8주차(3/3)

역전파 1

파이썬으로 배우는 기계학습

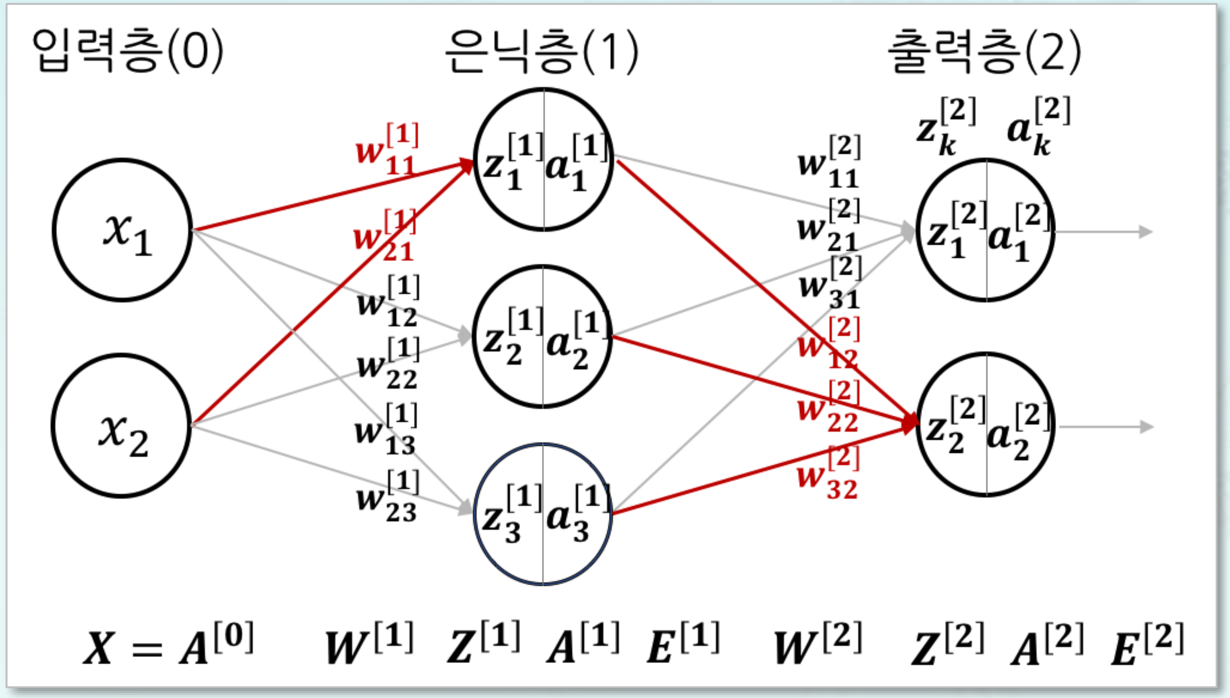
한동대학교
김영섭 교수

역전파 1

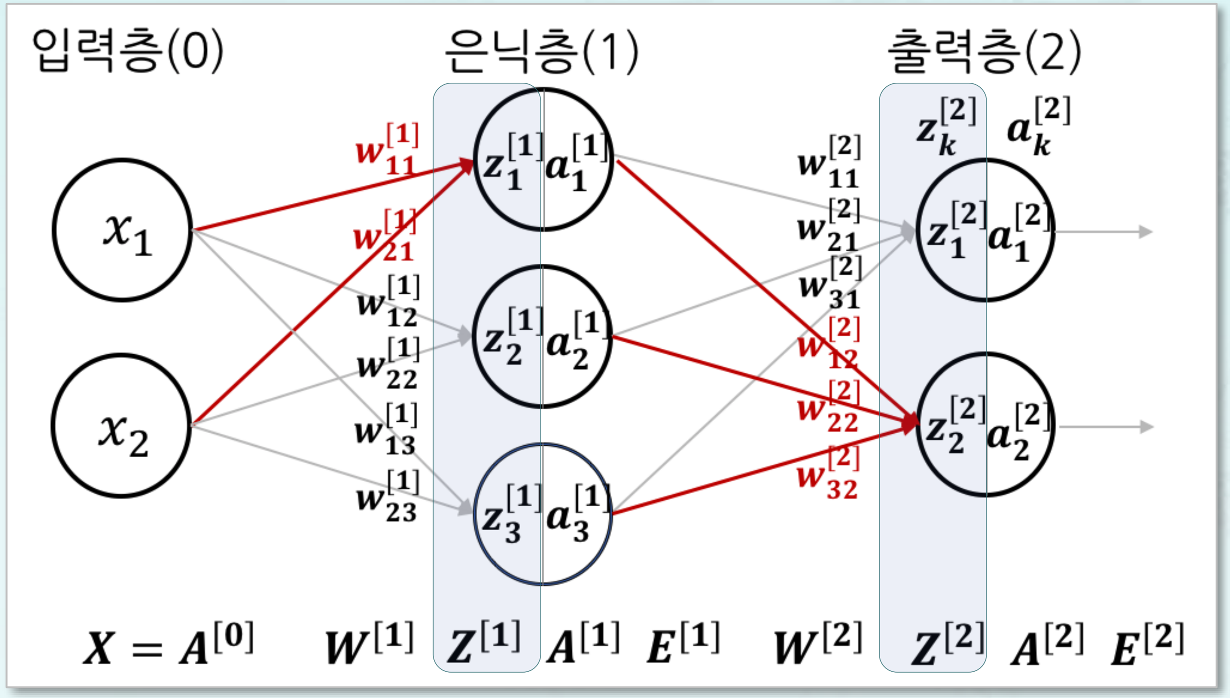
- 학습 목표
  - 역전파 개념을 이해한다.
  - 다층 신경망에서 은닉층 오차를 계산하는 방법을 배운다.
  - 은닉층 오차 계산을 단순화하고 일반화 하는 방법을 배운다.

- 학습 내용
  - 출력층 오차의 역전파
  - 역전파로 은닉층 오차 계산
  - 은닉층 오차 계산의 단순화 및 일반화

1. 다층 구조 인공 신경망: 신호처리 (복습)



1. 다층 구조 인공 신경망: 신호처리 (복습)

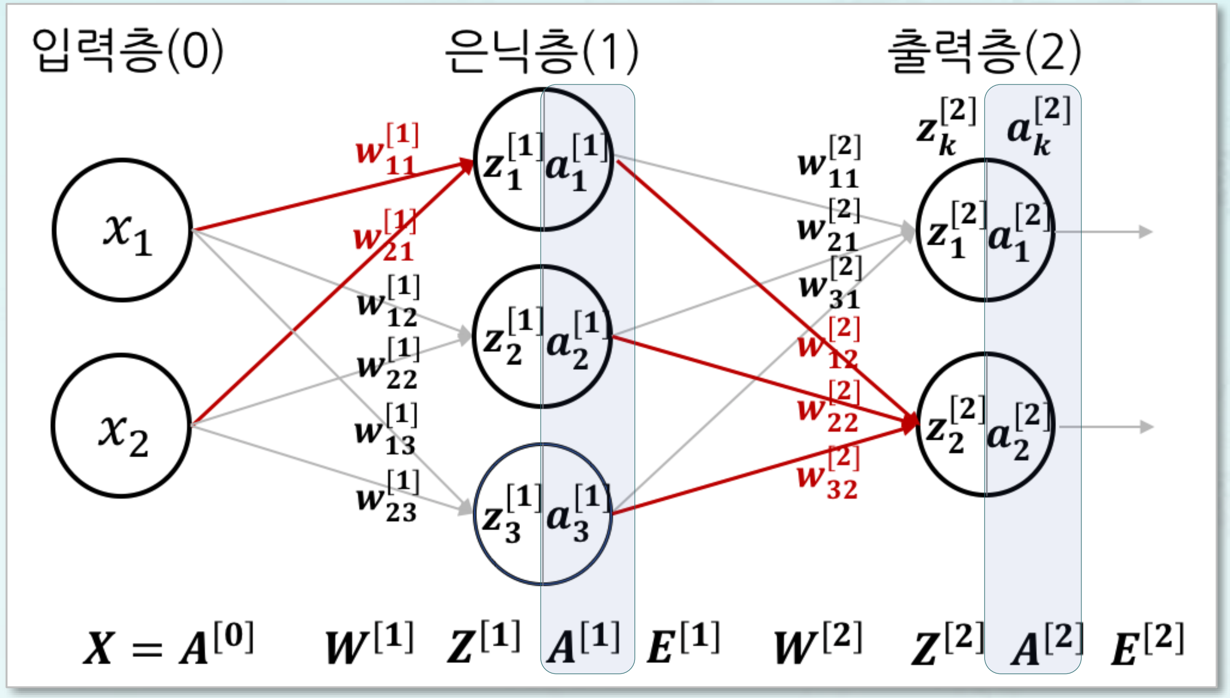


- **X** : 입력
- **W** : 가중치
- **Z : 순입력**

$$Z^{[l]} = W^{[l]T} \ A^{[l-1]}$$

$$Z^{[2]} = W^{[2]T} \ A^{[1]}$$

1. 다층 구조 인공 신경망: 신호처리 (복습)

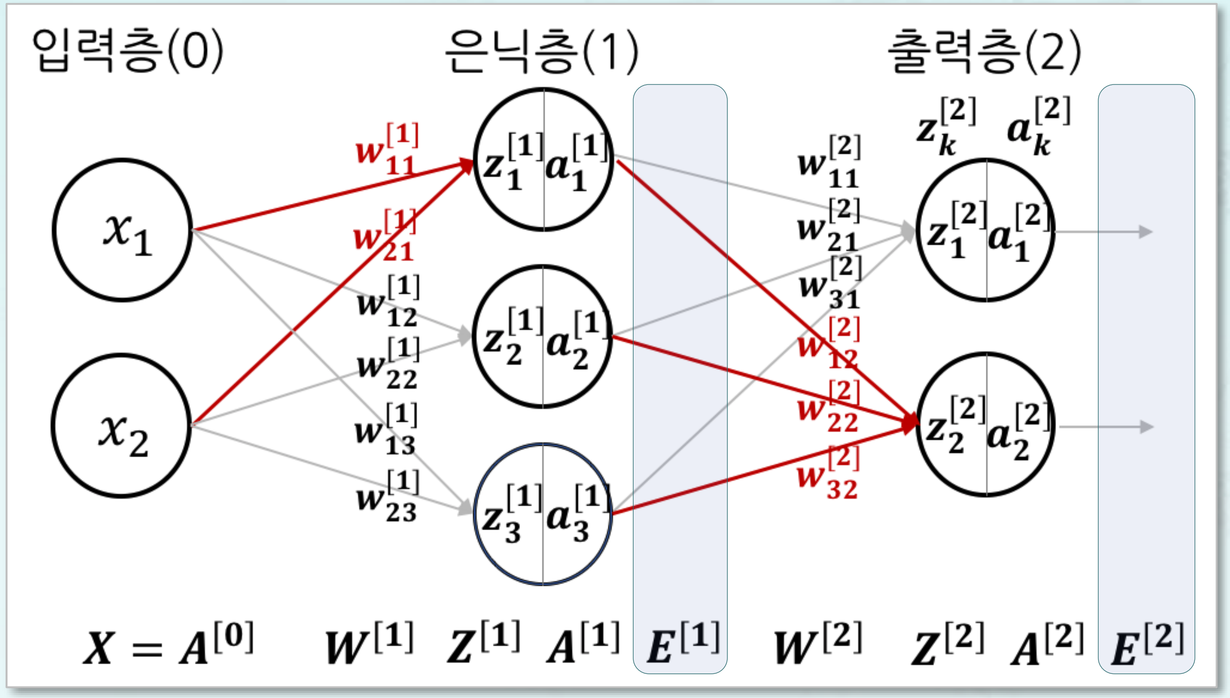


- **X** : 입력
- **W** : 가중치
- **Z** : 순입력
- **A : 출력**

$$\boxed{A^{[l]} = g(Z^{[l]})}$$

$$\boxed{A^{[2]} = sigmoid(Z^{[2]})}$$

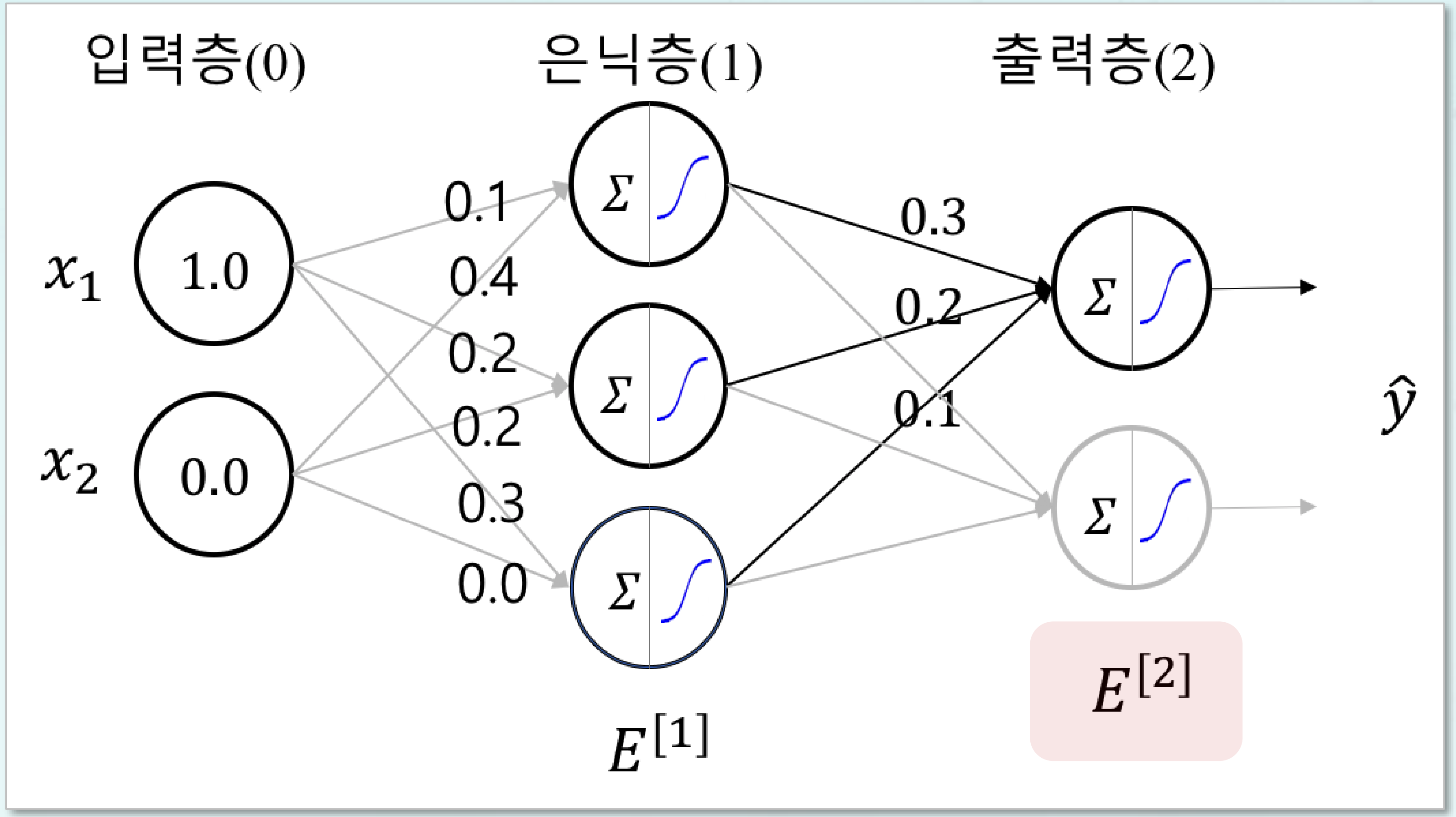
1. 다층 구조 인공 신경망: 신호처리 (복습)



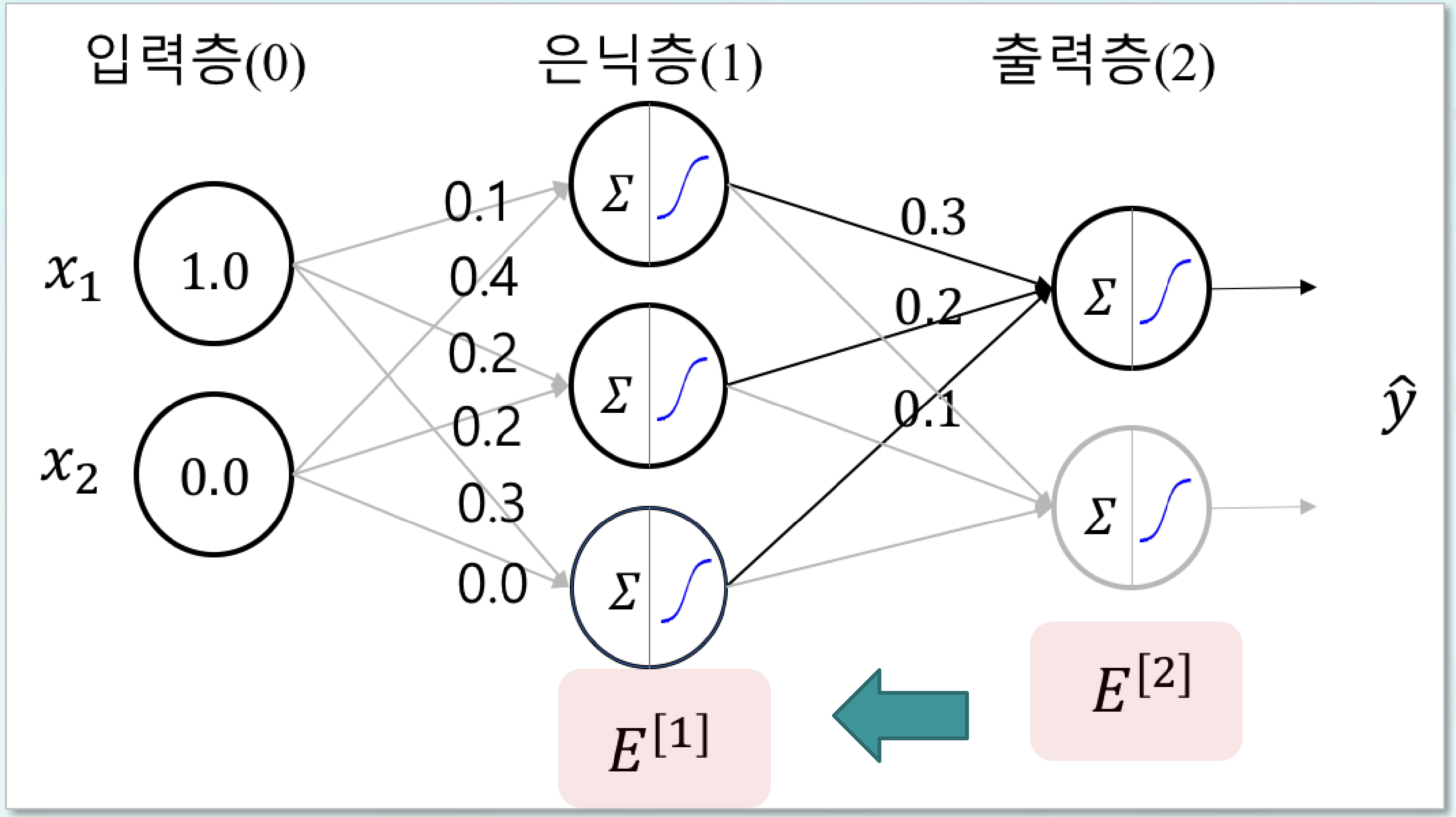
- **X** : 입력
- **W** : 가중치
- **Z** : 순입력
- **A** : 출력
- **E : 오차**

2. 역전파 1

2. 역전파 1: 은닉층의 오차 계산

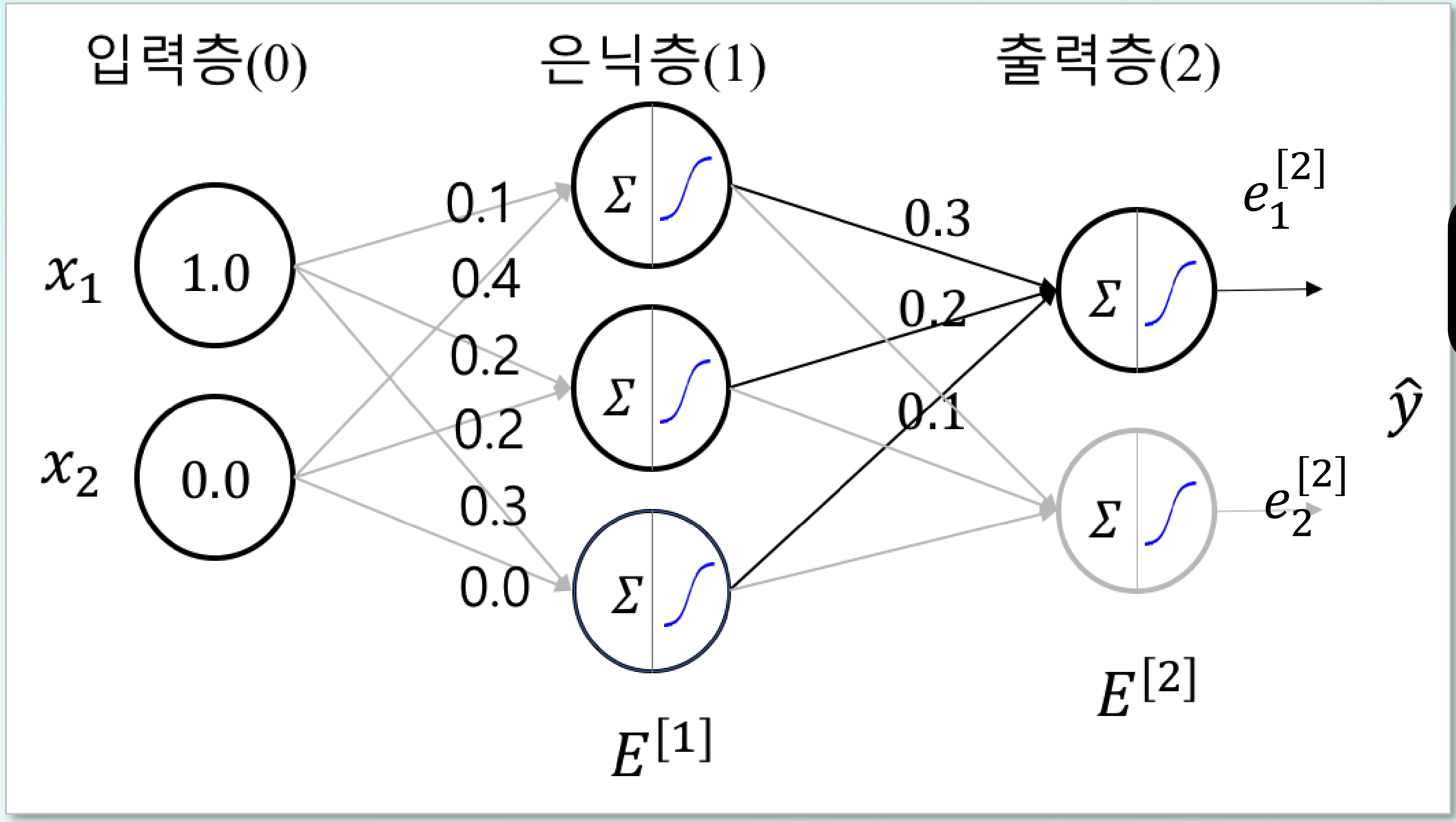


2. 역전파 1: 은닉층의 오차 계산



2. 역전파 1: 은닉층의 오차 계산

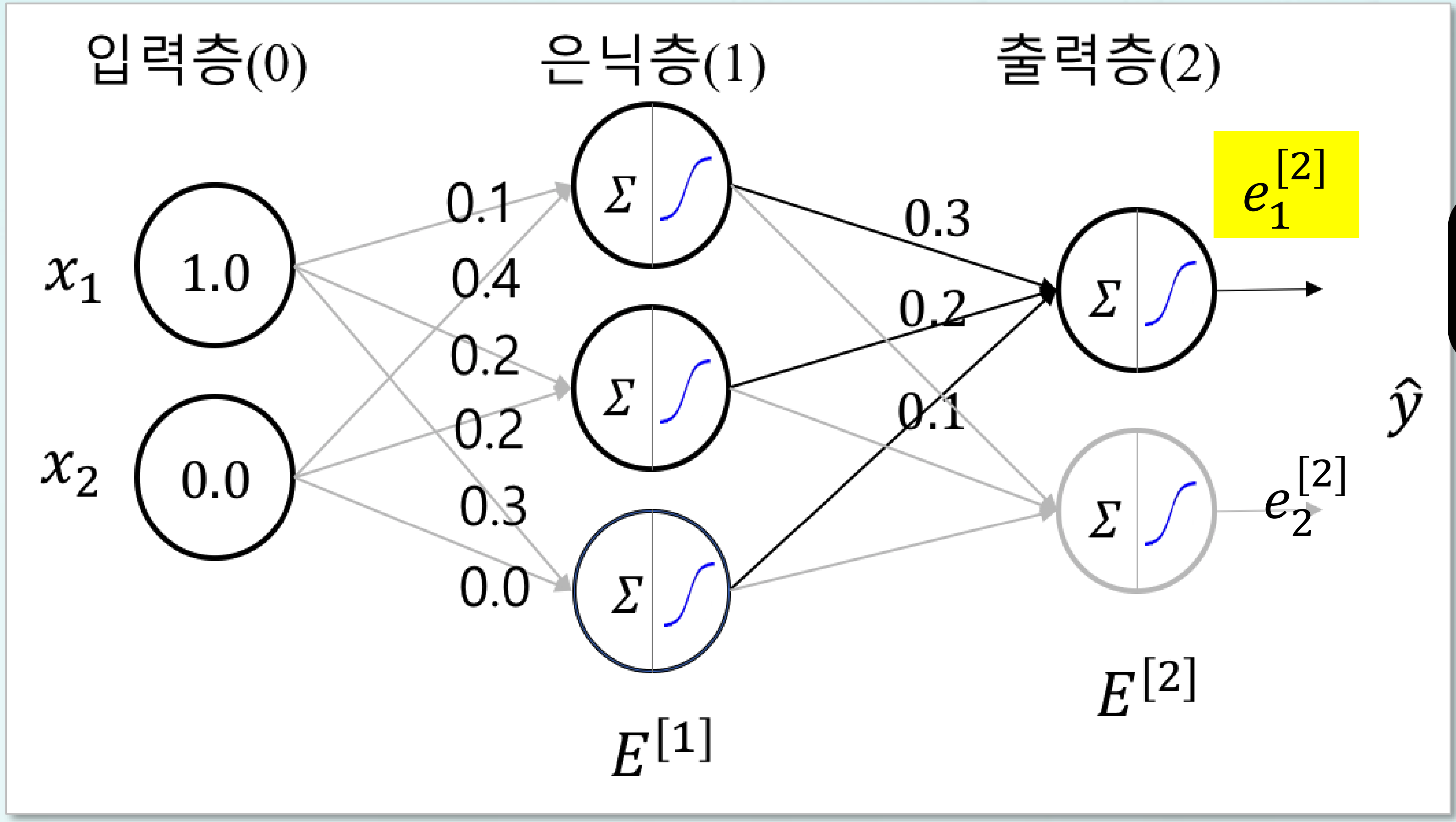
- 가정: $y_1 = 1, \ \hat{y}_1 = 0.58$
- 오차: $e_1 = y_1 - \hat{y}_1 = 0.42$



$$\begin{aligned} e_1 &= y_1 - \hat{y}_1 \\ &= 1.0 - 0.58 \\ &= 0.42 \end{aligned}$$

2. 역전파 1: 은닉층의 오차 계산

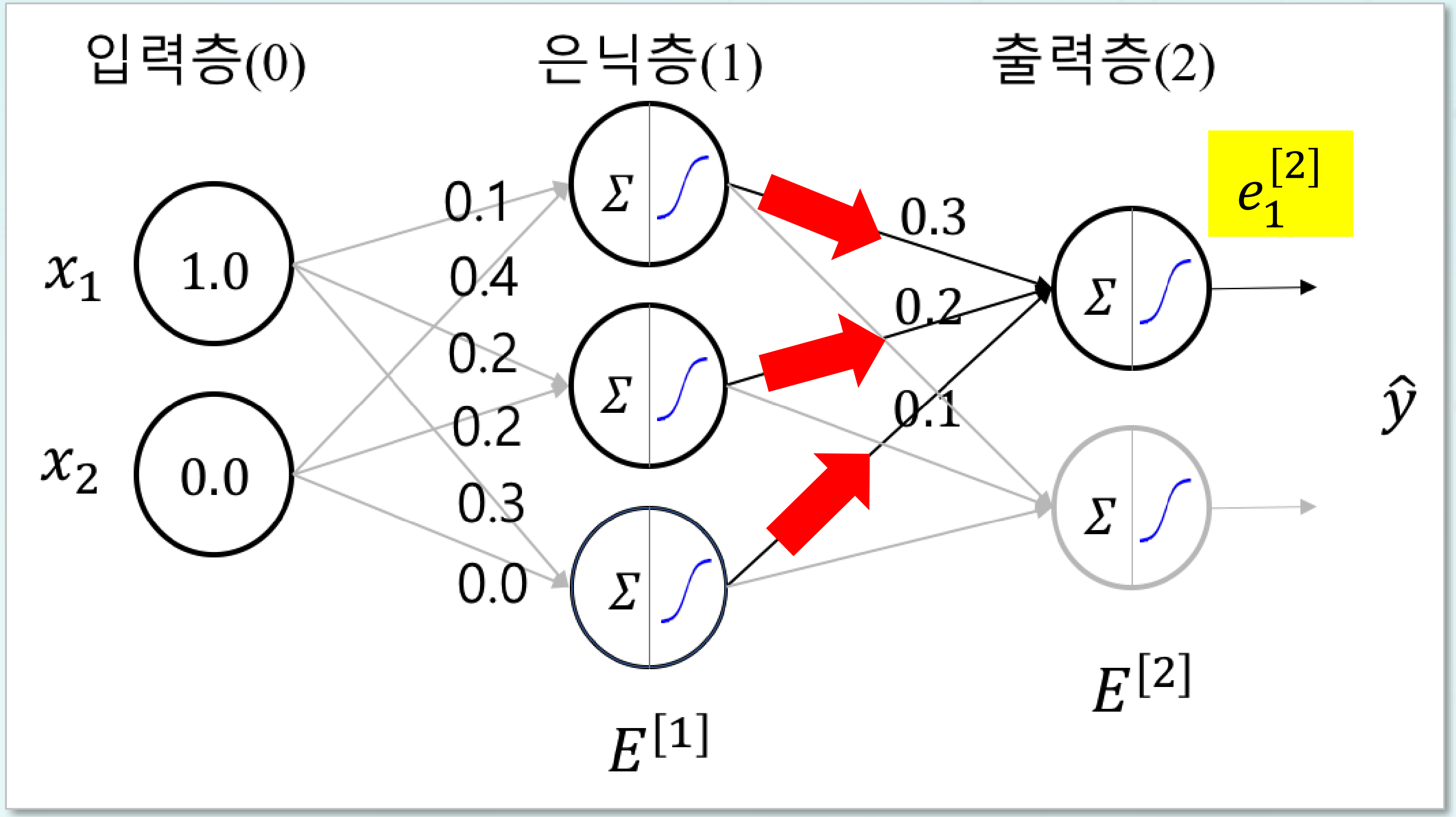
- 가정: $y_1 = 1, \ \hat{y}_1 = 0.58$
- 오차: $e_1 = y_1 - \hat{y}_1 = 0.42$



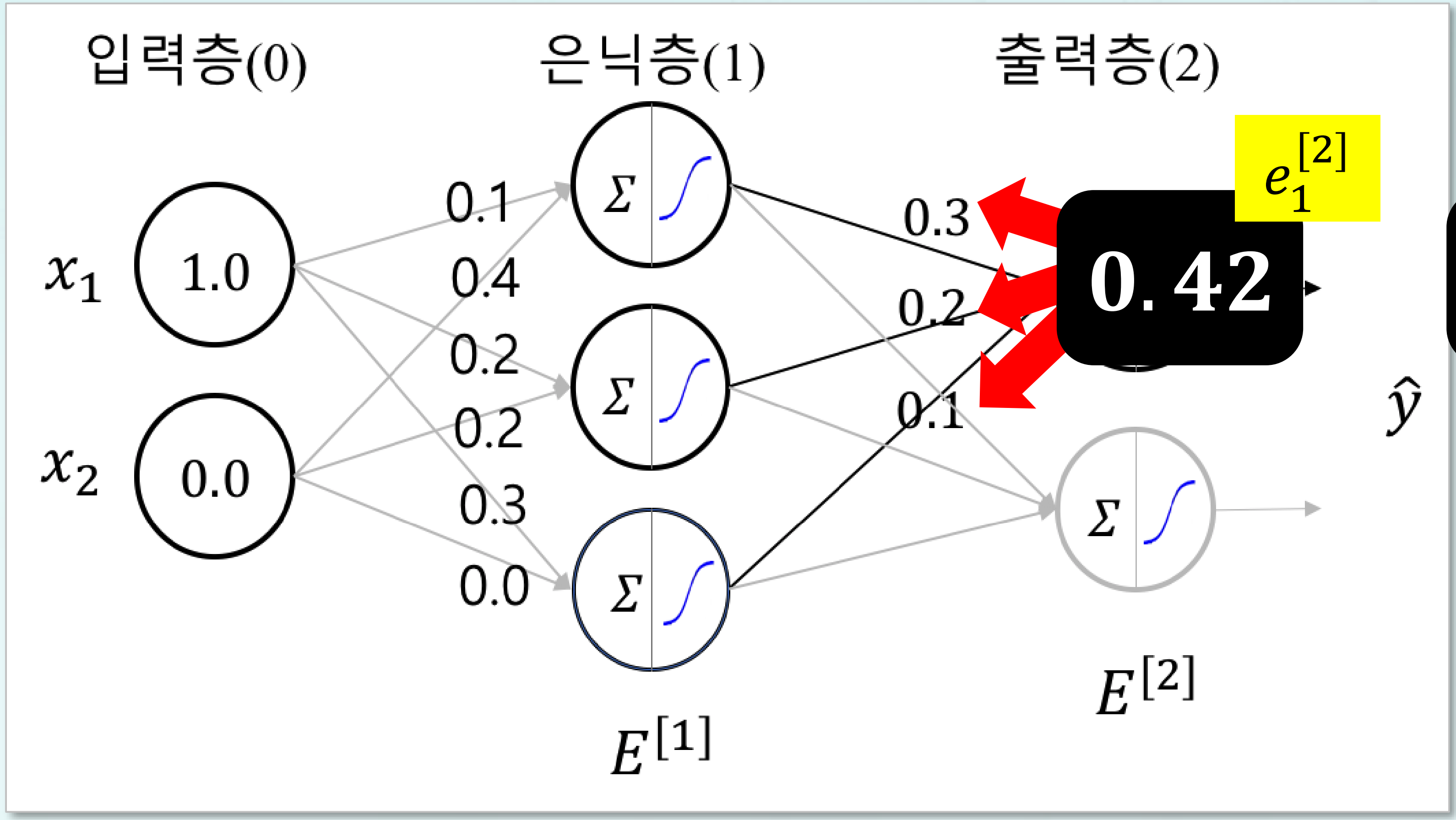
$$\begin{aligned} \boldsymbol{e_1} &= \boldsymbol{y_1 - \hat{y}_1} \\ &= \mathbf{1.0 - 0.58} \\ &= \mathbf{0.42} \end{aligned}$$

$$E^{[2]} = \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix} = \begin{pmatrix} 0.42 \\ \text{미정} \end{pmatrix}$$

2. 역전파 1: 은닉층의 오차 계산



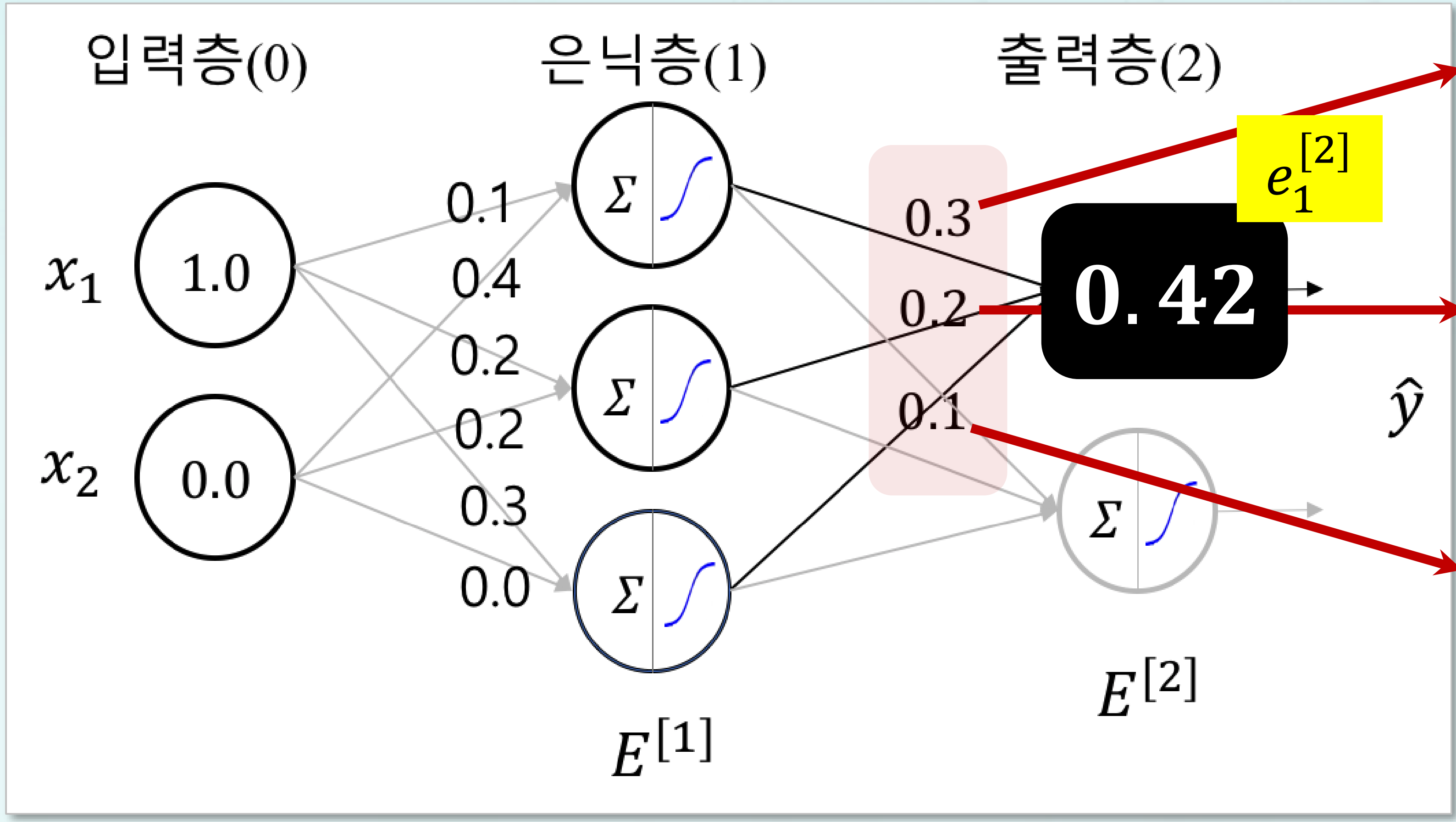
2. 역전파 1: 은닉층의 오차 계산



$$e_1 = y_1 - \hat{y}_1$$
$$= 1.0 - 0.58$$
$$= 0.42$$

2. 역전파 1: 은닉층의 오차 계산

오차 **0.42**를 가중치에 비례하여 배분



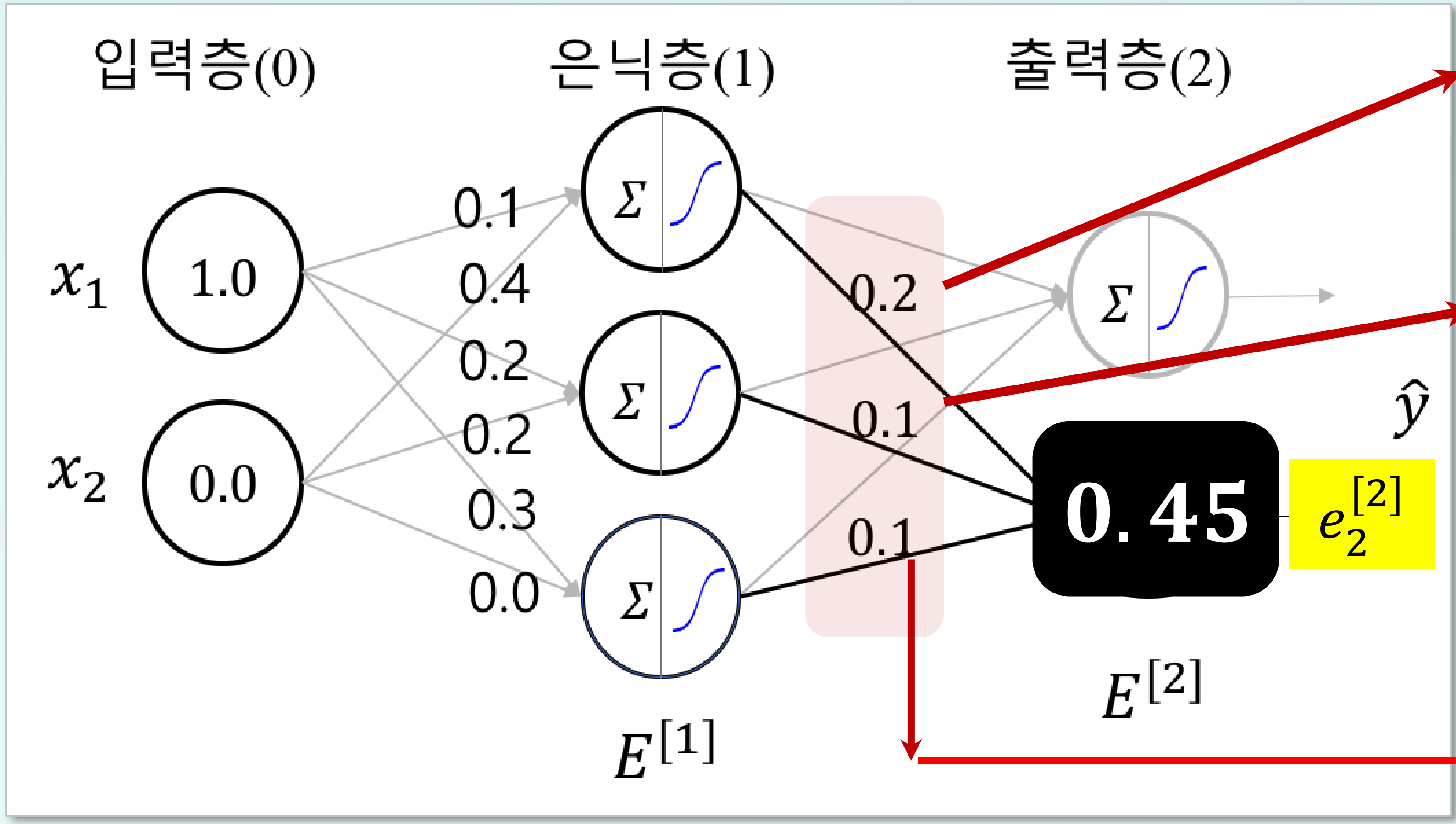
$$0.42 \times \frac{\mathbf{0.3}}{0.3 + 0.2 + 0.1} = 0.21$$

$$0.42 \times \frac{\mathbf{0.2}}{0.3 + 0.2 + 0.1} = 0.14$$

$$0.42 \times \frac{\mathbf{0.1}}{0.3 + 0.2 + 0.1} = 0.07$$

2. 역전파 1: 은닉층의 오차 계산

오차 **0.45**를 가중치에 비례하여 배분



$$0.45 \times \frac{\mathbf{0.2}}{0.2 + 0.1 + 0.1} = 0.23$$

$$0.45 \times \frac{\mathbf{0.1}}{0.2 + 0.1 + 0.1} = 0.11$$

$$0.45 \times \frac{\mathbf{0.1}}{0.2 + 0.1 + 0.1} = 0.11$$

2. 역전파 1: 은닉층의 오차 계산

$$E^{[2]} = \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix} = \begin{pmatrix} 0.42 \\ 0.45 \end{pmatrix}$$

$$\begin{aligned} e_1^{[2]} &= y_1 - \hat{y}_1 \\ &= 1 - 0.58 = 0.42 \end{aligned}$$

$$\begin{aligned} e_2^{[2]} &= y_2 - \hat{y}_2 \\ &= 1 - 0.55 = 0.45 \end{aligned}$$

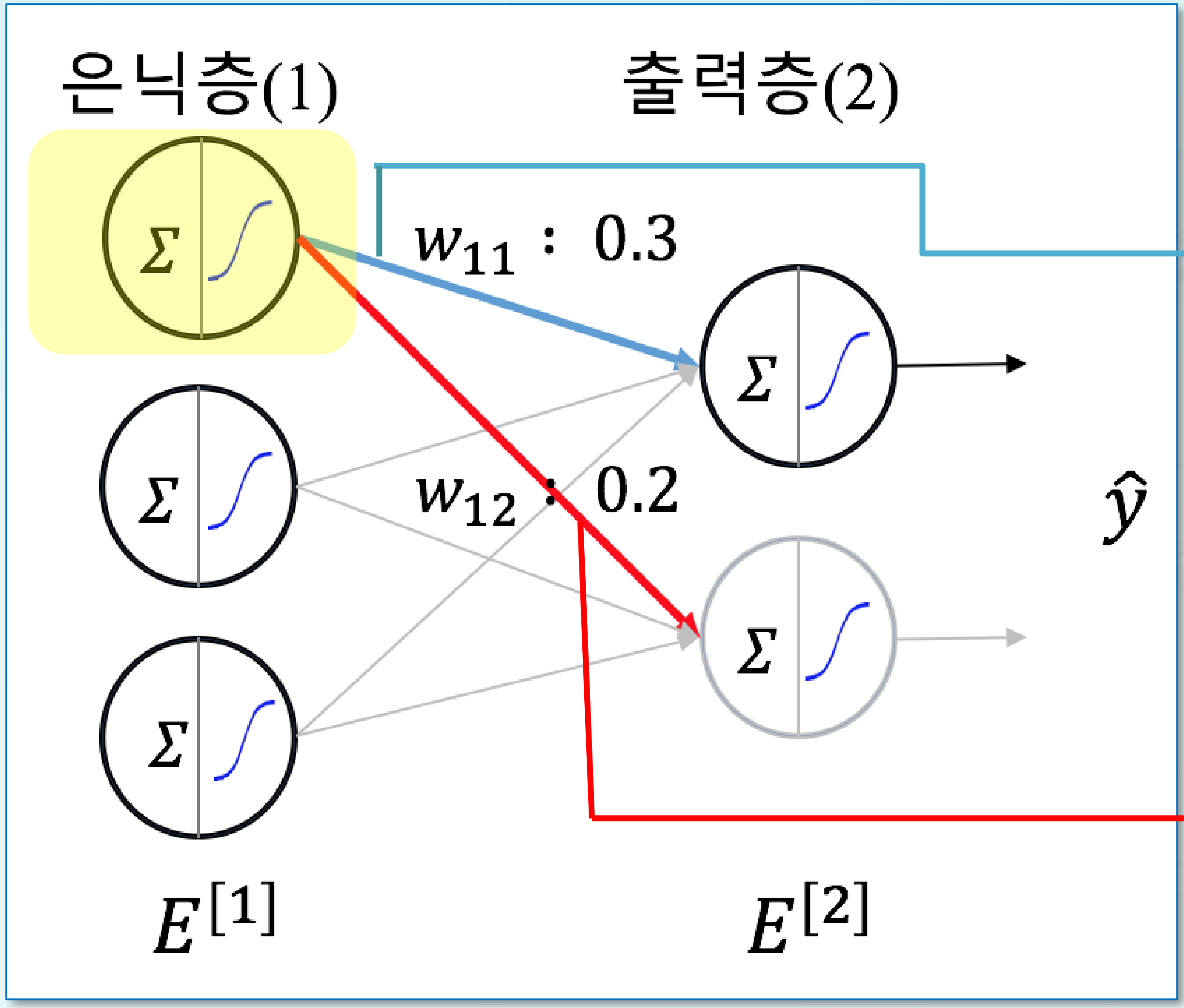
2. 역전파 1: 은닉층의 오차 계산

$$
E^{[1]} = \begin{pmatrix} e_1^{[1]} \\ e_2^{[1]} \\ e_3^{[1]} \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} 0.42 & & 0.45 & \end{array} \\ \begin{pmatrix} e_1^{[2]} \cdot \dfrac{w_{11}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \dfrac{w_{12}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \dfrac{w_{21}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \dfrac{w_{22}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \dfrac{w_{31}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \dfrac{w_{32}}{\sum_{i=1}^{3} w_{i2}} \end{pmatrix} \end{array}
$$

$E^{[1]}$

$$
E^{[2]} = \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix} = \begin{pmatrix} 0.42 \\ 0.45 \end{pmatrix}
$$

2. 역전파 1: 은닉층의 오차 계산



$$E^{[1]} = \begin{pmatrix} \boxed{e_1^{[2]} \cdot \frac{w_{11}}{\sum_{i=1}^{3} w_{i1}}} + \boxed{e_2^{[2]} \cdot \frac{w_{12}}{\sum_{i=1}^{3} w_{i2}}} \\ e_1^{[2]} \cdot \frac{w_{21}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{22}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \frac{w_{31}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{32}}{\sum_{i=1}^{3} w_{i2}} \end{pmatrix}$$

2. 역전파 1: 은닉층의 오차 계산

$$E^{[1]} = \begin{pmatrix} e_1^{[1]} \\ e_2^{[1]} \\ e_3^{[1]} \end{pmatrix} = \begin{pmatrix} e_1^{[2]} \cdot \frac{w_{11}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{12}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \frac{w_{21}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{22}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \frac{w_{31}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{32}}{\sum_{i=1}^{3} w_{i2}} \end{pmatrix}$$

$$= \begin{pmatrix} 0.42 \cdot \frac{0.3}{0.3 + 0.2 + 0.1} + 0.45 \cdot \frac{0.2}{0.2 + 0.1 + 0.1} \\ 0.42 \cdot \frac{0.3}{0.3 + 0.2 + 0.1} + 0.45 \cdot \frac{0.2}{0.2 + 0.1 + 0.1} \\ 0.42 \cdot \frac{0.3}{0.3 + 0.2 + 0.1} + 0.45 \cdot \frac{0.2}{0.2 + 0.1 + 0.1} \end{pmatrix} = \boxed{\begin{pmatrix} 0.44 \\ 0.25 \\ 0.18 \end{pmatrix}}$$

3. 오차 계산의 단순화와 일반화: 단순화

$$E^{[1]} = \begin{pmatrix} e_1^{[1]} \\ e_2^{[1]} \\ e_3^{[1]} \end{pmatrix} = \begin{pmatrix} e_1^{[2]} \cdot \frac{w_{11}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{12}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \frac{w_{21}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{22}}{\sum_{i=1}^{3} w_{i2}} \\ e_1^{[2]} \cdot \frac{w_{31}}{\sum_{i=1}^{3} w_{i1}} + e_2^{[2]} \cdot \frac{w_{32}}{\sum_{i=1}^{3} w_{i2}} \end{pmatrix}$$

$$= \boxed{\begin{pmatrix} \frac{w_{11}}{\sum_{i=1}^{3} w_{i1}} & \frac{w_{12}}{\sum_{i=1}^{3} w_{i2}} \\ \frac{w_{21}}{\sum_{i=1}^{3} w_{i1}} & \frac{w_{22}}{\sum_{i=1}^{3} w_{i2}} \\ \frac{w_{31}}{\sum_{i=1}^{3} w_{i1}} & \frac{w_{32}}{\sum_{i=1}^{3} w_{i2}} \end{pmatrix} \cdot \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix}}$$

3. 오차 계산의 단순화와 일반화: 단순화

$$E^{[1]} = \begin{pmatrix} \frac{w_{11}}{\boxed{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{12}}{\boxed{\sum_{i=1}^{3} w_{i2}}} \\ \frac{w_{21}}{\boxed{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{22}}{\boxed{\sum_{i=1}^{3} w_{i2}}} \\ \frac{w_{31}}{\boxed{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{32}}{\boxed{\sum_{i=1}^{3} w_{i2}}} \end{pmatrix} \cdot \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix}$$

3. 오차 계산의 단순화와 일반화: 단순화

$$E^{[1]} = \begin{pmatrix} \frac{w_{11}}{\cancel{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{12}}{\cancel{\sum_{i=1}^{3} w_{i2}}} \\ \frac{w_{21}}{\cancel{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{22}}{\cancel{\sum_{i=1}^{3} w_{i2}}} \\ \frac{w_{31}}{\cancel{\sum_{i=1}^{3} w_{i1}}} & \frac{w_{32}}{\cancel{\sum_{i=1}^{3} w_{i2}}} \end{pmatrix} \cdot \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix} \Rightarrow \boxed{E^{[1]} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix} \cdot \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix}}$$

3. 오차 계산의 단순화와 일반화: 일반화

$$E^{[1]} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix} \cdot \begin{pmatrix} e_1^{[2]} \\ e_2^{[2]} \end{pmatrix}$$

$$E^{[l]} = \begin{pmatrix} w_{11} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nm} \end{pmatrix}^T \cdot \begin{pmatrix} e_1^{[l+1]} \\ \vdots \\ e_n^{[l+1]} \end{pmatrix}$$

$$\boldsymbol{E}^{[1]} = W^{[2]\boldsymbol{T}} \cdot \boldsymbol{E}^{[2]}$$

$$E^{[l]} = W^{[l+1]T} \cdot E^{[l+1]}$$

8-3 역전파 1

- 학습 정리
  - 출력층 오차의 역전파
  - 역전파로 은닉층 오차 계산
  - 은닉층 오차 계산의 단순화 및 일반화