3$

$W2 = W2 - \alpha \times (A1^T \cdot loss\_2)$

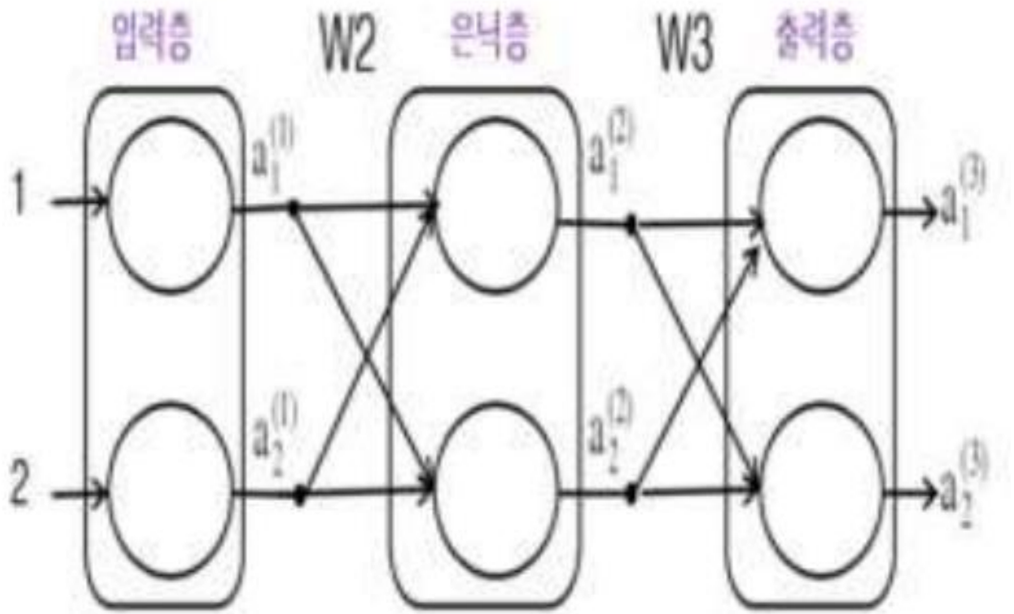
$b2 = b2 - \alpha \times loss\_2$

- **오차 역전파를 이용한 신경망 아키텍처**
  - **Neural Network 클래스**
    - def __init__(self, input_nodes, hidden_nodes, output_nodes, leraning_rate): 생성자
    - def feed_forward(self): 순전파 프로세스
    - def loss_val(self): 손실함수 계산 메소드
    - def train(self): 학습 메소드→가중치, 편향치 갱신 → 오류 역전파 수행
    - def predict(self, input_data): 예측 메소드
    - def accuracy(self, input_data, target_data): 정확도 측정 메소드

오차 역전파 실습 문제

- [문제 1] 아래의 가중치/바이어스/정답/학습율 초기 값과 가상의 시그모이드 함수는 다음과 같다
  - 1) 주어진 입력 값, 가중치 W2, W3, 바이어스 b2, b3, 가상의 sigmoid 함수를 이용하여 feed forward 1 회 수행 할 경우 $a_1^{(1)}$, $a_2^{(1)}$, $a_1^{(2)}$, $a_2^{(2)}$, $a_1^{(3)}$, $a_2^{(3)}$ 값을 계산하시오
  - 1)을 수행한 후에 back propagation 1 회 실행 할 경우 업데이트된 W3, b3, W2, b2를 계산 한 후에, 파이썬 코드로 결과를 확인하시오 (코드 구현시 A1, A2, A3, sigmoid 반환값 모두 행렬로 변환 후 계산하시오)

오차 역전파 실습 문제



$$W2 = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$$

$$W3 = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$T = (2 \quad 1)$$

$$b2 = (1 \quad 1)\ ,\ b3 = (0 \quad 2)$$

$$\text{learning\_rate} = 0.1$$

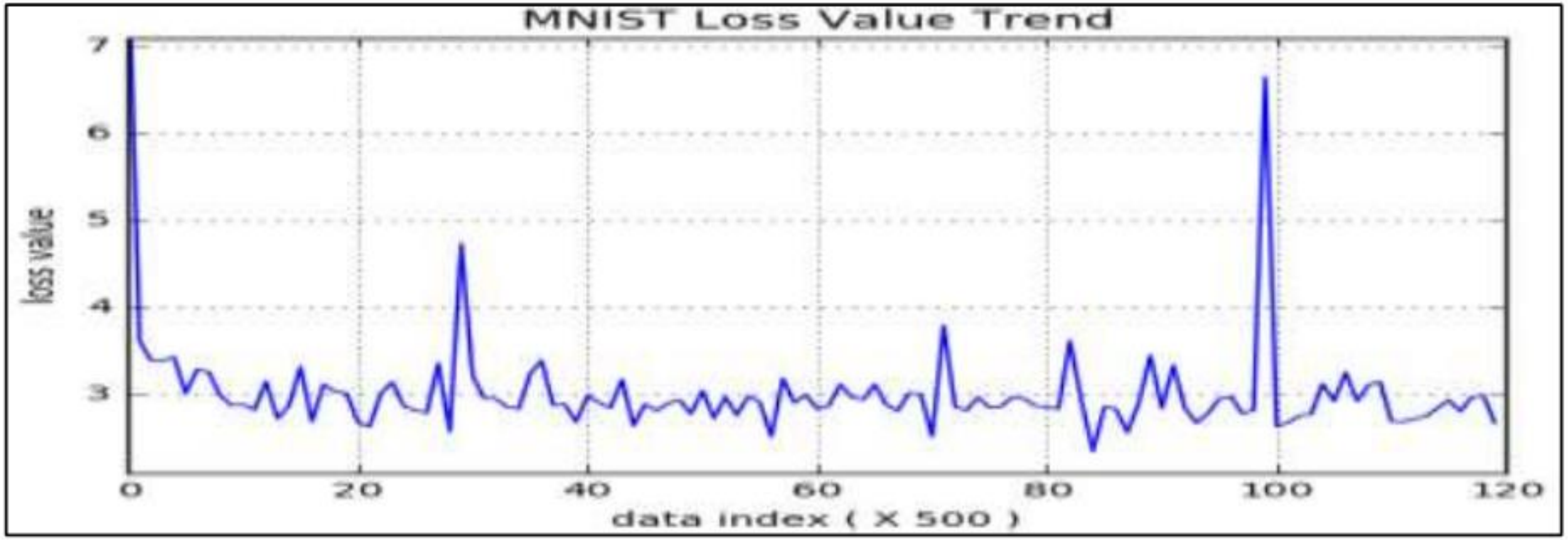
가상의
sigmoid 함수

sigmoid(0)	-3
sigmoid(1)	-2
sigmoid(2)	-1
sigmoid(3)	0
sigmoid(4)	1
sigmoid(5)	2
sigmoid(6)	3
sigmoid(7)	4

오차 역전파 실습 문제

- [문제 2] 오차역전파를 이용한 MNIST 검증
  - [은닉층 노드 100 개인 NeuralNetwork 객체 생성 및 학습]
    - obj = NeuralNetwork(i_nodes, h1_nodes, o_nodes, learning_rate)
    - 정규화 수행 후 입력데이타 / 정답데이터 분리 후, 반복횟수를 설정한 후
    - obj.train(input_data, target_data)
  - [정확도 검증 및 다음과 같은 손실함수 추세 확인]
    - obj.accuracy(test_input_data, test_target_data)

오차 역전파 실습 문제



MNIST Loss Value Trend
loss value
7
6
5
4
3
0
20
40
60
80
100
120
data index (X 500)

오차 역전파 실습 문제

- [문제 3] 오차역전파를 이용한 MNIST 검증
  - [은닉층 2 개를 가지는 NeuralNetwork 객체 생성 및 학습]
    - obj = NeuralNetwork(i_nodes, h1_nodes, h2_nodes, o_nodes, learning_rate)
  - [정확도 검증]
    - (accuracy_ret, false_list) = obj.accuracy(test_input_data, test_target_data)

오차 역전파 실습 문제

- [문제 4] 오차역전파를 이용한 MNIST 학습 overfitting 확인
  - [1] MNIST 학습을 위한 은닉층 1 개며 epochs = 20 을 가지는 객체 생성 및 학습
  - [2] validation 데이터를 이용하여 overfitting 확인
  - [3] epochs X training data 만큼 모든 데이터에 대해 손실함수 값 저장
  - [4] 손실함수 최대 값 , 최소 값 확인
  - [5] 정확도 검증
  - [6] 손실함수 최대 / 최소에 대응하는 실제 MNIST 데이터 확인
    - (해당 데이터의 정답과 예측 값 출력 및 이미지 확인)

오차 역전파 실습 문제

- **[문제 5] 오답에 대한 MNIST 데이터, 정답, 예측 값 확인**
  - **[오차역전파를 이용하여 NeuralNetwork 객체 생성 및 학습]**
    - obj = NeuralNetwork(…)
    - obj.train(…)
  - **[accuracy(…) 메서드는 다음과 같은 디버깅 정보 리턴]**
    - (accuracy_ret, index_label_prediction) = obj.accuracy(…)
    - 즉, 리턴되는 index_label_prediction 은
    - 오답에 대한 MNIST 데이터, 정답, 계산된 예측 값을 가지고 있어야 한다

오차 역전파 실습 문제

- [문제 6] 오답에 대한 MNIST 데이터 이미지 출력
  - [디버깅 정보 이용하여 MNIST 이미지 출력]
    - [문제 5] 에서 구현 내용을 바탕으로, 예측이 실패한 MNIST 데이터를 아래와 같이 index,
    - label, prediction 정보와 함께 이미지로 출력하시오 (예측이 실패한 데이터는 random
    - 하게 선택되어 출력되게 하시오)