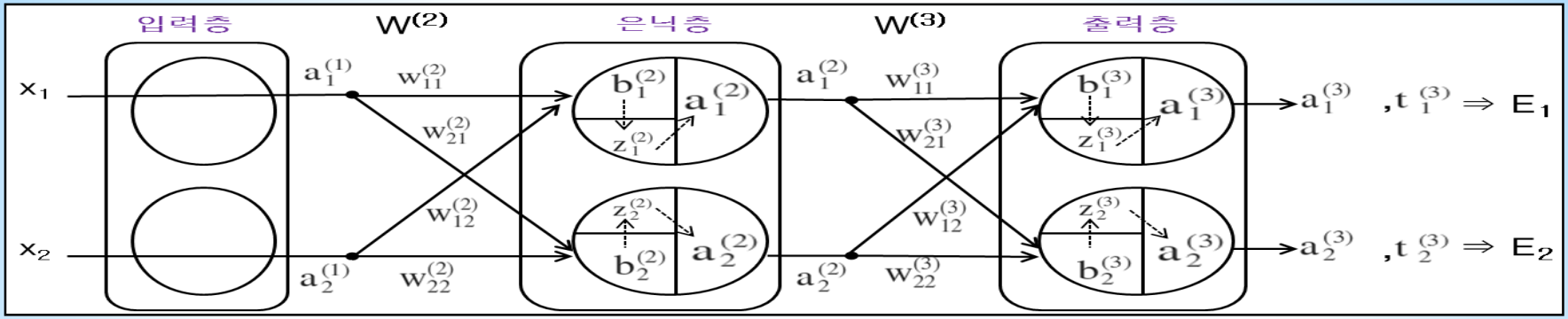
파이썬(Python)으로 구현하는

오차역전파
(Back Propagation)

– 은닉층에서의 오차역전파 공식–

Review – 선형회귀 값 (Z) • 출력 값 (A), 손실 값 (E), 가중치 (W), 바이어스 (b)



입력층 선형회귀 값 (Z1)	입력 층에는 가중치가 없기 때문에 선형회귀 값은 적용하지 않음	입력층 출력 값 (A1)	$a_2^{(1)} = x_2$ $a_1^{(1)} = x_1$
은닉층 선형회귀 값 (Z2)	$z_1^{(2)} = a_1^{(1)}w_{11}^{(2)} + a_2^{(1)}w_{12}^{(2)} + b_1^{(2)}$ $z_2^{(2)} = a_1^{(1)}w_{21}^{(2)} + a_2^{(1)}w_{22}^{(2)} + b_2^{(2)}$	은닉층 출력 값 (A2)	$a_1^{(2)} = \text{sigmoid}(z_1^{(2)})$ $a_2^{(2)} = \text{sigmoid}(z_2^{(2)})$
출력층 선형회귀 값 (Z3)	$z_1^{(3)} = a_1^{(2)}w_{11}^{(3)} + a_2^{(2)}w_{12}^{(3)} + b_1^{(3)}$ $z_2^{(3)} = a_1^{(2)}w_{21}^{(3)} + a_2^{(2)}w_{22}^{(3)} + b_2^{(3)}$	출력층 출력 값 (A3)	$a_1^{(3)} = \text{sigmoid}(z_1^{(3)})$ $a_2^{(3)} = \text{sigmoid}(z_2^{(3)})$
$W^{(2)}$, $W^{(3)}$	$W^{(2)} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{pmatrix}$ $W^{(3)} = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$	$b^{(2)}$, $b^{(3)}$	$b^{(2)} = \begin{pmatrix} b_1^{(2)} & b_2^{(2)} \end{pmatrix}$ $b^{(3)} = \begin{pmatrix} b_1^{(3)} & b_2^{(3)} \end{pmatrix}$
최종 손실 값 (E)	$E = \frac{1}{n}\sum_{i=1}^{n}(t_i^{(3)} - a_i^{(3)})^2 = \frac{1}{2}\{(t_1^{(3)} - a_1^{(3)})^2 + (t_2^{(3)} - a_2^{(3)})^2\} = E_1 + E_2$ ⇒		$E_1 = \frac{1}{2}(t_1^{(3)} - a_1^{(3)})^2$ $E_2 = \frac{1}{2}(t_2^{(3)} - a_2^{(3)})^2$

Review – 출력층 오차역전파 (Back Propagation) 공식

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

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

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

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

행렬곱 (dot product)

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

출력층 가중치 오차역전파 공식

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

$$\frac{\partial E}{\partial b^{(3)}} = \begin{pmatrix} \frac{\partial E}{\partial b_1^{(3)}} & \frac{\partial E}{\partial b_2^{(3)}} \end{pmatrix}$$

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

$$= loss\_3$$

출력층 바이어스 오차역전파 공식

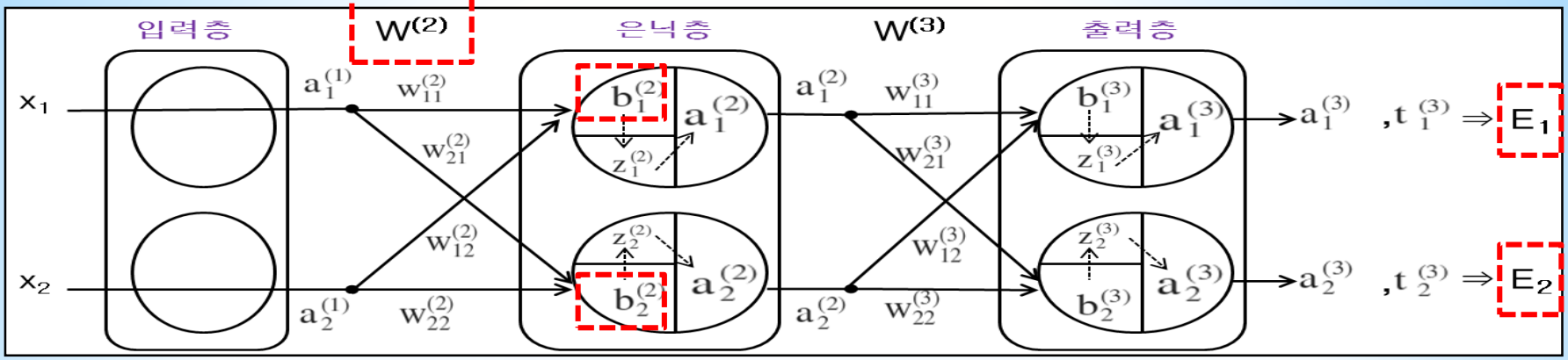
출력층 가중치 $W^{(3)}$ 업데이트

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

출력층 바이어스 $b^{(3)}$ 업데이트

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

은닉층 오차역전파 공식 유도 – 은닉층 가중치 $W^{(2)}$ / 은닉층 바이어스 $b^{(2)}$



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

지난시간

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

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

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

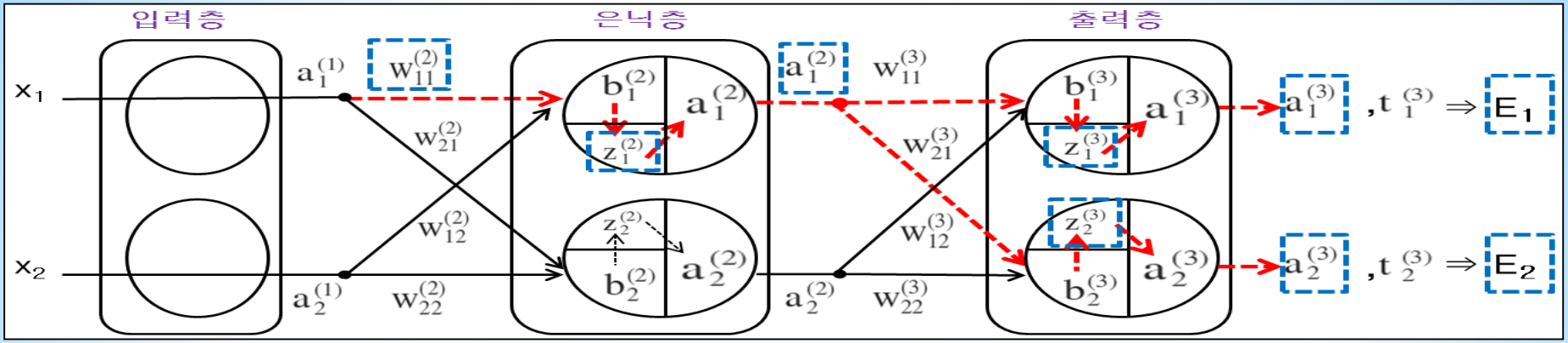
$$\frac{\partial E}{\partial w_{11}^{(2)}}, \frac{\partial E}{\partial w_{21}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \frac{\partial E}{\partial w_{22}^{(2)}} \qquad \frac{\partial E}{\partial b_{1}^{(2)}}, \frac{\partial E}{\partial b_{2}^{(2)}}$$

은닉층 가중치 $W^{(2)}$ 은닉층 바이어스 $b^{(2)}$

이번시간

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

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



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

체인 룰을 이용하여 국소미분으로 분리

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

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

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

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

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

[필요 수식]

$$E = E_1 + E_2$$

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

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

$$a_1^{(3)} = \mathrm{sigmoid}(z_1^{(3)})$$

$$a_2^{(3)} = \mathrm{sigmoid}(z_2^{(3)})$$

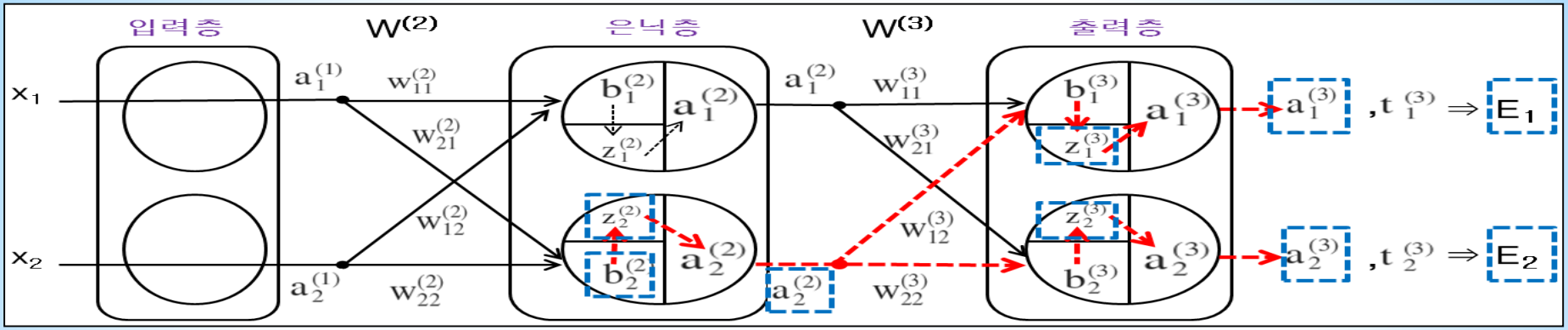
$$z_1^{(3)} = a_1^{(2)} w_{11}^{(3)} + a_2^{(2)} w_{12}^{(3)} + b_1^{(3)}$$

$$z_2^{(3)} = a_1^{(2)} w_{21}^{(3)} + a_2^{(2)} w_{22}^{(3)} + b_2^{(3)}$$

$$a_1^{(2)} = \mathrm{sigmoid}(z_1^{(2)})$$

$$z_1^{(2)} = a_1^{(1)} w_{11}^{(2)} + a_2^{(1)} w_{12}^{(2)} + b_1^{(2)}$$

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



$$\frac{\partial E}{\partial b_2^{(2)}} = \frac{\partial E_1}{\partial b_2^{(2)}} + \frac{\partial E_2}{\partial b_2^{(2)}}$$

체인 룰을 이용하여
국소미분으로 분리

$$= \frac{\partial E_1}{\partial a_1^{(3)}} \times \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial a_2^{(2)}} \times \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \times \frac{\partial z_2^{(2)}}{\partial b_2^{(2)}} + \frac{\partial E_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial a_2^{(2)}} \times \frac{\partial a_2^{(2)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(2)}}{\partial b_2^{(2)}}$$

$$= (a_1^{(3)} - t_1^{(3)}) \times \text{sigmoid}(z_1^{(3)})(1-\text{sigmoid}(z_1^{(3)})) \times w_{12}^{(3)} \times \text{sigmoid}(z_2^{(2)})(1-\text{sigmoid}(z_2^{(2)})) \times 1$$

$$+ (a_2^{(3)} - t_2^{(3)}) \times \text{sigmoid}(z_2^{(3)})(1-\text{sigmoid}(z_2^{(3)})) \times w_{22}^{(3)} \times \text{sigmoid}(z_2^{(2)})(1-\text{sigmoid}(z_2^{(2)})) \times 1$$

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

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

[필요 수식]

$$E = E_1 + E_2$$

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

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

$$a_1^{(3)} = \text{sigmoid}(z_1^{(3)})$$

$$a_2^{(3)} = \text{sigmoid}(z_2^{(3)})$$

$$z_1^{(3)} = a_1^{(2)}w_{11}^{(3)} + a_2^{(2)}w_{12}^{(3)} + b_1^{(3)}$$

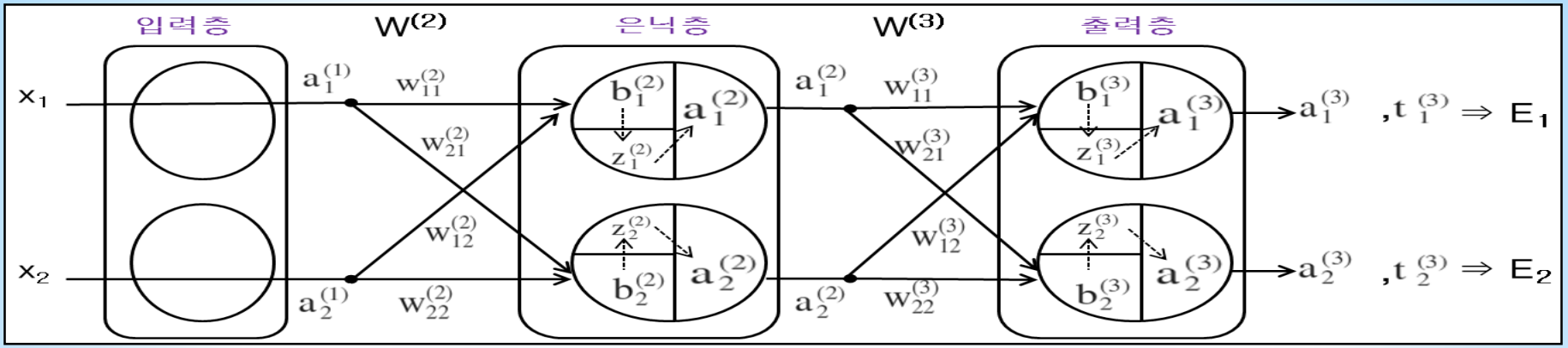
$$z_2^{(3)} = a_1^{(2)}w_{21}^{(3)} + a_2^{(2)}w_{22}^{(3)} + b_2^{(3)}$$

$$a_2^{(2)} = \text{sigmoid}(z_2^{(2)})$$

$$z_2^{(2)} = a_1^{(1)}w_{21}^{(2)} + a_2^{(1)}w_{22}^{(2)} + b_2^{(2)}$$

은닉층

Self Study



$$\frac{\partial E}{\partial w_{12}^{(2)}} = (a_1^{(3)}-t_1^{(3)}) \times a_1^{(3)}(1-a_1^{(3)}) \times w_{11}^{(3)} \times a_1^{(2)}(1-a_1^{(2)}) \times a_2^{(1)} + (a_2^{(3)}-t_2^{(3)}) \times a_2^{(3)}(1-a_2^{(3)}) \times w_{21}^{(3)} \times a_1^{(2)}(1-a_1^{(2)}) \times a_2^{(1)}$$

$$\frac{\partial E}{\partial w_{21}^{(2)}} = (a_1^{(3)}-t_1^{(3)}) \times a_1^{(3)}(1-a_1^{(3)}) \times w_{12}^{(3)} \times a_2^{(2)}(1-a_2^{(2)}) \times a_1^{(1)} + (a_2^{(3)}-t_2^{(3)}) \times a_2^{(3)}(1-a_2^{(3)}) \times w_{22}^{(3)} \times a_2^{(2)}(1-a_2^{(2)}) \times a_1^{(1)}$$

$$\frac{\partial E}{\partial w_{22}^{(2)}} = (a_1^{(3)}-t_1^{(3)}) \times a_1^{(3)}(1-a_1^{(3)}) \times w_{12}^{(3)} \times a_2^{(2)}(1-a_2^{(2)}) \times a_2^{(1)} + (a_2^{(3)}-t_2^{(3)}) \times a_2^{(3)}(1-a_2^{(3)}) \times w_{22}^{(3)} \times a_2^{(2)}(1-a_2^{(2)}) \times a_2^{(1)}$$

$$\frac{\partial E}{\partial b_1^{(2)}} = (a_1^{(3)}-t_1^{(3)}) \times a_1^{(3)}(1-a_1^{(3)}) \times w_{11}^{(3)} \times a_1^{(2)}(1-a_1^{(2)}) \times 1 + (a_2^{(3)}-t_2^{(3)}) \times a_2^{(3)}(1-a_2^{(3)}) \times w_{21}^{(3)} \times a_1^{(2)}(1-a_1^{(2)}) \times 1$$