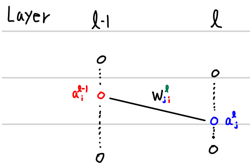


$a^l = \sigma(w^l a^{l-1} + b^l)$, a^l, b^l : $n_l \times 1$ 행렬 , w^l : $n_{l-1} \times n_l$ 행렬

a^l의 j번째 값 : $a_j^l = \sigma\left(\sum_{i=1}^{n_{l-1}} w_{ji}^l a_i^{l-1} + b_j^l\right)$

$\sigma(\cdot)$: 활성화 함수 (activation function)

C : 기준함수 (criterion function) +) 손실함수 Loss function 으로 생각해도 무방

[Input 과 Feedforward]

- Input : a^1
- $z^l = w^l \cdot a^{l-1} + b^l$ and $a^l = \sigma(z^l)$ where $l = 2, 3, \cdots, L$

[Output error 와 Backpropagate the error]

- Output error : $\delta^L = \frac{\partial C}{\partial z^L}$ 이라 하자.

l이 $L-1, L-2, \cdots, 2$ 에서 $\delta^l = \left((w^{l+1})^T \delta^{l+1}\right) \odot \sigma'(z^l)$ 을 만족한다.

pf) $$\delta^{l-1} = \frac{\partial C}{\partial z^{l-1}} = \left(\frac{\partial C}{\partial z_1^{l-1}}, \frac{\partial C}{\partial z_2^{l-1}}, \cdots, \frac{\partial C}{\partial z_{n_{l-1}}^{l-1}}\right)^T$$
$$= \left(\frac{\partial C}{\partial a_1^{l-1}} \frac{\partial a_1^{l-1}}{\partial z_1^{l-1}}, \frac{\partial C}{\partial a_2^{l-1}} \frac{\partial a_2^{l-1}}{\partial z_2^{l-1}}, \cdots \right)^T$$
$$= \left(\frac{\partial C}{\partial a_1^{l-1}}, \frac{\partial C}{\partial a_2^{l-1}}, \cdots \right)^T \odot \sigma'(z^{l-1}) \quad \cdots \text{ by } ①$$
$$= \left((w^l)^T \delta^l\right) \odot \sigma'(z^{l-1})$$

[Weight update]

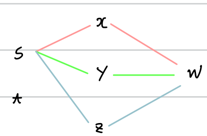
- $\frac{\partial C}{\partial w_{ji}^l} = \frac{\partial C}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial w_{ji}^l} = \delta_j^l a_i^{l-1} \Rightarrow \frac{\partial C}{\partial w^l} = \delta^l (a^{l-1})^T$
- $\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial b_j^l} = \delta_j^l \Rightarrow \frac{\partial C}{\partial b^l} = \delta^l$

① i가 $1, 2, \cdots, n_{l-1}$ 일때

$$\frac{\partial C}{\partial a_i^{l-1}} = \sum_{j=1}^{n_l} \frac{\partial C}{\partial a_j^l} \cdot \frac{\partial a_j^l}{\partial a_i^{l-1}} \quad \cdots \text{ by } ②$$
$$= \sum_{j=1}^{n_l} \frac{\partial C}{\partial a_j^l} \cdot \frac{\partial a_j^l}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial a_i^{l-1}} = \sum_{j=1}^{n_l} \delta_j^l \cdot w_{ji}^l = \left[(w^l)^T \delta^l\right]_i$$

② Chain rule : 연쇄법칙

case) $w = f(x, y, z)$, $x = P(s,t)$, $y = Q(s,t)$, $z = R(s,t)$



$$\frac{dw}{ds} = \frac{dw}{dx} \cdot \frac{dx}{ds} + \frac{dw}{dy} \cdot \frac{dy}{ds} + \frac{dw}{dz} \cdot \frac{dz}{ds}$$

[문제 1]

input : a^1과 target : t 가 $a^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $t = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ 이고

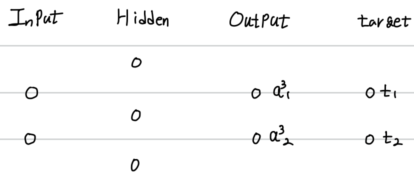
$W^2 = \begin{pmatrix} 1 & -2 \\ 2 & 4 \\ -3 & 1 \end{pmatrix}$, $b^2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

$W^3 = \begin{pmatrix} 1 & 2 & -3 \\ 2 & -1 & 3 \end{pmatrix}$, $b^3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, activation : $\sigma(x) = x$

$\begin{pmatrix} a^3_1 \\ a^3_2 \end{pmatrix} = \sigma\left(w^3 \cdot \sigma(w^2 \cdot a^1 + b^2) + b^3\right)$ 일때

$Loss = \frac{1}{2}\left[(a^3_1 - t_1)^2 + (a^3_2 - t_2)^2\right]$에 대하여 $lr = 0.01$로

Weight update를 한 결과를 구해보자.



[forward]

Layer : l	forward $\xrightarrow{Wa+b}$	z^l	$\xrightarrow{\sigma(z)}$	a^l
1				$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
2	$W^2a^1 + b^2 \longrightarrow$	$\begin{pmatrix} 0 \\ 8 \\ 1 \end{pmatrix}$	$\longrightarrow$	$\begin{pmatrix} 0 \\ 8 \\ 1 \end{pmatrix}$
3	$W^3a^2 + b^3 \longrightarrow$	$\begin{pmatrix} 15 \\ -4 \end{pmatrix}$	$\longrightarrow$	$\begin{pmatrix} 15 \\ -4 \end{pmatrix}$

[Backward]

Layer : l	backward	$\longrightarrow$	δ^l
3	$\frac{\partial L}{\partial z^3}$	$\longrightarrow$	$\begin{pmatrix} 13 \\ -7 \end{pmatrix}$
2	$((W^3)^T \cdot \delta^3) \odot \sigma'(z^2)$	$\longrightarrow$	$\begin{pmatrix} -1 \\ 33 \\ -60 \end{pmatrix}$

[Weight update]

$\frac{\partial L}{\partial W^3} = \begin{pmatrix} 13 \\ -7 \end{pmatrix} \begin{pmatrix} 0 & 8 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 104 & 13 \\ 0 & -56 & -7 \end{pmatrix}$, $\frac{\partial L}{\partial b^3} = \begin{pmatrix} 13 \\ -7 \end{pmatrix}$

$W^3 - 0.01 \frac{dL}{dW^3} = \begin{pmatrix} 1 & 0.96 & -3.13 \\ 2 & -0.44 & 3\ 07 \end{pmatrix}$

$b^3 - 0.01 \frac{dL}{db^3} = \begin{pmatrix} 1.87 \\ 1.07 \end{pmatrix}$

$\frac{\partial L}{\partial W^2} = \begin{pmatrix} -1 \\ 33 \\ -60 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 33 & 33 \\ -60 & -60 \end{pmatrix}$, $\frac{\partial L}{\partial b^2} = \begin{pmatrix} -1 \\ 33 \\ 60 \end{pmatrix}$

$W^2 - 0.01 \frac{\partial L}{\partial W^2} = \begin{pmatrix} 1.01 & -1.99 \\ 1.67 & 3.67 \\ -2\ 4 & 1.6 \end{pmatrix}$

$b^2 - 0.01 \frac{\partial L}{\partial b^2} = \begin{pmatrix} 1.01 \\ 1.67 \\ 3.6 \end{pmatrix}$

[문제 2]

input : a^1과 target : t 가 $a^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $t = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ 이고

$W^2 = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}$, $b^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$W^3 = \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix}$, $b^3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, activation : $\sigma(x) = x^2$

$\begin{pmatrix} a_1^3 \\ a_2^3 \end{pmatrix} = \sigma\left(W^3 \cdot \sigma(W^2 \cdot a^1 + b^2) + b^3\right)$ 일때

$Loss = \frac{1}{2}\left[(a_1^3 - t_1)^2 + (a_2^3 - t_2)^2\right]$ 에 대하여 $lr = 0.01$로

weight update를 한 결과를 구해보자.

Input	Hidden	Output	target
o	o	o a_1^3	o t_1
o	o	o a_2^3	o t_2

[forward]

Layer : l	forward $\xrightarrow{Wa+b}$	z^l $\xrightarrow{\sigma(z)}$	a^l
1			$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
2	$W^2 a^1 + b^2 \longrightarrow$	$\begin{pmatrix} -1 \\ 2 \end{pmatrix} \longrightarrow$	$\begin{pmatrix} 1 \\ 4 \end{pmatrix}$
3	$W^3 a^2 + b^3 \longrightarrow$	$\begin{pmatrix} -3 \\ -2 \end{pmatrix} \longrightarrow$	$\begin{pmatrix} 9 \\ 4 \end{pmatrix}$

[Backward]

Layer : l	backward $\longrightarrow$	δ^l
3	$\frac{\partial L}{\partial z^3} \longrightarrow$	$\begin{pmatrix} -42 \\ -4 \end{pmatrix}$
2	$((W^3)^T \cdot \delta^3) \odot \sigma'(z^2) \longrightarrow$	$\begin{pmatrix} 100 \\ 184 \end{pmatrix}$

$$\frac{\partial L}{\partial z^3} = \frac{\partial L}{\partial a^3} \cdot \frac{\partial a^3}{\partial z^3} = \begin{pmatrix} a_1^3 - t_1 \\ a_2^3 - t_2 \end{pmatrix} \odot \begin{pmatrix} 2 \cdot z_1^3 \\ 2 \cdot z_2^3 \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \end{pmatrix} \odot \begin{pmatrix} -6 \\ -4 \end{pmatrix} = \begin{pmatrix} -42 \\ -4 \end{pmatrix}$$

$$((W^3)^T \cdot \delta^3) \odot \sigma'(z^2) = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} -42 \\ -4 \end{pmatrix} \odot \begin{pmatrix} -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 100 \\ 184 \end{pmatrix}$$

[Weight update]

$$\frac{\partial L}{\partial W^3} = \begin{pmatrix} -42 \\ -4 \end{pmatrix} \begin{pmatrix} 1 & 4 \end{pmatrix} = \begin{pmatrix} -42 & -168 \\ -4 & -16 \end{pmatrix} , \quad \frac{\partial L}{\partial b^3} = \begin{pmatrix} -42 \\ -4 \end{pmatrix}$$

$$W^3 - 0.01 \frac{dL}{dW^3} = \begin{pmatrix} 1.42 & 0.68 \\ 2.04 & -0.84 \end{pmatrix}$$

$$b^3 - 0.01 \frac{dL}{db^3} = \begin{pmatrix} 0.42 \\ 0.04 \end{pmatrix}$$

$$\frac{\partial L}{\partial W^2} = \begin{pmatrix} 100 \\ 184 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 100 & 100 \\ 184 & 184 \end{pmatrix} , \quad \frac{\partial L}{\partial b^2} = \begin{pmatrix} 100 \\ 184 \end{pmatrix}$$

$$W^2 - 0.01 \frac{\partial L}{\partial W^2} = \begin{pmatrix} 0 & -3 \\ -0.84 & -0.84 \end{pmatrix}$$

$$b^2 - 0.01 \frac{\partial L}{\partial b^2} = \begin{pmatrix} -1 \\ -1.84 \end{pmatrix}$$

[문제 3]

input : a^1와 target : t 가 $a^1 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix}$, $t = (-1, 0, 1)$

$W^2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 2 \end{pmatrix}$, $b^2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$W^3 = \begin{pmatrix} 2 & 3 & 1 \\ 2 & -1 & 2 \end{pmatrix}$, $b^3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$W^4 = (1, 1)$, $b^4 = (0)$, activation : $\sigma(x) = x$

$$a^4 = \sigma(W^4 \cdot \sigma(W^3 \cdot \sigma(W^2 \cdot a^1 + b^2) + b^3)) + b^4$$

$$Loss = \frac{1}{3}\left[(a^4_1 - t_1)^2 + (a^4_2 - t_2)^2 + (a^4_3 - t_3)^2\right]$$

weight update를 한 결과를 구해보자.

[forward]

Layer : l	forward $\xrightarrow{Wa+b}$ z^l	$\xrightarrow{\sigma(z)}$ a^l
1		$\begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix}$
2	$W^2a^1 + b^2 \longrightarrow \begin{pmatrix} 5 & -1 & 3 \\ 4 & 1 & 3 \\ 3 & -3 & 1 \end{pmatrix}$	$\rightarrow \begin{pmatrix} 5 & -1 & 3 \\ 4 & 1 & 3 \\ 3 & -3 & 1 \end{pmatrix}$
3	$W^3a^2 + b^3 \longrightarrow \begin{pmatrix} 25 & -2 & 16 \\ 12 & -9 & 5 \end{pmatrix}$	$\rightarrow \begin{pmatrix} 25 & -2 & 16 \\ 12 & -9 & 5 \end{pmatrix}$
4	$W^4a^3 + b^4 \longrightarrow (37, -11, 21)$	$\rightarrow (37, -11, 21)$

[Backward]

Layer : l	backward	$\longrightarrow \delta^l$
4	$\frac{\partial L}{\partial z^4}$	$\longrightarrow \frac{2}{3}(38, -11, 20)$
3	$((W^4)^T \cdot \delta^4) \odot \sigma'(z^3)$	$\longrightarrow \frac{2}{3}\begin{pmatrix} 38 & -11 & 20 \\ 38 & -11 & 20 \end{pmatrix}$
2	$((W^3)^T \cdot \delta^3) \odot \sigma'(z^2)$	$\longrightarrow \frac{2}{3}\begin{pmatrix} 152 & -44 & 80 \\ 76 & -22 & 40 \\ 114 & -33 & 60 \end{pmatrix}$

$$\frac{\partial L}{\partial z^3} = \frac{\partial L}{\partial a^3} \cdot \frac{\partial a^3}{\partial z^3} = \frac{2}{3}(a^4 - t) \odot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{2}{3}(38, -11, 20)$$

$$((W^4)^T \cdot \delta^4) \odot \sigma'(z^3) = \frac{2}{3}\begin{pmatrix} 38 & -11 & 20 \\ 38 & -11 & 20 \end{pmatrix} \odot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 38 & -11 & 20 \\ 38 & -11 & 20 \end{pmatrix}$$

$$((W^3)^T \cdot \delta^3) \odot \sigma'(z^2) = \frac{2}{3}\begin{pmatrix} 2 & 2 \\ 3 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 38 & -11 & 20 \\ 38 & -11 & 20 \end{pmatrix} \odot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$= \frac{2}{3}\begin{pmatrix} 152 & -44 & 80 \\ 76 & -22 & 40 \\ 114 & -33 & 60 \end{pmatrix}$$

[Weight update]

$$\frac{\partial L}{\partial W^4} = \delta^4 \cdot (a^3)^T = \frac{2}{3}(38, -11, 20) \cdot \begin{pmatrix} 25 & 12 \\ -2 & -9 \\ 16 & 5 \end{pmatrix} = \frac{2}{3}(1252, 655)$$

$$\frac{\partial L}{\partial W^3} = \delta^3 \cdot (a^2)^T = \frac{2}{3}\begin{pmatrix} 38 & -11 & 20 \\ 38 & -11 & 20 \end{pmatrix}\begin{pmatrix} 5 & 4 & 3 \\ -1 & 1 & -3 \\ 3 & 3 & 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 261 & 201 & 167 \\ 261 & 201 & 167 \end{pmatrix}$$

$$\frac{\partial L}{\partial W^2} = \delta^2 \cdot (a^1)^T = \frac{2}{3}\begin{pmatrix} 152 & -44 & 80 \\ 76 & -22 & 40 \\ 114 & -33 & 60 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 188 & 428 \\ 94 & 214 \\ 141 & 321 \end{pmatrix}$$