오차역전파 계산하기

By. JongSeobJ

1. 변수 설명

$$X = (x_1 \quad \cdots \quad x_n)$$

$$W_0 = \begin{pmatrix} {w^0}_{11} & \cdots & {w^0}_{1h_0} \\ \vdots & \ddots & \vdots \\ {w^0}_{n1} & \cdots & {w^0}_{nh_0} \end{pmatrix} (n \times h_0)$$

$$N_0 = (N_{01} \quad \cdots \quad N_{0h_0})(1 \times h_0)$$

$$O_0 = (O_{01} \quad \cdots \quad O_{0h_0})(1 \times h_0)$$

layer0

$$W_1 = \begin{pmatrix} {w^1}_{11} & \cdots & {w^1}_{1h_1} \\ \vdots & \ddots & \vdots \\ {w^1}_{h_0 1} & \cdots & {w^1}_{h_0 h_1} \end{pmatrix} (h_0 \times h_1)$$

$$N_1 = (N_{11} \quad \cdots \quad N_{1h_1})(1 \times h_1)$$

$$O_1 = (O_{11} \quad \cdots \quad O_{1h_1})(1 \times h_1)$$

layer1

$$W_2 = \begin{pmatrix} {w^2}_{11} & \cdots & {w^2}_{1h_2} \\ \vdots & \ddots & \vdots \\ {w^2}_{h_1 1} & \cdots & {w^2}_{h_1 h_2} \end{pmatrix} (h_1 \times h_2)$$

$$N_2 = (N_{21} \quad \cdots \quad N_{2h_2})(1 \times h_2)$$

$$O_2 = (O_{21} \quad \cdots \quad O_{2h_2})(1 \times h_2)$$

layer2

$$O_i = active(N_{01} \quad \cdots \quad N_{0h_i})$$

$$Loss = \frac{1}{2}\sum(target - O_2)^2$$

2. W_2

$$\frac{\partial Loss}{\partial W_2} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

$$\frac{\partial Loss}{\partial O_2} = \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \end{pmatrix} (1 \times h_2) \qquad \frac{\partial Scalar}{\partial Vector}$$

$$\frac{\partial O_2}{\partial N_2} = \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & \frac{\partial O_{21}}{\partial N_{2h_2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial O_{2h_2}}{\partial N_{21}} & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} = \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} (h_2 \times h_2) \qquad \frac{\partial Vector}{\partial Vector}$$

$$O_2 = (O_{21} \quad \cdots \quad O_{2h_2}) = active(N_{01} \quad \cdots \quad N_{0h_2})$$

$$\frac{\partial O_{2i}}{\partial N_{2j}} = 0, (i \neq j)$$

활성화 함수는 자기 자신의 미분값만을 가진다.
자기 자신 외에는 0

2. W_2

$$\frac{\partial Loss}{\partial W_2} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

$$\frac{\partial N_2}{\partial W_2} = \left(\begin{pmatrix} \frac{\partial N_{21}}{\partial w^2{}_{11}} & \cdots & \frac{\partial N_{21}}{\partial w^2{}_{1h_2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_{2h_2}}{\partial w^2{}_{1h_2}} & \cdots & \frac{\partial N_{2h_2}}{\partial w^2{}_{1h_2}} \end{pmatrix} \cdots \begin{pmatrix} \frac{\partial N_{21}}{\partial w^2{}_{h_1 1}} & \cdots & \frac{\partial N_{21}}{\partial w^2{}_{h_1 h_2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_{2h_2}}{\partial w^2{}_{h_1 h_2}} & \cdots & \frac{\partial N_{2h_2}}{\partial w^2{}_{h_1 h_2}} \end{pmatrix} \right)$$

$$= \left(\begin{pmatrix} O_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{11} \end{pmatrix} \cdots \begin{pmatrix} O_{1h_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{1h_1} \end{pmatrix} \right) (h_2 \times (h_2 \times h_2)) \quad \frac{\partial Vector}{\partial Matrix}$$

$$N_2 = (O_{11} \quad \cdots \quad O_{1h_1}) \circ \begin{pmatrix} w^2{}_{11} & \cdots & w^2{}_{1h_2} \\ \vdots & \ddots & \vdots \\ w^2{}_{h_1 1} & \cdots & w^2{}_{h_1 h_2} \end{pmatrix}$$

$$= \left(O_{11} w^2{}_{11} + \cdots + O_{1h_1} w^2{}_{h_1 1} \quad \cdots \quad O_{11} w^2{}_{1h_2} + \cdots + O_{1h_1} w^2{}_{h_1 h_2} \right)$$

$$= (N_{21} \quad \cdots \quad N_{2h_2})$$

2. W_2

$$\frac{\partial Loss}{\partial W_2} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

$$= \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \end{pmatrix} \circ \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} \circ \left(\begin{pmatrix} O_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{11} \end{pmatrix} \cdots \begin{pmatrix} O_{1h_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{1h_1} \end{pmatrix} \right)$$

$(1 \times h_2) \circ (h_2 \times h_2) \circ (h_2 \times (h_1 \times h_2))$

$$= \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \times \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} \circ \left(\begin{pmatrix} O_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{11} \end{pmatrix} \cdots \begin{pmatrix} O_{1h_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{1h_1} \end{pmatrix} \right)$$

$(1 \times h_2) \circ (h_2 \times (h_1 \times h_2))$

$$= \left(\begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} \end{pmatrix} \cdots \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} \end{pmatrix} \right)$$

$(1 \times (h_1 \times h_2))$

2. W_2

$$\frac{\partial Loss}{\partial W_2} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

$$= \left(\left(\frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} \quad \cdots \quad \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} \right) \quad \cdots \quad \left(\frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} \quad \cdots \quad \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} \right) \right)$$

$(1 \times (h_1 \times h_2))$ → reshape →

$$\begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} \\ \vdots & \ddots & \vdots \\ \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} \end{pmatrix} \quad (h_1 \times h_2)$$

2. W_2

$$\frac{\partial Loss}{\partial W_2} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

계산의 편리함을 위한 trick

$$= \begin{pmatrix} O_{11} \\ \vdots \\ O_{1h_1} \end{pmatrix} \circ \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \end{pmatrix} \circ \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix}$$

$(h_1 \times 1) \circ (1 \times h_2) \circ (h_2 \times h_2) = (h_1 \times h_2)$

$$= \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{11} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{2h_2}} \times O_{11} \\ \vdots & \ddots & \vdots \\ \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{21}} \times O_{1h_1} & \cdots & \frac{\partial Loss}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial N_{2h_2}} \times O_{1h_1} \end{pmatrix}$$

$$\rightarrow \quad \frac{\partial Loss}{\partial W_2} = {O_1}^T \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial W_2}$$

2. W_1

$$\frac{\partial Loss}{\partial W_1} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial O_1} \times \frac{\partial O_1}{\partial N_1} \times \frac{\partial N_1}{\partial W_1}$$

$\frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2}$: 위와 동일

$$\frac{\partial N_2}{\partial O_1} = \begin{pmatrix} \frac{\partial N_{21}}{\partial O_{11}} & \cdots & \frac{\partial N_{21}}{\partial O_{1h_1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_{2h_2}}{\partial O_{11}} & \cdots & \frac{\partial N_{2h_2}}{\partial O_{1h_1}} \end{pmatrix} = \begin{pmatrix} {w^2}_{11} & \cdots & {w^2}_{h_1 1} \\ \vdots & \ddots & \vdots \\ {w^2}_{1h_2} & \cdots & {w^2}_{h_1 h_2} \end{pmatrix} = {W_2}^T (h_2 \times h_2)$$

$$N_2 = (O_{11} \quad \cdots \quad O_{1h_1}) \circ \begin{pmatrix} {w^2}_{11} & \cdots & {w^2}_{1h_2} \\ \vdots & \ddots & \vdots \\ {w^2}_{h_1 1} & \cdots & {w^2}_{h_1 h_2} \end{pmatrix}$$

$$= \left(O_{11}{w^2}_{11} + \cdots + O_{1h_1}{w^2}_{h_1 1} \quad \cdots \quad O_{11}{w^2}_{1h_2} + \cdots + O_{1h_1}{w^2}_{h_1 h_2} \right)$$

$$= (N_{21} \quad \cdots \quad N_{2h_2})$$

2. W_1

$$\frac{\partial Loss}{\partial W_1} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial O_1} \times \frac{\partial O_1}{\partial N_1} \times \frac{\partial N_1}{\partial W_1}$$

$$\frac{\partial O_1}{\partial N_1} = \begin{pmatrix} \frac{\partial O_{11}}{\partial N_{11}} & \cdots & \frac{\partial O_{11}}{\partial N_{1h_1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial O_{1h_1}}{\partial N_{11}} & \cdots & \frac{\partial O_{1h_1}}{\partial N_{1h_1}} \end{pmatrix} = \begin{pmatrix} \frac{\partial O_{11}}{\partial N_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{1h_1}}{\partial N_{1h_1}} \end{pmatrix}$$

$$\frac{\partial N_1}{\partial W_1} = \begin{pmatrix} \begin{pmatrix} \frac{\partial N_{11}}{\partial w^1{}_{11}} & \cdots & \frac{\partial N_{11}}{\partial w^1{}_{1h_1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_{1h_1}}{\partial w^1{}_{1h_1}} & \cdots & \frac{\partial N_{1h_1}}{\partial w^1{}_{1h_1}} \end{pmatrix} & \cdots & \begin{pmatrix} \frac{\partial N_{11}}{\partial w^1{}_{h_0 1}} & \cdots & \frac{\partial N_{11}}{\partial w^1{}_{h_0 h_1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial N_{1h_1}}{\partial w^1{}_{h_0 h_1}} & \cdots & \frac{\partial N_{1h_1}}{\partial w^1{}_{h_0 h_1}} \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{pmatrix} O_{01} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{01} \end{pmatrix} & \cdots & \begin{pmatrix} O_{0h_0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{0h_0} \end{pmatrix} \end{pmatrix} \quad (h_1 \times (h_0 \times h_1))$$

2. W_1

$$\frac{\partial Loss}{\partial W_1} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial O_1} \times \frac{\partial O_1}{\partial N_1} \times \frac{\partial N_1}{\partial W_1}$$

$$= \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \end{pmatrix} \circ \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} \circ {W_2}^T \circ \begin{pmatrix} \frac{\partial O_{11}}{\partial N_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{1h_1}}{\partial N_{1h_1}} \end{pmatrix} \circ \left(\begin{pmatrix} O_{01} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{01} \end{pmatrix} \quad \cdots \quad \begin{pmatrix} O_{0h_0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{0h_0} \end{pmatrix} \right)$$

$$(1 \times h_2) \circ (h_2 \times h_2) \circ (h_2 \times h_1) \circ (h_1 \times h_1) \circ (h_1 \times (h_0 \times h_1))$$

2. W_1

$$\frac{\partial Loss}{\partial W_1} = \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial O_1} \times \frac{\partial O_1}{\partial N_1} \times \frac{\partial N_1}{\partial W_1}$$

$$= \begin{pmatrix} \frac{\partial Loss}{\partial O_{21}} & \cdots & \frac{\partial Loss}{\partial O_{2h_2}} \end{pmatrix} \circ \begin{pmatrix} \frac{\partial O_{21}}{\partial N_{21}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{2h_2}}{\partial N_{2h_2}} \end{pmatrix} \circ {W_2}^T \circ \begin{pmatrix} \frac{\partial O_{11}}{\partial N_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial O_{1h_1}}{\partial N_{1h_1}} \end{pmatrix} \circ \left(\begin{pmatrix} O_{01} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{01} \end{pmatrix} \quad \cdots \quad \begin{pmatrix} O_{0h_0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & O_{0h_0} \end{pmatrix} \right)$$

$$(1 \times h_2) \circ (h_2 \times h_2) \circ (h_2 \times h_1) \circ (h_1 \times h_1) \circ (h_1 \times (h_0 \times h_1))$$

계산의 편리함을 위한 trick

$$\frac{\partial Loss}{\partial W_1} = {O_0}^T \times \frac{\partial Loss}{\partial O_2} \times \frac{\partial O_2}{\partial N_2} \times \frac{\partial N_2}{\partial O_1} \times \frac{\partial O_1}{\partial N_1}$$

3. W_i

$i = 1, \dots n, (n = \#\ of\ layer)$

$$\frac{\partial Loss}{\partial W_n} = {O_{n-1}}^T \times \frac{\partial Loss}{\partial O_n} \times \frac{\partial O_n}{\partial N_n}$$

$$\frac{\partial Loss}{\partial W_i} = {O_{i-1}}^T \times \frac{\partial Loss}{\partial O_n} \times \left(\prod_{j=0}^{n-i-1} \frac{\partial O_{n-j}}{\partial N_{n-j}} \times \frac{\partial N_{n-j}}{\partial O_{n-j-1}} \right) \times \frac{\partial O_i}{\partial N_i}$$