한국항공대학교
KOREA AEROSPACE UNIVERSITY

Back-propagation II

한국항공대학교
소프트웨어학과

1

오늘 수업 목표

1. Chain Rule
2. Backpropagation
3. Backpropagation with numbers

2

Chain Rule

3

Example 0

$$y = f(u) \quad u = g(x)$$

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

4

Example 0-3

$y = f(\vec{u}) \qquad \vec{u} = g(\vec{x})$

$$\frac{dy}{dx_1} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x_1} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x_1}$$

$$\frac{dy}{dx_2} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x_2} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x_2}$$

∇

y

$u_1 + u_2$

$x_1 \quad x_2$

5

Example 0-3

$y = f(\vec{u}) \qquad \vec{u} = g(\vec{x})$

$$\left[\frac{dy}{dx_1}, \frac{dy}{dx_2}\right] = \nabla_{\vec{u}}\, y \cdot \begin{bmatrix} \frac{du_1}{dx_1}, & \frac{du_1}{dx_2} \\ \frac{du_2}{dx_1}, & \frac{du_2}{dx_2} \end{bmatrix}$$

$\nabla_{\vec{x}}\, y$

$$\nabla_{\vec{x}}\, y = \nabla_{\vec{u}}\, y \cdot \nabla_{\vec{x}}\, \vec{u} \qquad J$$

6

Gradient & Jacobian

$f: \mathbb{R}^n \to \mathbb{R}^1$

$$\nabla_{\vec{x}}\, l = \begin{bmatrix} \frac{\partial l}{\partial x_0} & \frac{\partial l}{\partial x_1} \end{bmatrix}$$

$f: \mathbb{R}^{n=2} \to \mathbb{R}^{m=3}$

$$\nabla_{\vec{x}}\, \vec{V} = \begin{bmatrix} \frac{\partial V_0}{\partial x_0} & \frac{\partial V_2}{\partial x_0} \\ \frac{\partial V_1}{\partial x_0} & \frac{\partial V_2}{\partial x_1} \\ \frac{\partial V_2}{\partial x_0} & \frac{\partial V_2}{\partial x_1} \end{bmatrix}$$

3×2

$m \times n$

7

Gradient & Jacobian

$f: \mathbb{R}^{n \times m} \to \mathbb{R}^k \qquad f: \mathbb{R}^n \to \mathbb{R}^m$

$$\nabla_{\vec{W}}\, \vec{V} = \begin{bmatrix} \frac{\partial V_0}{\partial w_{00}} & \frac{\partial V_0}{\partial w_{01}} & \frac{\partial V_0}{\partial w_{10}} & \frac{\partial V_0}{\partial w_{11}} \\ \frac{\partial V_1}{\partial w_{00}} & \frac{\partial V_1}{\partial w_{01}} & \frac{\partial V_1}{\partial w_{10}} & \frac{\partial V_1}{\partial w_{11}} \\ \frac{\partial V_2}{\partial w_{00}} & \frac{\partial V_2}{\partial w_{01}} & \frac{\partial V_2}{\partial w_{10}} & \frac{\partial V_2}{\partial w_{11}} \end{bmatrix} \quad \nabla_{\vec{x}}\, \vec{V} = \begin{bmatrix} \frac{\partial V_0}{\partial x_0} & \frac{\partial V_2}{\partial x_0} \\ \frac{\partial V_1}{\partial x_0} & \frac{\partial V_2}{\partial x_1} \\ \frac{\partial V_2}{\partial x_0} & \frac{\partial V_2}{\partial x_1} \end{bmatrix}$$

$k \times (n \times m)$

8

Example 0-Final

$$y = f(\vec{u}) \qquad \vec{u} = g(\vec{v}) \qquad \vec{v} = h(\vec{x})$$

$$\nabla_{\vec{x}} y = \nabla_{\vec{u}} y \cdot \nabla_{\vec{v}} \vec{u} \cdot \nabla_{\vec{x}} \vec{v}$$

$$y = f(u) \qquad u = g(v) \qquad v = g(x)$$

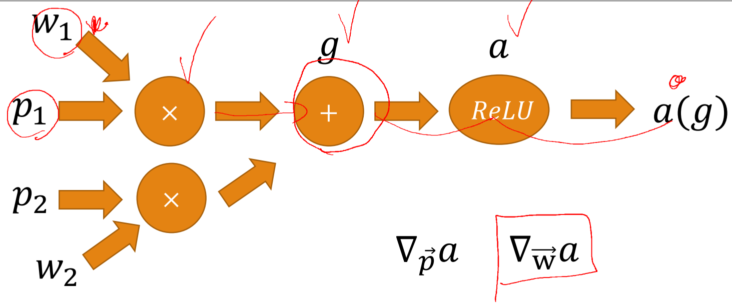
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}\frac{dv}{dx}$$

9

Back-propagation with Chain Rule

10

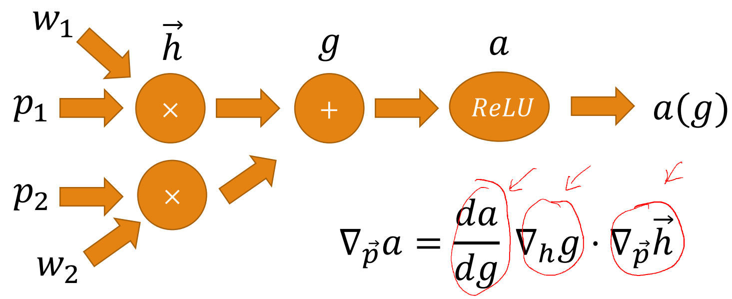
A Unit – case 3



$\nabla_{\vec{p}} a \qquad \boxed{\nabla_{\vec{w}} a}$

11

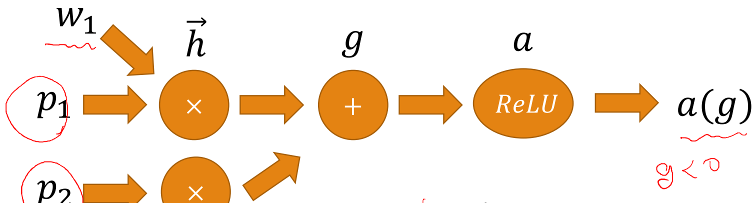
A Unit – case 3



$$\nabla_{\vec{p}} a = \frac{da}{dg} \nabla_h g \cdot \nabla_{\vec{p}} \vec{h}$$

12

A Unit – case 3



$$\nabla_{\vec{w}} a = \begin{cases} [0,0] & \text{if } a = 0 \\ [p_1, p_2] & \text{if } a > 0 \end{cases}$$

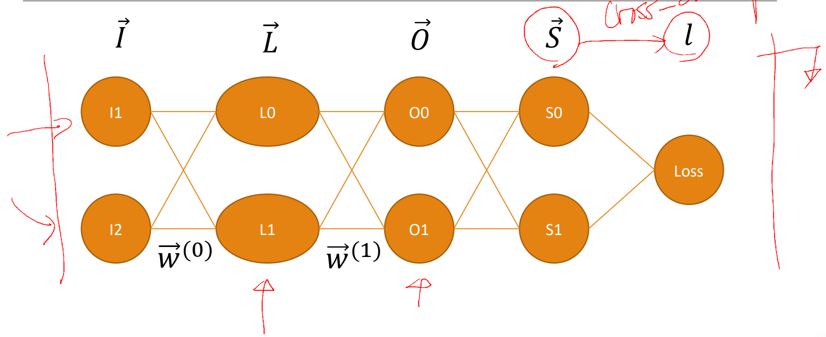
13

1 neuron
↓
network

Back-propagation with Chain Rule

14

Example



15

Example

$\frac{\partial l}{\partial W_{00}^{0}}$, ...

$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$

$\vec{L} = ReLU(\vec{h})$

$\vec{O} = \vec{L} \cdot \vec{W}^{(1)}$

$\vec{S} = softmax(\vec{O})$

$l = cross\_entropy(\vec{S})$

$\vec{W}^{(1)}$

$\nabla_{\vec{w}^{(1)}} l$

$\nabla_{\vec{w}^{(0)}} l$

16

Example

$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$

$\vec{L} = ReLU(\vec{h})$

$\nabla\ 3$ $\vec{O} = \vec{L} \cdot \vec{W}^{(1)}$

$\nabla\ 2$ $\vec{S} = softmax(\vec{O})$

$\nabla\ 1$ $l = cross\_entropy(\vec{S})$

$$\nabla_{\vec{W}^{(1)}} l = \overset{1}{\nabla_{\vec{S}} l} \cdot \overset{2}{\nabla_{\vec{O}} \vec{S}} \cdot \overset{3}{\nabla_{\vec{W}^{(1)}} \vec{O}}$$

$$\nabla_{\vec{W}^{(0)}} l = \underbrace{\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}}_{5}$$

Example $\nabla_{\vec{W}^{(1)}} l = \boxed{\nabla_{\vec{S}} l} \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{W}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[\frac{\partial l}{\partial S_0}, \frac{\partial l}{\partial S_1}, \frac{\partial l}{\partial S_2}, \cdots \frac{\partial l}{\partial S_{n-1}}\right]$$

Example $\nabla_{\vec{W}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{W}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[\frac{\partial l}{\partial S_0}, \frac{\partial l}{\partial S_1}, \frac{\partial l}{\partial S_2}, \cdots \frac{\partial l}{\partial S_{n-1}}\right]$$

$$\frac{\partial l}{\partial S_0}$$

$$l(\vec{S}, label) = -\sum_{i=0}^{n} label_i \log \vec{S}_i$$

$$= \frac{-\partial [label_0 \log S_0 + label_1 \log S_1 + \cdots]}{\partial S_0}$$

Example $\nabla_{\vec{W}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{W}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[\frac{\partial l}{\partial S_0}, \frac{\partial l}{\partial S_1}, \frac{\partial l}{\partial S_2}, \cdots \frac{\partial l}{\partial S_{n-1}}\right]$$

$$\frac{d \log x}{d\ x} = \frac{1}{x}$$

$$l(\vec{S}, label) = -\sum_{i=0}^{n} label_i \log \vec{S}_i$$

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[\frac{\partial l}{\partial S_0}, \frac{\partial l}{\partial S_1}, \frac{\partial l}{\partial S_2}, \cdots \frac{\partial l}{\partial S_{n-1}}\right]$$

$$\frac{d \log x}{d\, x} = \frac{1}{x}$$

$$l(\vec{S}, label) = -\sum_{i=0}^{n} label_i \log \vec{S}_i$$

$$\frac{\partial l}{\partial \vec{S}_0} = -\frac{\partial(label_0 \log \vec{S}_0 + label_1 \log \vec{S}_1 + \cdots + label_n \log \vec{S}_n)}{\partial \vec{S}_0}$$

21

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[\frac{\partial l}{\partial S_0}, \frac{\partial l}{\partial S_1}, \frac{\partial l}{\partial S_2}, \cdots \frac{\partial l}{\partial S_{n-1}}\right]$$

$$\frac{d \log x}{d\, x} = \frac{1}{x}$$

$$\frac{\partial l}{\partial \vec{S}_0} = -\frac{\partial(label_0 \log \vec{S}_0 + \cancel{label_1 \log \vec{S}_1} + \cdots + \cancel{label_n \log \vec{S}_n})}{\partial \vec{S}_0}$$

$$= -\left(\frac{label_0}{\vec{S}_0}\right) \checkmark$$

$- label_0 \dfrac{\partial \log S_0}{\partial S_0}$

$-\dfrac{label_0}{S_0}$

22

Example

$$\frac{\partial l}{\partial s_i} = -\left(\frac{label_i}{s_i}\right)$$

label - [0, 1]

$$\nabla_{\vec{S}} l = \left[-\frac{label_0}{\vec{S}_0}, \ldots, -\frac{label_{n-1}}{\vec{S}_{n-1}}\right]$$

0

$$\nabla_{\vec{S}} l = \left[-\frac{label_0}{\vec{S}_0}, -\frac{label_1}{\vec{S}_1}\right] \leftarrow \text{2차원}$$

23

Example $\nabla_{\vec{w}^{(1)}} l = \boxed{\nabla_{\vec{S}} l} \cdot \boxed{\nabla_{\vec{O}} \vec{S}} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{O}} \vec{S} = \begin{bmatrix} \frac{\partial S_0}{\partial O_0} & \frac{\partial S_0}{\partial O_1} & \cdots & \frac{\partial S_0}{\partial O_{n-1}} \\ \frac{\partial S_1}{\partial O_0} & \frac{\partial S_1}{\partial O_1} & \cdots & \frac{\partial S_1}{\partial O_{n-1}} \\ \frac{\partial S_2}{\partial O_0} & \cdots & \cdots & \frac{\partial S_2}{\partial O_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial S_{n-1}}{\partial O_0} & \cdots & \cdots & \frac{\partial S_{n-1}}{\partial O_{n-1}} \end{bmatrix}$$

$\vec{S} \rightarrow \vec{O}$

24

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\vec{S}_i(\vec{O}) = \frac{e^{o_i}}{\sum_{j=0}^{n} e^{o_j}}$$
$$= \frac{e^{o_i}}{e^{o_0}+e^{o_1}+\cdots+e^{o_n}}$$

$$\frac{\partial S_i}{\partial O_j} = \partial\left(\frac{e^{O_i}}{e^{O_0}+e^{O_1}+\cdots+e^{O_n}}\right) \Big/ \partial O_0$$

$$\frac{d\vec{S}_i}{d\vec{O}_j} = \frac{d\left(\frac{e^{o_i}}{e^{o_0}+e^{o_1}+\cdots+e^{o_n}}\right)}{do_j}$$

25

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\frac{d\vec{S}_i}{d\vec{O}_j} = \frac{d\left(\frac{e^{o_i}}{e^{o_0}+e^{o_1}+\cdots+e^{o_n}}\right)}{do_j}$$

$$\frac{\partial \frac{a}{b}}{\partial c} = \frac{\partial a b^{-1}}{\partial c} = a\frac{\partial b^{-1}}{\partial c} + b^{-1}\frac{\partial a}{\partial c}$$

$$= e^{o_i}\frac{d\left(\frac{1}{e^{o_0}+e^{o_1}+\cdots+e^{o_n}}\right)}{do_j} + \frac{1}{e^{o_0}+e^{o_1}+\cdots+e^{o_n}}\frac{d(e^{o_i})}{do_j}$$

$$= -e^{o_i}\frac{1}{(e^{o_0}+e^{o_1}+\cdots+e^{o_n})^2}\frac{d(e^{o_j})}{do_j} +$$

$$\frac{\partial (a+b+c)}{\partial c} = \frac{\partial \square^{-1}}{\partial c} = \frac{\partial \square^{-1}}{\partial \square}\frac{\partial \square}{\partial c}$$

$$= -\frac{e^{o_i+o_j}}{(e^{o_0}+e^{o_1}+\cdots+e^{o_n})^2} \quad \text{if } i \neq j$$

$$= \frac{e^{o_i}(e^{o_0}+e^{o_1}+\cdots+e^{o_{i-1}}+\cdots+e^{o_n})}{(e^{o_0}+e^{o_1}+\cdots+e^{o_n})^2} \quad \text{if } i = j$$

$$= -\square^{-2}\frac{\partial \square}{\partial c}$$

26

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

if $i \neq j$

$$-\frac{e^{o_i+o_j}}{(e^{o_0}+e^{o_1}+\cdots+e^{o_n})^2} = -\vec{S}_i \cdot \vec{S}_j$$

$e^{O_i} \times 1 - e^{O_i}$

if $i = j$

$$\frac{e^{o_i}(e^{o_0}+e^{o_1}+\cdots+e^{o_{i-1}}+\cdots+e^{o_n})}{(e^{o_0}+e^{o_1}+\cdots+e^{o_n})^2} = \vec{S}_i \cdot (1-\vec{S}_i)$$

$$\frac{e^{O_i}}{\Sigma e^{O_i}}\left(1-\frac{e^{O_i}}{\Sigma e^{O_i}}\right)$$

27

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$\frac{\partial S_0}{\partial O_0}$

$$\nabla_{\vec{O}}\vec{S} = \begin{bmatrix} S_0(1-S_0) & -S_0S_1 & \cdots & -S_0S_{n-1} \\ -S_1S_0 & S_1(1-S_1) & \cdots & -S_1S_{n-1} \\ -S_2S_0 & \cdots & \cdots & -S_2S_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ -S_nS_0 & \cdots & \cdots & S_{n-1}(1-S_{n-1}) \end{bmatrix}$$

28

Example $\nabla_{\vec{w}^{(1)}} l = \boxed{\nabla_{\vec{S}} l} \cdot \boxed{\nabla_{\vec{O}} \vec{S}} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{S}} l = \left[-\frac{label_0}{\vec{S}_0}, -\frac{label_1}{\vec{S}_1} \right]$$

$$\nabla_{\vec{O}} \vec{S} = \begin{bmatrix} S_0(1-S_0) & -S_0 S_1 \\ -S_1 S_0 & S_1(1-S_1) \end{bmatrix}$$

0

0

29

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} \frac{\partial O_0}{\partial w_{00}} & \frac{\partial O_0}{\partial w_{01}} & \cdots & \frac{\partial O_0}{\partial w_{kl}} \\ \frac{\partial O_1}{\partial w_{00}} & \frac{\partial O_1}{\partial w_{01}} & \cdots & \frac{\partial O_1}{\partial w_{kl}} \\ \frac{\partial O_2}{\partial w_{00}} & \frac{\partial O_2}{\partial w_{01}} & \cdots & \frac{\partial O_2}{\partial w_{kl}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial O_{n-1}}{\partial w_{00}} & \frac{\partial O_{n-1}}{\partial w_{01}} & \cdots & \frac{\partial O_{n-1}}{\partial w_{kl}} \end{bmatrix}$$

$$\vec{O} = \vec{L} \cdot \vec{W}^{(1)}$$

30

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\vec{O_j} = \sum_i L_i W_{ij} \rightarrow [L_0 \quad L_1 \quad L_2] \cdot \begin{bmatrix} W_{00} & W_{01} & W_{02} \\ W_{10} & W_{11} & W_{12} \\ W_{20} & W_{21} & W_{22} \end{bmatrix} = [L_0 W_{00} + L_1 W_{10} + L_2 W_{20}, \ldots]$$

O_0 ''

$$\frac{\partial \vec{O_j}}{\partial W_{kl}} = \frac{\partial \left(\sum_i L_i W_{ij} \right)}{\partial W_{kl}} = \begin{cases} 0 & if\ j \neq l \\ L_k & if\ j = l \end{cases}$$

$$\frac{\partial\, L_0 W_{00} + L_1 \cancel{W_{10}} + L_2 \cancel{W_{20}}}{\partial w_{01}}$$

??

$= L_0$ $\qquad = 0$

31

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & \ldots & 0 \\ 0 & L_0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & L_{n-1} \end{bmatrix}$$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} \frac{\partial O_0}{\partial W_{00}} & \frac{\partial O_0}{\partial W_{01}} & \frac{\partial O_0}{\partial W_{10}} & \frac{\partial O_0}{\partial W_{11}} \\ \frac{\partial O_1}{\partial W_{00}} & \frac{\partial O_1}{\partial W_{01}} & \frac{\partial O_1}{\partial W_{10}} & \frac{\partial O_1}{\partial W_{11}} \end{bmatrix}$$

L_0 0 L_1 L_1 0 L_0

2×2 4 2 2×4

32

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & \dots & 0 \\ 0 & L_0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_{n-1} \end{bmatrix}$$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & L_1 & 0 \\ 0 & L_0 & 0 & L_1 \end{bmatrix}$$

33

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & \dots & 0 \\ 0 & L_0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_{n-1} \end{bmatrix}$$

$$\underset{2\text{x}2 \quad 2}{\nabla_{\vec{w}^{(1)}} \vec{O}} = \begin{bmatrix} L_0 & 0 & L_1 & 0 \\ 0 & L_0 & 0 & L_1 \end{bmatrix}$$

34

Example $\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & \dots & 0 \\ 0 & L_0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_{n-1} \end{bmatrix}$$

$W^{(1)}$

$\nabla_W \vec{O} =$

2x3 3

$$\underset{2\text{x}2 \quad 2}{\nabla_{\vec{w}^{(1)}} \vec{O}} = \begin{bmatrix} L_0 & 0 & L_1 & 0 \\ 0 & L_0 & 0 & L_1 \end{bmatrix}$$

$$\underset{3\text{x}2 \quad 2}{\nabla_{\vec{w}^{(1)}} \vec{O}} = \begin{bmatrix} L_0 & 0 & L_1 & 0 & L_2 & 0 \\ 0 & L_0 & 0 & L_1 & 0 & L_2 \end{bmatrix}$$

2x6

35

Example $\nabla_{\vec{w}^{(1)}} l = \boxed{\nabla_{\vec{S}} l} \cdot \boxed{\nabla_{\vec{O}} \vec{S}} \cdot \boxed{\nabla_{\vec{w}^{(1)}} \vec{O}}$

$$\nabla_{\vec{S}} l = \left[-\frac{label_0}{\vec{S}_0}, -\frac{label_1}{\vec{S}_1} \right]$$

$$\nabla_{\vec{O}} \vec{S} = \begin{bmatrix} S_0(1 - S_0) & -S_0 S_1 \\ -S_1 S_0 & S_1(1 - S_1) \end{bmatrix}$$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix} L_0 & 0 & L_1 & 0 \\ 0 & L_0 & 0 & L_1 \end{bmatrix}$$

36

Example

$\vec{h} = \vec{I} \cdot \overrightarrow{W}^{(0)}$
$\vec{L} = ReLU(\vec{h})$
$\vec{O} = \vec{L} \cdot \overrightarrow{W}^{(1)}$
$\vec{S} = softmax(\vec{O})$
$l = cross\_entropy(\vec{S})$

$\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$

$\nabla_{\vec{w}^{(0)}} l =$
$\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\overrightarrow{W}^{(0)}} \vec{h}$

37

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\overrightarrow{W}^{(0)}} \vec{h}$

$\nabla_{\vec{S}} l \quad \nabla_{\vec{O}} \vec{S} \quad \nabla_{\vec{w}^{(1)}} \vec{O}$

$\nabla_{\vec{L}} \vec{O} = ?$

$\nabla_{\vec{h}} \vec{L} = ?$

$\nabla_{\overrightarrow{W}^{(0)}} \vec{h} = ?$

38

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\overrightarrow{W}^{(0)}} \vec{h}$

$$\vec{O_j} = \sum_i L_i W_{ij}$$

$O_0 = L_0 W_{00} + L_1 W_{10} + L_2 W_{20} + \cdots$

$$\nabla_{\vec{L}} \vec{O} = \begin{bmatrix} \frac{\partial O_0}{\partial L_0} & \cdots & \frac{\partial O_0}{\partial L_{n-1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial O_{m-1}}{\partial L_0} & \cdots & \frac{\partial O_{m-1}}{\partial L_{n-1}} \end{bmatrix} = \begin{bmatrix} W_{00} & \cdots & W_{n-1\,0} \\ \vdots & \ddots & \vdots \\ W_{0\,m-1} & \cdots & W_{n-1\,m-1} \end{bmatrix}$$

n m

$m \times n$

39

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\overrightarrow{W}^{(0)}} \vec{h}$

W_{01}

$$\nabla_{\vec{L}} \vec{O} = \begin{bmatrix} W_{00} & W_{10} \\ W_{01} & W_{11} \end{bmatrix} = \boxed{\overrightarrow{W}^{(1)T}}$$

40

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}}\vec{S} \cdot \nabla_{\vec{L}}\vec{O} \cdot \nabla_{\vec{h}}\vec{L} \cdot \nabla_{\vec{W}^{(0)}}\vec{h}$

$\vec{h} = ReLU(\vec{L})$

$h_0 = ReLU(L_0)$

hxn

$$\nabla_{\vec{h}}\vec{L} = \begin{bmatrix} \begin{cases}0\\1\end{cases} & 0 & 0 & \cdots \\ 0 & \begin{cases}0\\1\end{cases} & 0 & \cdots \\ 0 & 0 & \begin{cases}0\\1\end{cases} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

41

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}}\vec{S} \cdot \nabla_{\vec{L}}\vec{O} \cdot \nabla_{\vec{h}}\vec{L} \cdot \nabla_{\vec{W}^{(0)}}\vec{h}$

$$\nabla_{\vec{L}}\vec{O} = \begin{bmatrix} W_{00} & W_{10} \\ W_{01} & W_{11} \end{bmatrix} = \vec{W}^{(1)^T}$$

$$\nabla_{\vec{h}}\vec{L} = \begin{bmatrix} 0 \; or \; 1 & 0 \\ 0 & 0 \; or \; 1 \end{bmatrix}$$

42

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}}\vec{S} \cdot \nabla_{\vec{L}}\vec{O} \cdot \nabla_{\vec{h}}\vec{L} \cdot \nabla_{\vec{W}^{(0)}}\vec{h}$

$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$

$\nabla_{\vec{W}^{(0)}}\vec{h} =$

43

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}}\vec{S} \cdot \nabla_{\vec{L}}\vec{O} \cdot \nabla_{\vec{h}}\vec{L} \cdot \nabla_{\vec{W}^{(0)}}\vec{h}$

$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$

$$\frac{\partial \vec{O_j}}{\partial W_{kl}} = \frac{\partial\left(\sum_i L_i W_{ij}\right)}{\partial W_{kl}} = \begin{cases} 0 & if \; j \neq l \\ L_k & if \; j = l \end{cases}$$

$\nabla_{\vec{W}^{(0)}}\vec{h} =$

$$\frac{\partial \vec{h_j}}{\partial W_{kl}} = \frac{\partial\left(\sum_i I_i W_{ij}\right)}{\partial W_{kl}} = \begin{cases} 0 & if \; j \neq l \\ I_k & if \; j = l \end{cases}$$

44

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$

$$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$$

$$\nabla_{\vec{w}^{(o)}} \vec{h} = \begin{bmatrix} I_0 & 0 & \dots & 0 \\ 0 & I_0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{n-1} \end{bmatrix}$$

$$\underset{3\times2 \quad 2}{\nabla_{\vec{w}^{(0)}} \vec{h}} = \begin{bmatrix} I_0 & 0 & I_1 & 0 & I_2 & 0 \\ 0 & I_0 & 0 & I_1 & 0 & I_2 \end{bmatrix}$$

2x6

45

Example $\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$

$$\nabla_{\vec{L}} \vec{O} = \begin{bmatrix} W_{00} & W_{10} \\ W_{01} & W_{11} \end{bmatrix} = \vec{W}^{(1)T}$$

$$\nabla_{\vec{h}} \vec{L} = \begin{bmatrix} 0 \text{ or } 1 & 0 \\ 0 & 0 \text{ or } 1 \end{bmatrix}$$

$$\nabla_{\vec{w}^{(0)}} \vec{h} = \begin{bmatrix} I_0 & 0 & I_1 & 0 \\ 0 & I_0 & 0 & I_1 \end{bmatrix}$$

46

Example

$$\vec{h} = \vec{I} \cdot \vec{W}^{(0)}$$
$$\vec{L} = ReLU(\vec{h})$$
$$\vec{O} = \vec{L} \cdot \vec{W}^{(1)}$$
$$\vec{S} = softmax(\vec{O})$$
$$l = cross\_entropy(\vec{S})$$

$$\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$$

$$\nabla_{\vec{w}^{(0)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$$

Back

OO

47

Example

$$\nabla_{\vec{w}^{(0)}} l =$$
$$\nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}^{(1)}} \vec{O} \cdot \nabla_{\vec{h}^{(1)}} \vec{L}^{(1)} \cdot \nabla_{\vec{L}^{(1)}} \vec{h}^{(1)} \cdot \nabla_{\vec{w}^{(0)}} \vec{I}$$

48

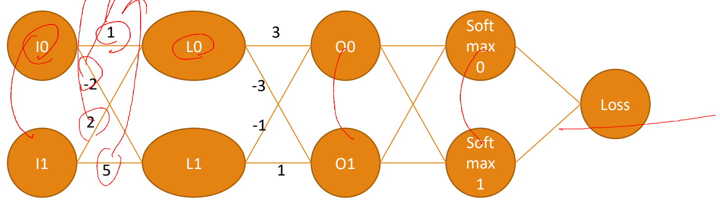
일반화

W = ? ←
I = ? ←

Back-propagation with Numbers

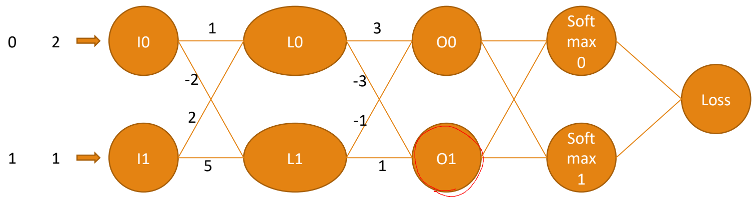
49

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}$$



50

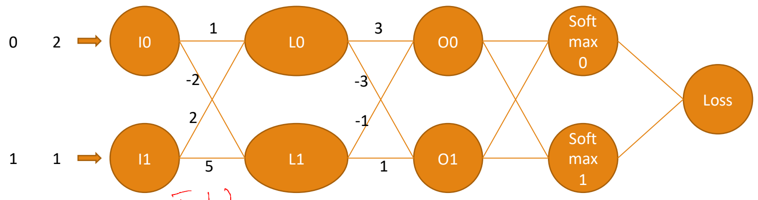
$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}$$



$I = [2 \quad 1]$
$label = [0 \quad 1]$

51

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}$$

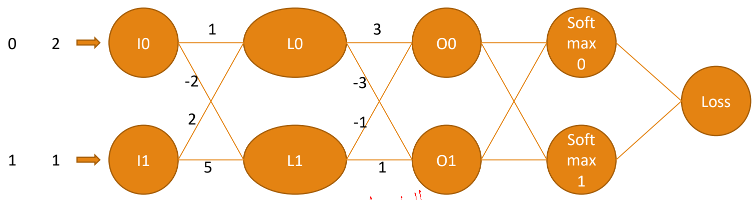


I·W ?

$I = [2 \quad 1]$ $\quad h = [4 \quad 1]$
$label = [0 \quad 1]$ $\quad L = [4 \quad 1]$

52

$$W^0 = \begin{bmatrix}1 & -2\\ 2 & 5\end{bmatrix} \quad W^1 = \begin{bmatrix}3 & -3\\ -1 & 1\end{bmatrix}$$



$L \cdot W^1$

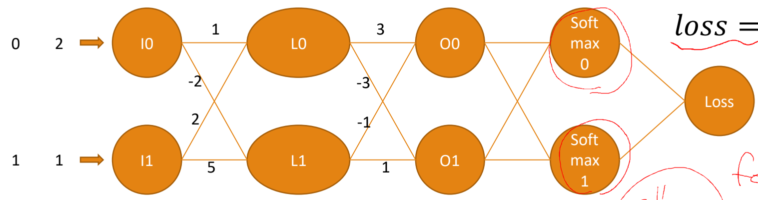
$I = [2 \quad 1]$ $h = [4 \quad 1]$

$label = [0 \quad 1]$ $L = [4 \quad 1]$ $O = [11 \quad -11]$

53

$$W^0 = \begin{bmatrix}1 & -2\\ 2 & 5\end{bmatrix} \quad W^1 = \begin{bmatrix}3 & -3\\ -1 & 1\end{bmatrix}$$

$l = -(0 \times \log + 1 \times 2.7 \times 10^{-10})$



$loss = 22$

feed-forward

$I = [2 \quad 1]$ $h = [4 \quad 1]$ $O = [11 \quad -11]$

$label = [0 \quad 1]$ $L = [4 \quad 1]$ $S = [1 \quad 2.7E-10]$

$\frac{e^{11}}{e^{11}+e^{-11}}$ $\frac{e^{-11}}{e^{11}+e^{-11}}$

54

$I = [2 \quad 1]$ $h = [4 \quad 1]$ $O = [11 \quad -11]$

$label = [0 \quad 1]$ $L = [4 \quad 1]$ $S = [1 \quad 2.7E-10]$

$$W^0 = \begin{bmatrix}1 & -2\\ 2 & 5\end{bmatrix} \quad W^1 = \begin{bmatrix}3 & -3\\ -1 & 1\end{bmatrix} \quad l = 6.9$$

$$\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$$

$$\nabla_{\vec{S}} l = \left[-\frac{label_0}{S_0}, -\frac{label_1}{S_1}\right]$$

$$\nabla_{\vec{O}} \vec{S} = \begin{bmatrix}S_0(1-S_0) & -S_0S_1\\ -S_1S_0 & S_1(1-S_1)\end{bmatrix}$$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix}L_0 & 0 & L_1 & 0\\ 0 & L_0 & 0 & L_1\end{bmatrix}$$

55

$I = [2 \quad 1]$ $h = [4 \quad 1]$ $O = [11 \quad -11]$

$label = [0 \quad 1]$ $L = [4 \quad 1]$ $S = [1 \quad 2.7E-10]$

$$W^0 = \begin{bmatrix}1 & -2\\ 2 & 5\end{bmatrix} \quad W^1 = \begin{bmatrix}3 & -3\\ -1 & 1\end{bmatrix} \quad l = 22$$

$$\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$$

$$\nabla_{\vec{S}} l = \left[-\frac{0}{1}, -\frac{1}{2.7e-10}\right] = [0 \quad -3.58e10]$$

$$\nabla_{\vec{O}} \vec{S} = \begin{bmatrix}0 & -2.7e-10\\ -2.7e-10 & 2.7e-10(1-2.7e-10)\end{bmatrix} = \begin{bmatrix}0 & -2.7e-10\\ -2.7e-10 & 2.7e-10\end{bmatrix}$$

$$\nabla_{\vec{w}^{(1)}} \vec{O} = \begin{bmatrix}4 & 0 & 1 & 0\\ 0 & 4 & 0 & 1\end{bmatrix}$$

56

$I = [2 \quad 1]$ $\quad h = [4 \quad 1]$ $\quad O = [11 \quad -11]$

$label = [0 \quad 1]$ $\quad L = [4 \quad 1]$ $\quad S = [0.999 \quad 0.001]$

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} \quad l = 6.9$$

$$\nabla_{\vec{w}^{(1)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{w}^{(1)}} \vec{O}$$

$$\nabla_{\vec{w}^{(1)}} l = [0 \quad -3.58e10] \cdot \begin{bmatrix} 0 & -2.7e-10 \\ -2.7e-10 & 2.7e-10 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 & 1 & 0 \\ 0 & 4 & 0 & 1 \end{bmatrix}$$

$$= [4, -4, 1, -1]$$

이것의 의미는 무엇인가?

$$\left[\frac{\partial l}{\partial w_{00}} \quad \frac{\partial l}{\partial w_{01}} \quad \frac{\partial l}{\partial w_{10}} \quad \frac{\partial l}{\partial w_{00}} \right]$$

57

$I = [2 \quad 1]$ $\quad h = [4 \quad 1]$ $\quad O = [11 \quad -11]$

$label = [0 \quad 1]$ $\quad L = [4 \quad 1]$ $\quad S = [0.999 \quad 0.001]$

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} \quad l = 6.9$$

$$\nabla_{\vec{w}^{(0)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$$

$$\nabla_{\vec{L}} \vec{O} = \begin{bmatrix} W_{00} & W_{10} \\ W_{01} & W_{11} \end{bmatrix} = \vec{W}^{(1)T}$$

$$\begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix}$$

$$\nabla_{\vec{h}} \vec{L} = \begin{bmatrix} 0 \text{ or } 1 & 0 \\ 0 & 0 \text{ or } 1 \end{bmatrix} \quad \nabla_{\vec{w}^{(0)}} \vec{h} = \begin{bmatrix} I_0 & 0 & I_1 & 0 \\ 0 & I_0 & 0 & I_1 \end{bmatrix}$$

2 1 2 1

58

$I = [2 \quad 1]$ $\quad h = [4 \quad 1]$ $\quad O = [11 \quad -11]$

$label = [0 \quad 1]$ $\quad L = [4 \quad 1]$ $\quad S = [0.999 \quad 0.001]$

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} \quad l = 6.9$$

$$\nabla_{\vec{w}^{(0)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$$

$$\nabla_{\vec{L}} \vec{O} = \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix}$$

$$\nabla_{\vec{h}} \vec{L} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\nabla_{\vec{w}^{(0)}} \vec{h} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix}$$

59

$I = [2 \quad 1]$ $\quad h = [4 \quad 1]$ $\quad O = [11 \quad -11]$

$label = [0 \quad 1]$ $\quad L = [4 \quad 1]$ $\quad S = [0.999 \quad 0.001]$

$$W^0 = \begin{bmatrix} 1 & -2 \\ 2 & 5 \end{bmatrix} \quad W^1 = \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} \quad l = 6.9$$

$$\nabla_{\vec{w}^{(0)}} l = \nabla_{\vec{S}} l \cdot \nabla_{\vec{O}} \vec{S} \cdot \nabla_{\vec{L}} \vec{O} \cdot \nabla_{\vec{h}} \vec{L} \cdot \nabla_{\vec{W}^{(0)}} \vec{h}$$

$$\nabla_{\vec{W}^{(0)}} l = [0 \quad -3.58e10] \cdot \begin{bmatrix} 0 & -2.7e-10 \\ -2.7e-10 & 2.7e-10 \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix}$$

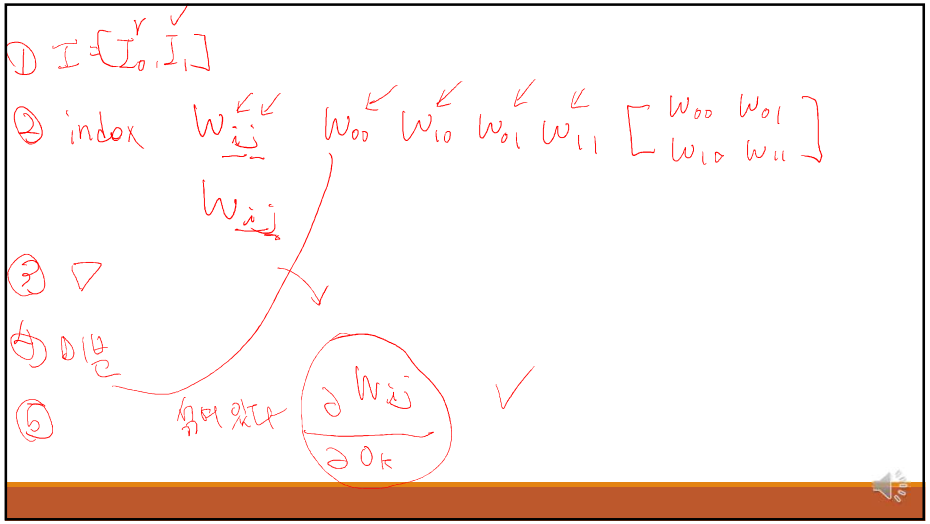
$$= [12, -4, 6, -2]$$

$$w - \alpha \frac{dl}{dw}$$

$$1 - \alpha 12 = w_{00}$$

60

61


① $I = [I_0, I_1]$
② index W_{ij}
W_{00} W_{10} W_{01} W_{11}
$\begin{bmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{bmatrix}$
W_{ij}
③ ∇
④ 미분
⑤
숨어있다
$\frac{\partial W_{ij}}{\partial O_k}$

62