Deep Learning Basic

Jaewon Kim, Dankook Univ.

Chapter 2-1

Part 1

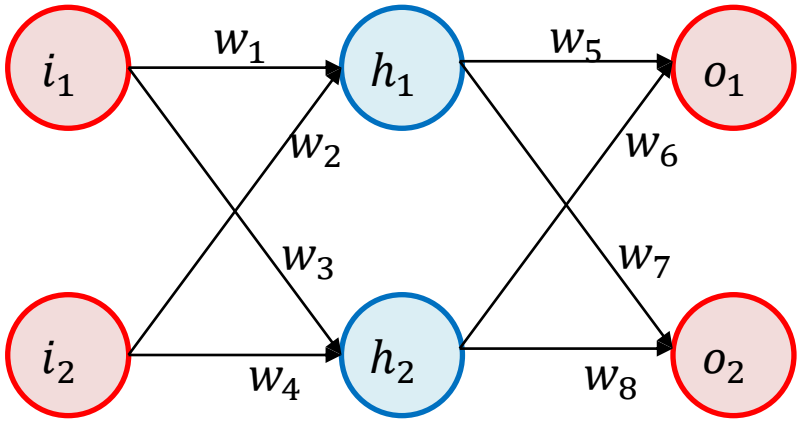
Backpropagation

Backpropagation

How to update weight parameter?

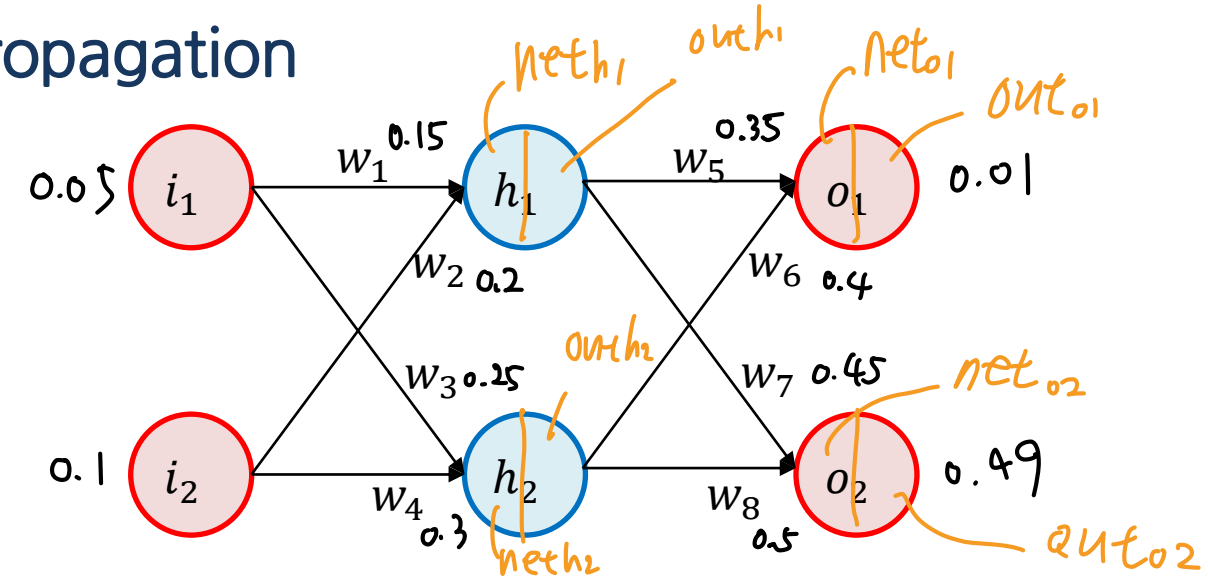
$$W := W - \frac{\partial}{\partial W} cost(W, b)$$

Backpropagation



Input : 0.05, 0.1
Output : 0.01, 0.99
Weight initialization

Backpropagation



h_1의 input : $0.05 \times 0.15 + 0.1 \times 0.2 = 0.0275$ (net_{h_1})

$sig = \frac{1}{1+e^{-x}} \longrightarrow Out\ h_1 = \frac{1}{1+e^{-0.0275}} = 0.5069$

h_2의 input : $0.05 \times 0.25 + 0.1 \times 0.3 = 0.0425$

(net_{h_2})

$out\ h_2 = 0.5106$

$out\ h_1$

$out\ h_2$

Backpropagation

$net_{o1} = 0.5069 \times 0.35 + 0.5106 \times 0.4 = 0.3817$

$out_{o1} = \frac{1}{1 + e^{-0.3817}} = 0.5943...$

$net_{o2} = 0.5069 \times 0.45 + 0.5069 \times 0.5 = 0.4834$

$out_{o2} = \frac{1}{1 + e^{-0.4834}} = 0.6186$

$out_{o1} = 0.5943 \longrightarrow 0.01$ target...

$out_{o2} = 0.6186 \longrightarrow 0.99$

Backpropagation

$$E_{total} = E_{o1} + E_{o2}$$

$$= \frac{1}{2}\left[(out_{o1} - target_{o1})^2 + (out_{o2} - target_{o2})^2 \right]$$

대입 후 계산

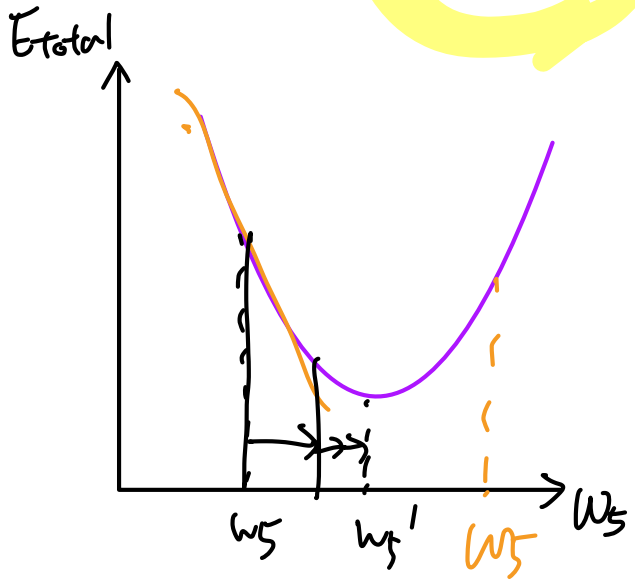
$$E_{total} = 0.2391$$

Our Goal : 0.2391 ⟶ 0

"Gradient Decent"

w_5 Update Processing

- Gradient Decent $\frac{\partial E_{total}}{\partial w_5}$



if $w_5 < w_5'$ → 오른쪽 이동.

$w_5 > w_5'$ → 왼쪽 이동.

$$\frac{\partial E_{total}}{\partial w_5}$$

$$w := w + \frac{\partial}{\partial w_5} E_{total}$$

w_5 Update Processing

- Gradient Decent

$$E_{total} = \frac{1}{2}\left[(out_{o1} - target_{o1})^2 + (out_{o2} - target_{o2})^2 \right]$$

w_5 Update Processing

- Chain Rule . "연쇄 미분법"

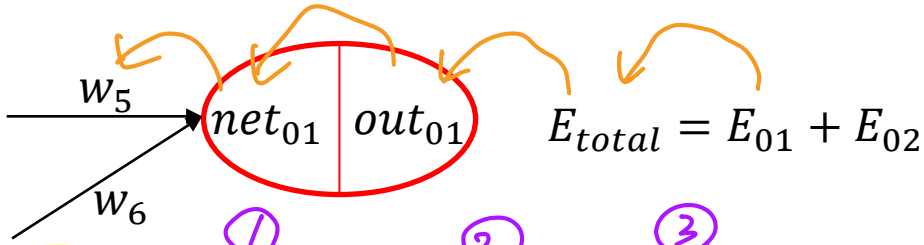
$$Sigmoid(x) = \frac{1}{1+e^{-x}}$$

↓ 미분.

$$Sig(x)\left(1 - Sig(x)\right)$$

w_5 Update Processing

- Chain Rule



$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{01}} \times \frac{\partial out_{01}}{\partial net_{01}} \times \frac{\partial net_{01}}{\partial w_5}$$

① $\dfrac{\partial E_{total}}{\partial out_{01}} = -(target_{01} - out_{01}) = -[0.01 - 0.5943) = 0.5843$

② $\dfrac{\partial out_{01}}{\partial net_{01}} = out_{01}(1 - out_{01}) = 0.5943(1 - 0.5943) = 0.2411$

③ $\dfrac{\partial net_{01}}{\partial w_5} = out_{h1} \times w_5 + out_{h2} \times w_6 \Rightarrow out_{h1} = 0.5069$

w_5 Update Processing

- Chain Rule

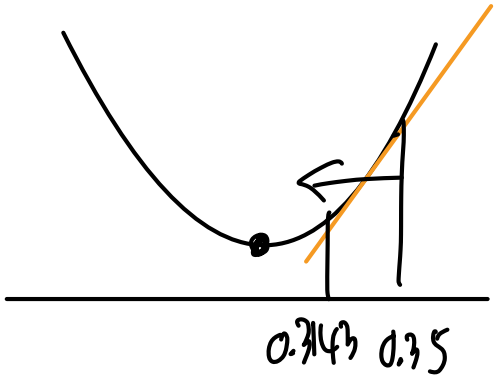
① × ② × ③ = $\frac{\partial E_{total}}{\partial w_5}$

$$\frac{\partial E_{total}}{\partial w_5} = 0.5843 \times 0.2411 \times 0.5069 = 0.0714$$

$$w_5' := w_5 - \alpha \frac{\partial E_{total}}{\partial w_5} \qquad \alpha = 0.5 \text{ 이라고 하자}$$

$$w_5' = 0.35 - 0.5 \times 0.0714$$

$$= 0.3143.$$



$w_5 = 0.35 \longrightarrow 0.3143$

$target_{o1} = 0.01$ $\quad$ $out_{o1} = 0.5943$

w_6 Update Processing

$$w_6^+ = w_6 - \alpha \frac{\partial E_{total}}{\partial w_6}$$

$$\frac{\partial E_{total}}{\partial w_6} = \frac{\partial E_{total}}{\partial out_{o1}} \times \frac{\partial out_{o1}}{\partial net_{o1}} \times \frac{\partial net_{o1}}{\partial w_6}$$

$\frac{\partial E_{total}}{\partial out_{o1}} = 0.7843$

$\frac{\partial out_{o1}}{\partial net_{o1}} = 0.2411$

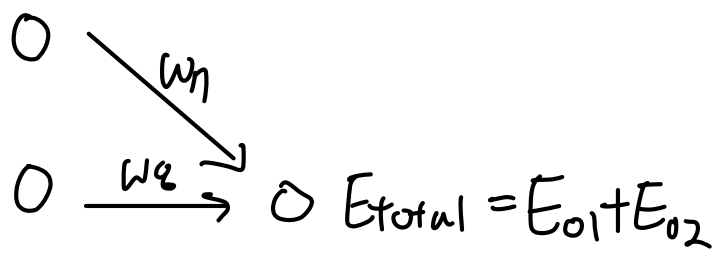
$$\frac{\partial net_{o1}}{\partial w_6} = out_{h2} = 0.5106$$

$$\frac{\partial E_{total}}{\partial w_6} = 0.7843 \times 0.2411 \times 0.5106 = 0.07193$$

$$w_6^+ = 0.4 - 0.5 \times 0.07193 = 0.364$$

w_7 Update Processing

$$W_7^+ = W_7 - \alpha \frac{\partial E_{total}}{\partial W_7}$$



$$\frac{\partial E_{total}}{\partial W_7} = \frac{\partial E_{total}}{\partial out_{O2}} \times \frac{\partial out_{O2}}{\partial net_{O2}} \times \frac{\partial net_{O2}}{\partial W_7}$$

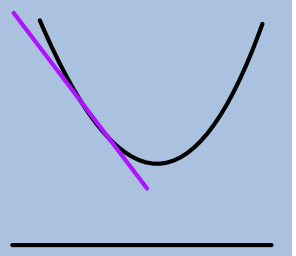
$$\frac{\partial E_{total}}{\partial out_{O2}} = -(target_{O2} - out_{O2}) = -(0.99 - 0.6186) = -0.3714$$

$$\frac{\partial out_{O2}}{\partial net_{O2}} = out_{O2}(1 - out_{O2}) = 0.6186\ (1 - 0.618) = 0.2359$$

$$\frac{\partial net_{O2}}{\partial W_7} = out_{H1} = 0.5069$$

$$\frac{\partial E_{total}}{\partial W_7} = (-0.3714) \times [0.2359) \times (0.5069) = -0.04441$$

$$W_7^+ = W_7 - \alpha \frac{\partial E_{total}}{\partial W_7} = 0.45 + (0.5) \times (+0.04441) = 0.4722$$



w_8 Update Processing

$$W_8^+ = W_8 - \alpha \frac{\partial E_{total}}{\partial W_8}$$

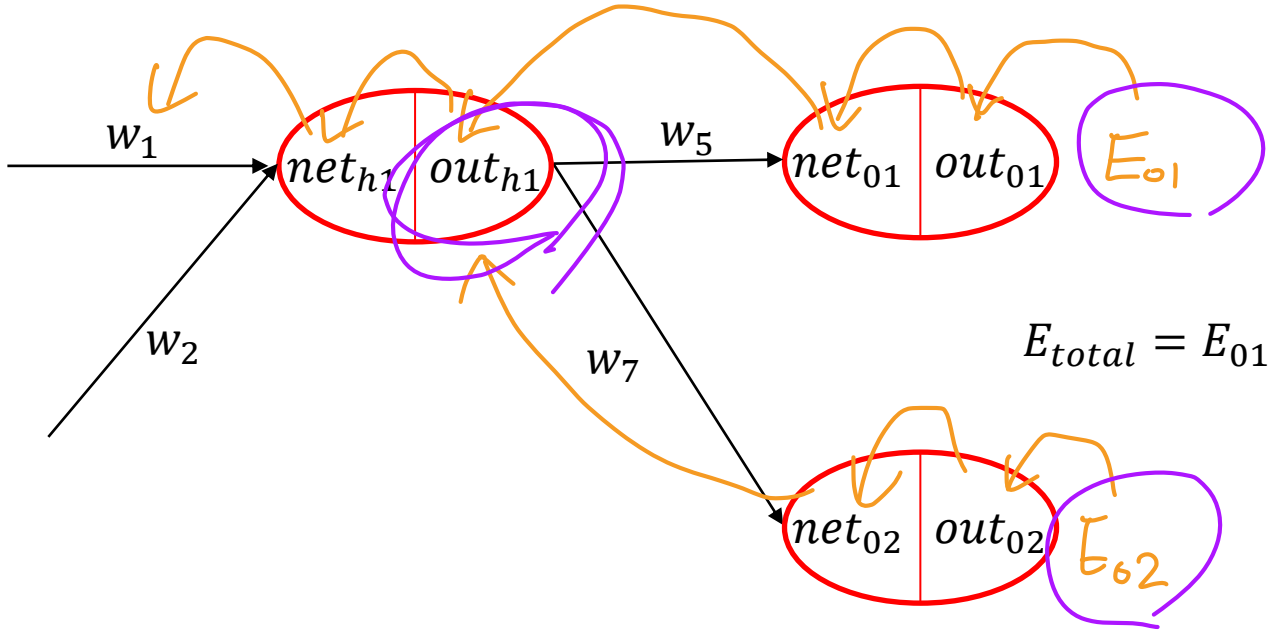
$$\frac{\partial E_{total}}{\partial W_8} = \frac{\partial E_{total}}{\partial out_{o2}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \times \frac{\partial net_{o2}}{\partial W_8}$$

$$= -0.3714 \times 0.2359 \times (out_{h2} = 0.5106)$$

$$= -0.044474$$

$$W_8^+ = 0.5 - (0.5) \times (-0.044474) = 0.5224$$

w_1 Update Processing



$$E_{total} = E_{01} + E_{02}$$

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} \times \frac{\partial out_{h1}}{\partial net_{h1}} \times \frac{\partial net_{h1}}{\partial w_1}$$

w_1 Update Processing

$$\frac{\partial E_{total}}{\partial out_{h_1}} = \frac{\partial E_{o1}}{\partial out_{h_1}} + \frac{\partial E_{o2}}{\partial out_{h_1}}$$

$$= \frac{\partial E_{o1}}{\partial net_{o1}} \times \frac{\partial net_{o1}}{\partial out_{h_1}} + \frac{\partial E_{o2}}{\partial net_{o2}} \times \frac{\partial net_{o2}}{\partial out_{h_1}}$$

$$= \left(\frac{\partial E_{o1}}{\partial out_{o1}} \times \frac{\partial out_{o1}}{\partial net_{o1}} \right) \times \frac{\partial net_{o1}}{\partial out_{h_1}} + \left(\frac{\partial E_{o2}}{\partial out_{o2}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \right) \times \frac{\partial net_{o2}}{\partial out_{h_1}}$$

w_1 Update Processing

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h_1}} \times \frac{\partial out_{h_1}}{\partial net_{h_1}} \times \frac{\partial net_{h_1}}{\partial w_1}$$

$$\left[\left(\frac{\partial E_{o1}}{\partial out_{o1}} \times \frac{\partial out_{o1}}{\partial net_{o1}} \right) \times \frac{\partial net_{o1}}{\partial out_{h_1}} + \left(\frac{\partial E_{o2}}{\partial out_{o2}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \right) \times \frac{\partial net_{o2}}{\partial out_{h_1}} \right] \times \frac{\partial out_{h_1}}{\partial net_{h_1}} \times \frac{\partial net_{h_1}}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial w_1} = \left[0.5843 \times 0.2411 \times 0.35 + (-0.3714) \times 0.2359 \times 0.45 \right] \times 0.25 \times 0.05$$

$$= 0.0001235$$

$$w_1^+ = w_1 - \alpha \frac{\partial E_{total}}{\partial w_1} = 0.15 - 0.5 \times (0.0001235) = 0.1499$$

$$0.15 \longrightarrow 0.1499$$

"gradient vanishing"

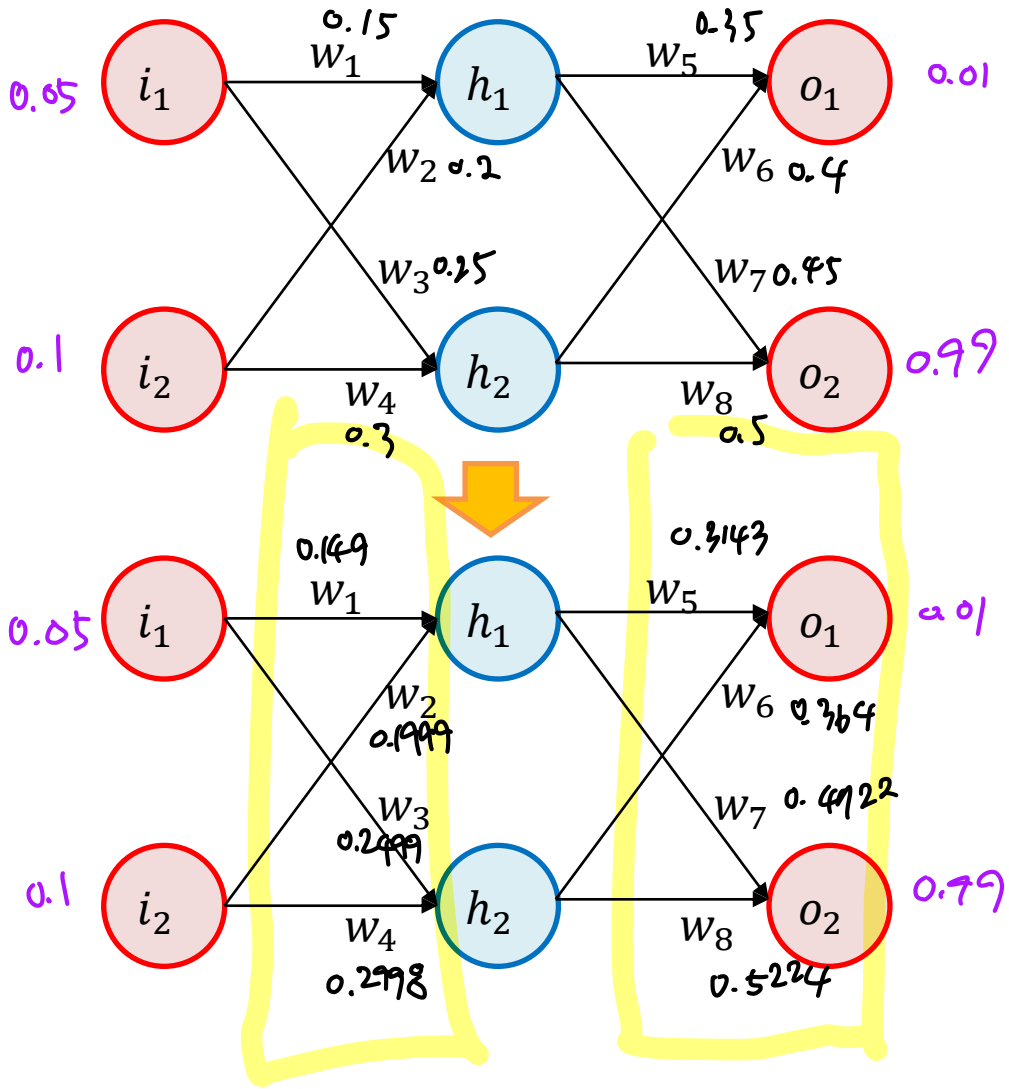
$$w_2, w_3, w_4$$

Do it yourself !

After All weights are Updated,

0.2371

0.2326



Etotal = 0.2326

Thank you..!!!

Reference : https://bskyvision.com/718?category=635506