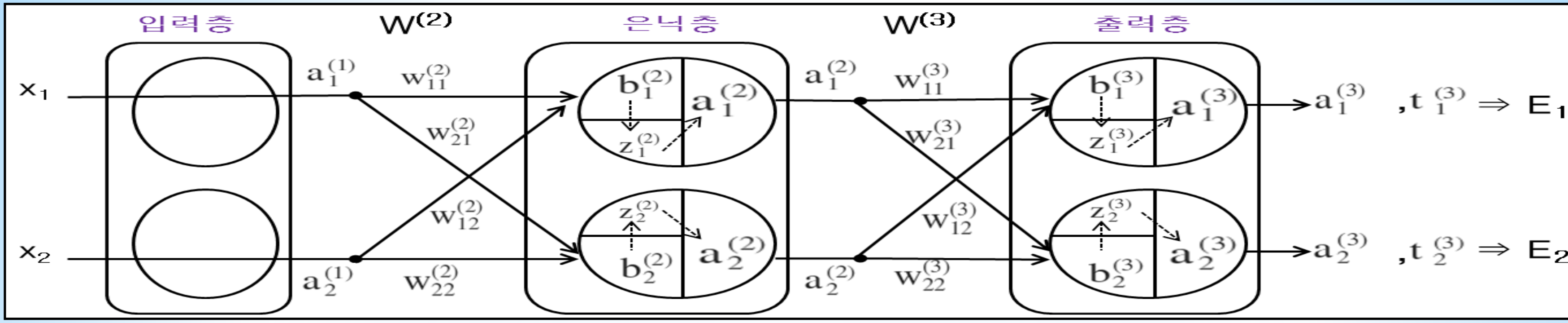
파이썬(Python)으로 구현하는

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)	$a_1^{(2)} = \mathrm{sigmoid}(z_1^{(2)})$ $a_2^{(2)} = \mathrm{sigmoid}(z_2^{(2)})$
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)	$a_1^{(3)} = \mathrm{sigmoid}(z_1^{(3)})$ $a_2^{(3)} = \mathrm{sigmoid}(z_2^{(3)})$
$W^{(2)}$, $W^{(3)}$	$W^{(2)} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} \end{pmatrix}$ $W^{(3)} = \begin{pmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \end{pmatrix}$	$b^{(2)}$, $b^{(3)}$	$b^{(2)} = \begin{pmatrix} b_1^{(2)} & b_2^{(2)} \end{pmatrix}$ $b^{(3)} = \begin{pmatrix} b_1^{(3)} & b_2^{(3)} \end{pmatrix}$
최종 손실 값 (E)	$E = \frac{1}{n}\sum_{i=1}^{n}(t_i^{(3)} - a_i^{(3)})^2 = \frac{1}{2}\{(t_1^{(3)} - a_1^{(3)})^2 + (t_2^{(3)} - a_2^{(3)})^2\} = E_1 + E_2$	➡	$E_1 = \frac{1}{2}(t_1^{(3)} - a_1^{(3)})^2$ $E_2 = \frac{1}{2}(t_2^{(3)} - a_2^{(3)})^2$

은닉층 가중치 오차역전파 (Back Propagation) 공식

입력층 출력	$A1 = (\, a_1^{(1)} \quad a_2^{(1)} \,)$	은닉층 출력	$A2 = (\, a_1^{(2)} \quad a_2^{(2)} \,)$
출력층 손실	$loss\_3 = \left((a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)}) \quad (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)}) \right)$		
은닉층 손실	$loss\_2 = (loss\_3 \bullet W3^{T}) \times (A2\times(1-A2))$		

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

$$= \begin{pmatrix} \begin{matrix} a_1^{(1)}(a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{11}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \\ +\ a_1^{(1)}(a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{21}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \end{matrix} & \begin{matrix} a_1^{(1)}(a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{12}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \\ +\ a_1^{(1)}(a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{22}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \end{matrix} \\ \begin{matrix} a_2^{(1)}(a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{11}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \\ +\ a_2^{(1)}(a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{21}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \end{matrix} & \begin{matrix} a_2^{(1)}(a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{12}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \\ +\ a_2^{(1)}(a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{22}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \end{matrix} \end{pmatrix}$$

$$= \begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \end{pmatrix} \bullet \left(\left((a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)}) \quad (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)}) \right) \bullet \begin{pmatrix} w_{11}^{(3)} & w_{12}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} \end{pmatrix} \times \left(a_1^{(2)}(1-a_1^{(2)}) \quad a_2^{(2)}(1-a_2^{(2)}) \right) \right)$$

행렬곱

$$= A1^{T} \bullet \left((loss\_3 \bullet W3^{T}) \times (A2\times(1-A2)) \right)$$

$$= \boxed{A1^{T} \bullet loss\_2}$$ 은닉층 가중치 오차역전파 공식

은닉층 가중치 $W^{(2)}$ 업데이트

$$\boxed{W^{(2)} = W^{(2)} - \alpha \frac{\partial E}{\partial W^{(2)}} = W^{(2)} - \alpha \times (A1^{T} \bullet loss\_2)}$$

은닉층 바이어스 오차역전파 (Back Propagation) 공식

입력층 출력	$A1 = \begin{pmatrix} a_1^{(1)} & a_2^{(1)} \end{pmatrix}$	은닉층 출력	$A2 = \begin{pmatrix} a_1^{(2)} & a_2^{(2)} \end{pmatrix}$
출력층 손실	$loss\_3 = \begin{pmatrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)}) & (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)}) \end{pmatrix}$		
은닉층 손실	$loss\_2 = (loss\_3 \bullet W3^T) \times (A2 \times (1-A2))$		

$$\boxed{\frac{\partial E}{\partial b^{(2)}}} = \begin{pmatrix} \dfrac{\partial E}{\partial b_1^{(2)}} & \dfrac{\partial E}{\partial b_2^{(2)}} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{matrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{11}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \\ +\ (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{21}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \end{matrix} & \begin{matrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{12}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \\ +\ (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{22}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \end{matrix} \\ \begin{matrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{11}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \\ +\ (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{21}^{(3)}a_1^{(2)}(1-a_1^{(2)}) \end{matrix} & \begin{matrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)})w_{12}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \\ +\ (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)})w_{22}^{(3)}a_2^{(2)}(1-a_2^{(2)}) \end{matrix} \end{pmatrix}$$

$$= \begin{pmatrix} (a_1^{(3)}-t_1^{(3)})a_1^{(3)}(1-a_1^{(3)}) & (a_2^{(3)}-t_2^{(3)})a_2^{(3)}(1-a_2^{(3)}) \end{pmatrix} \bullet \begin{pmatrix} w_{11}^{(3)} & w_{12}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} \end{pmatrix} \times \begin{pmatrix} a_1^{(2)}(1-a_1^{(2)}) & a_2^{(2)}(1-a_2^{(2)}) \end{pmatrix}$$

$$= \big((loss\_3 \bullet W3^T) \times (A2 \times (1-A2)) \big)$$

$$= \boxed{loss\_2}$$ 은닉층 바이어스 오차역전파 공식

은닉층 바이어스 $b^{(2)}$ 업데이트

$$\boxed{b^{(2)} = b^{(2)} - \alpha \frac{\partial E}{\partial b^{(2)}} = b^{(2)} - \alpha \times loss\_2}$$