Multi-Layered Perceptron

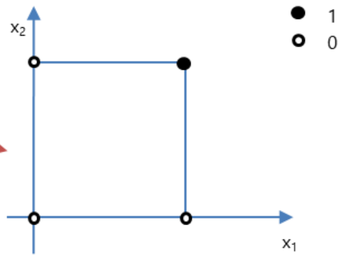
Hyerim Bae
Department of Industrial Engineering, Pusan National University
hrbae@pusan.ac.kr

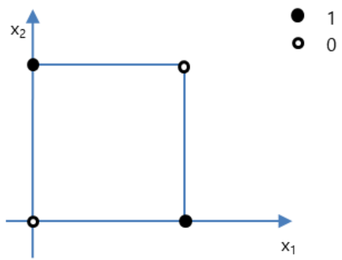
Contents

Perceptron의 한계

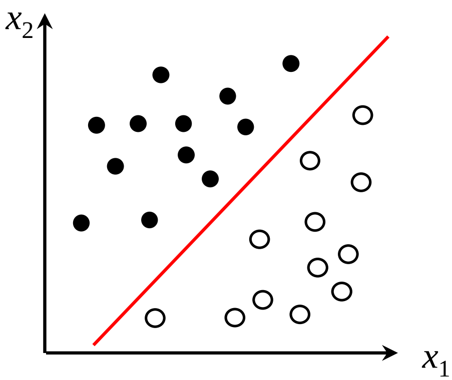
- Linearly Separable 문제만 풀 수 있다.

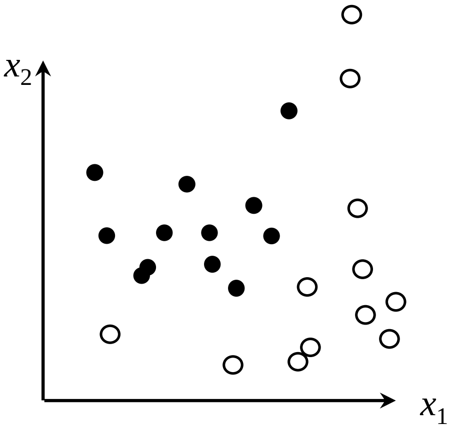
input		Output (by f)		
X_0	X_1	AND	OR	XOR
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0





Linearly separable

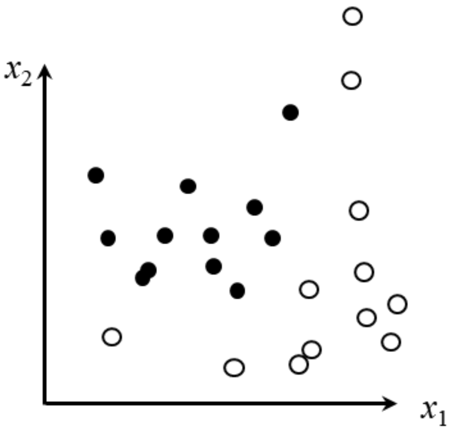




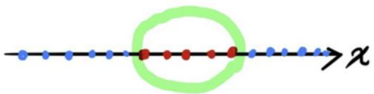
Overcome the linearity

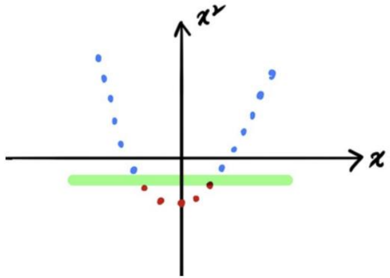
- In order to solve more complex problem
  - Using multiple linear classifier
  - Using non-linear classifier
  - Transforming into higher dimension

Using non-linear function

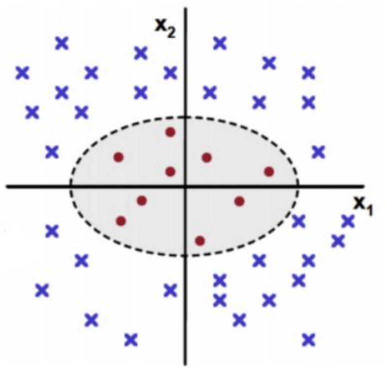


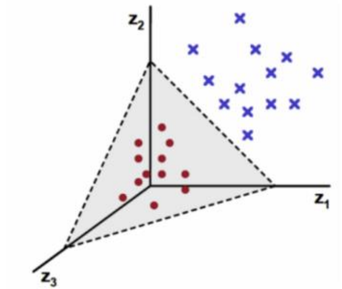
Transforming data





$$x \rightarrow \{x, x^2\}$$

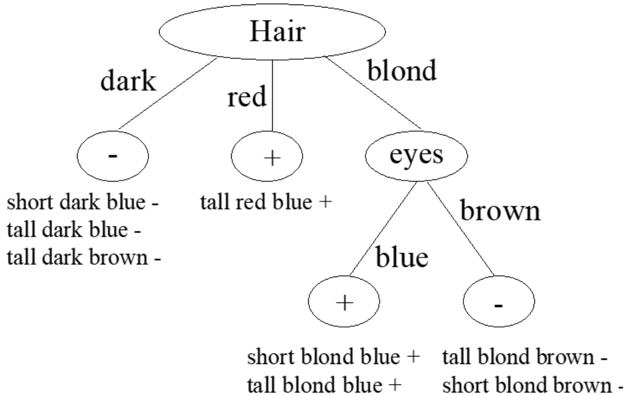


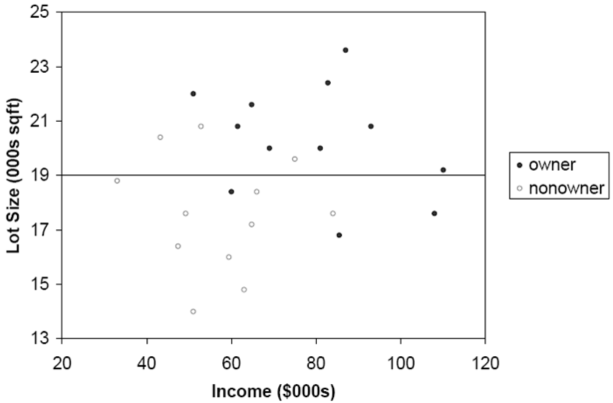


$$x = \{x_1, x_2\} \rightarrow z = \{x_1^2, \sqrt{2}x_1x_2, x_2^2\}$$

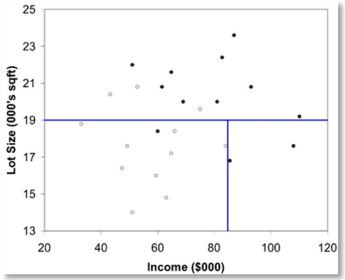
Using multiple linear classifier

	Height	Hair	Eyes	Class
1	short	blond	blue	+
2	tall	blond	brown	-
3	tall	red	blue	+
4	short	dark	blue	-
5	tall	dark	blue	-
6	tall	blond	blue	+
7	tall	dark	brown	-
8	short	blond	brown	-

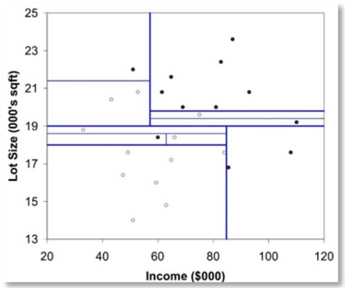




Lot Size (000s sqft)
Income ($000s)
owner
nonowner



Lot Size (000's sqft)
Income ($000)



Lot Size (000's sqft)
Income ($000)

How to solve XOR problem?

- XOR implementation

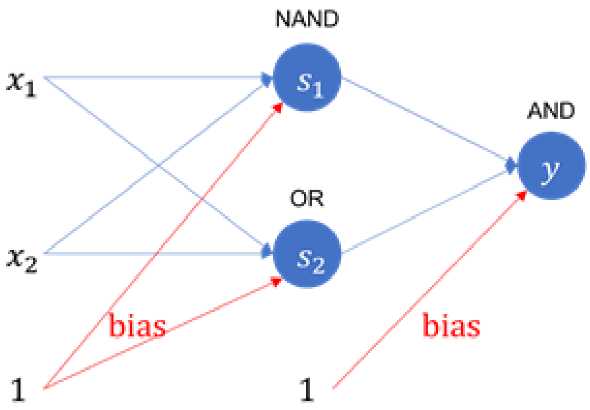
x_1	x_2	s_1	s_2	y
0	0	1	0	0
0	1	1	1	1
1	0	1	1	1
1	1	0	1	0

```
def XORGate(x1, x2):
    s1 = NANDGate(x1, x2)
    s2 = ORGate(x1, x2)
    y = ANDGate(s1, s2)
    return y
```

[1.1.7] XOR Gate

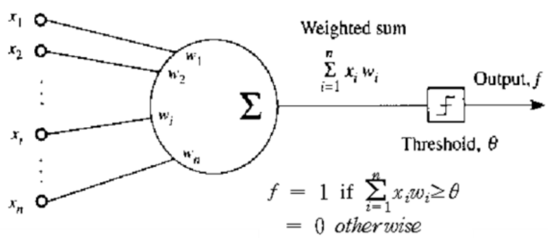
```
XORGate(0,0)
>> 0
XORGate(0,1)
>> 1
XORGate(1,0)
>> 1
XORGate(1,1)
>> 0
```

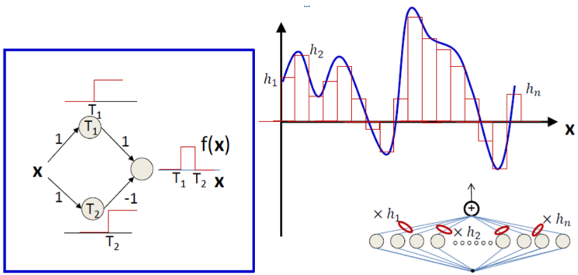
[1.1.8] XOR Gate 결과



Universal Approximation Theorem

- Activation functions
  - Sign

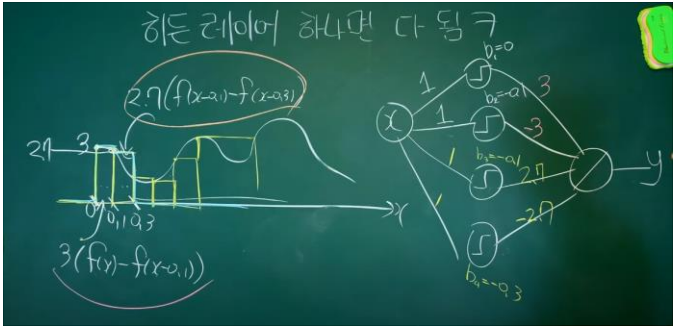




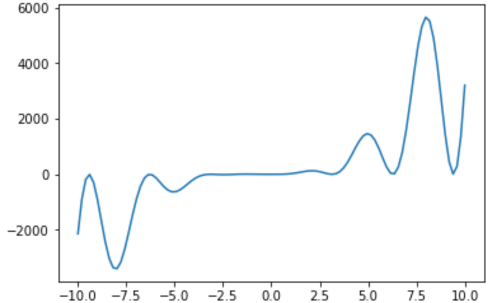
  - Sigmoid

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

    - Use $f(\boldsymbol{M}x)$



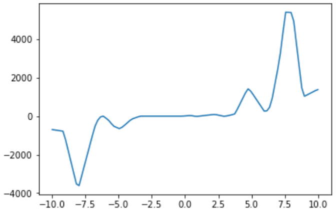
https://www.youtube.com/watch?v=vnkGn4r62Q8&list=PL_iJu012NOxdDZEygsVG4jS8srnSdIgdn&index=22



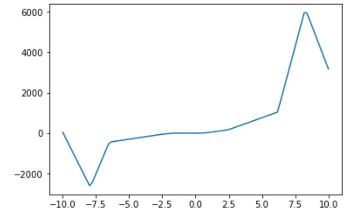
$$y = 3\sin(x)\cos(x)\,(6x^2 + 3x^3 + x)\tan(x)$$

MLP

Single hidden layer perceptron



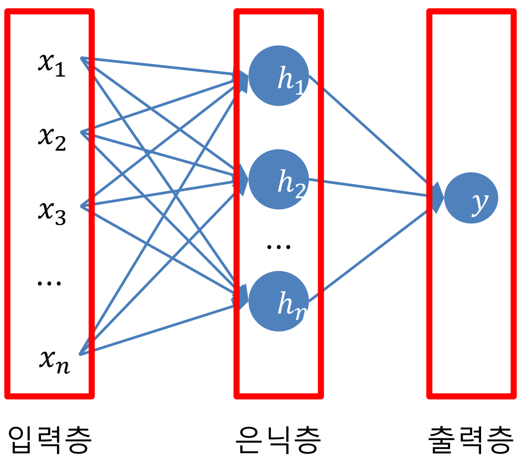
4-hidden layers with 10 perceptrons

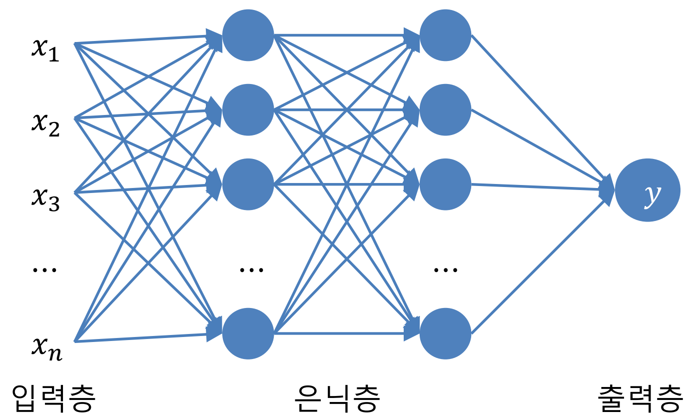


1-hidden layers with 1000 perceptrons

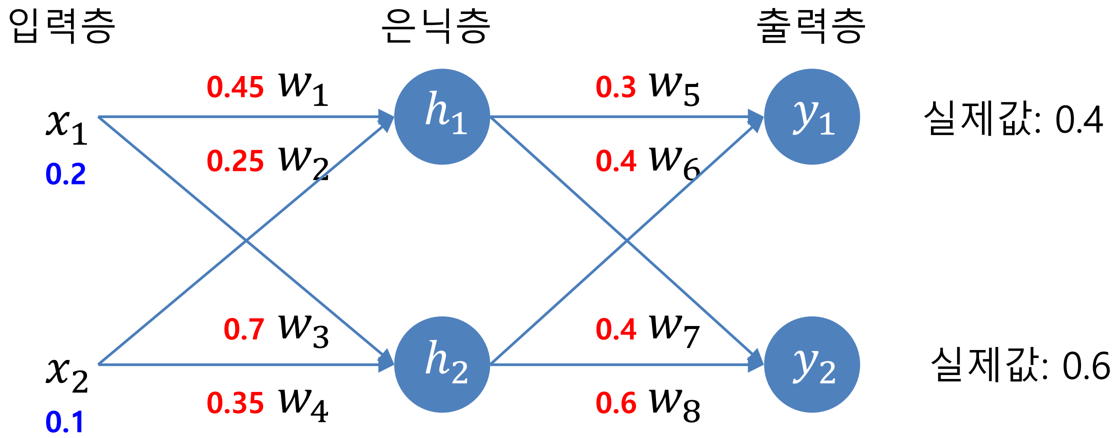
출처: https://3months.tistory.com/140

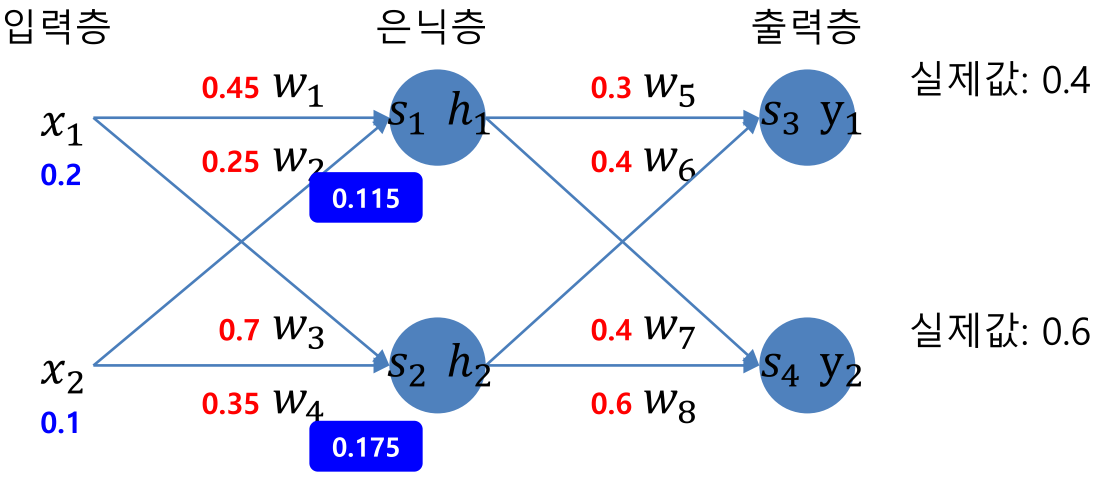
MLP





Feed forward



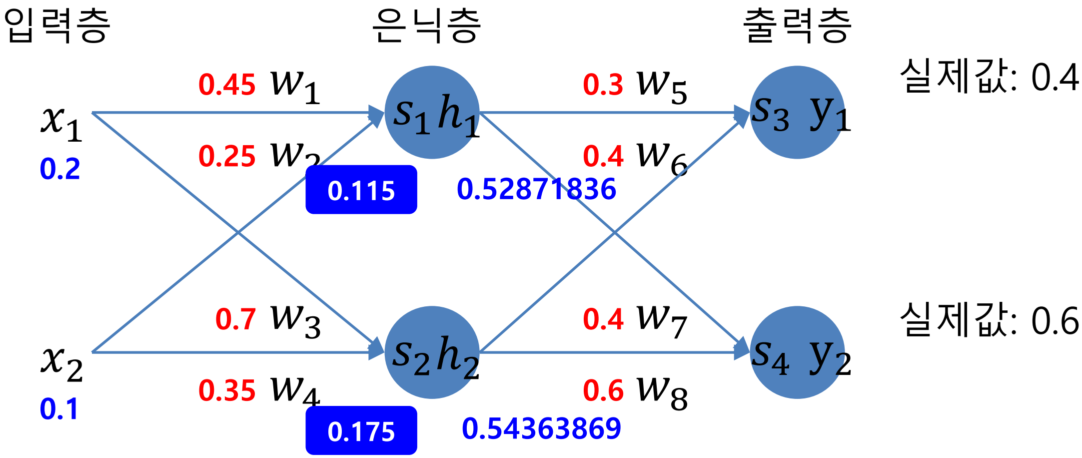


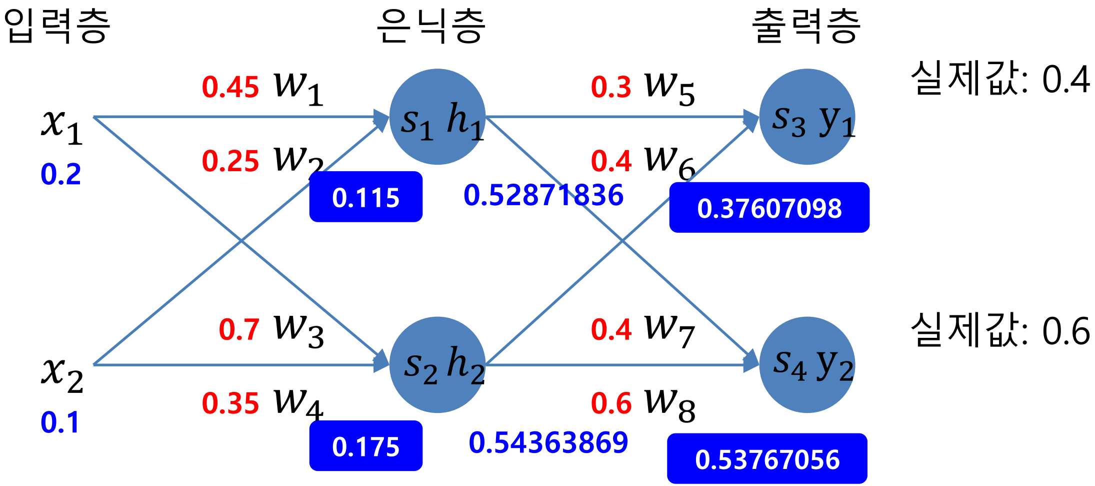
$$s_1 = w_1x_1 + w_2x_2 = 0.45 \times 0.2 + 0.25 \times 0.1 = 0.115$$

$$s_2 = w_3x_1 + w_4x_2 = 0.7 \times 0.2 + 0.35 \times 0.1 = 0.175$$

$$h_1 = sigmoid(s_1) = sigmoid(0.115) = 0.52871836$$

$$h_2 = sigmoid(s_2) = sigmoid(0.175) = 0.54363869$$

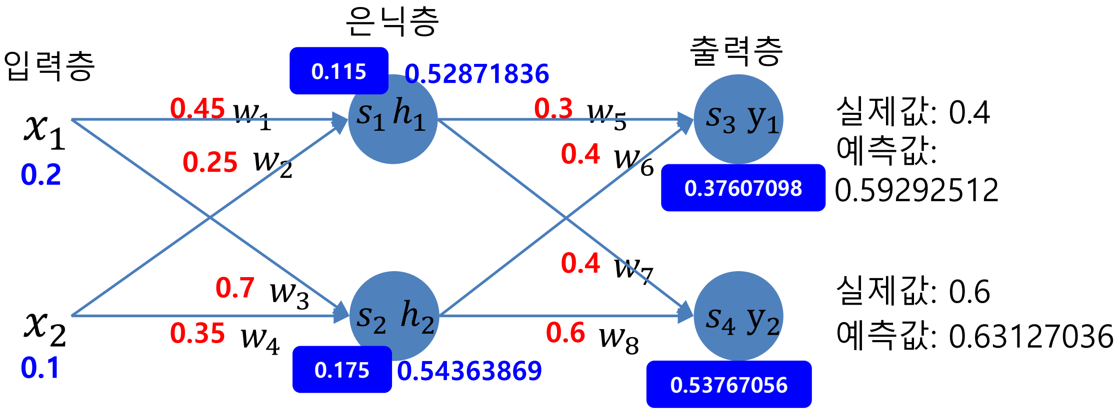




입력층
은닉층
출력층
실제값: 0.4
실제값: 0.6
0.45 w_1
0.25 w_2
0.7 w_3
0.35 w_4
0.3 w_5
0.4 w_6
0.4 w_7
0.6 w_8
x_1
0.2
x_2
0.1
s_1 h_1
s_2 h_2
s_3 y_1
s_4 y_2
0.115
0.175
0.52871836
0.54363869
0.37607098
0.53767056

$$y_1 = sigmoid(s_3) = sigmoid(0.37607098) = 0.59292512$$

$$y_2 = sigmoid(s_4) = sigmoid(0.53767056) = 0.63127036$$



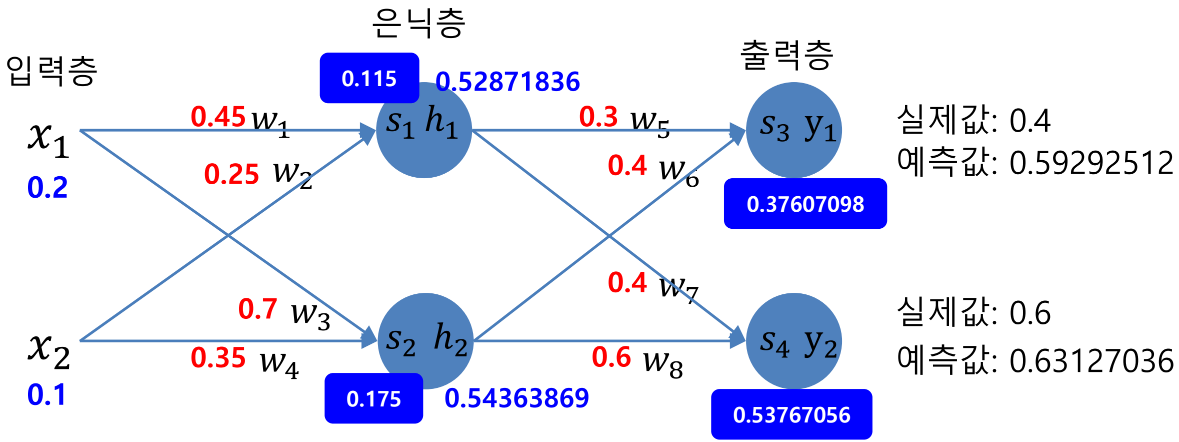
$$E_{y_1} = \frac{1}{2}(target_{y_1} - output_{y_1})^2 = 0.01861005$$

$$E_{y_2} = \frac{1}{2}(target_{y_2} - output_{y_2})^2 = 0.00048892$$

$$E_{total} = E_{y_1} + E_{y_2} = 0.01861005 + 0.00048892 = 0.01909897$$

Backpropagation

- Update w5, w6, w7, w8

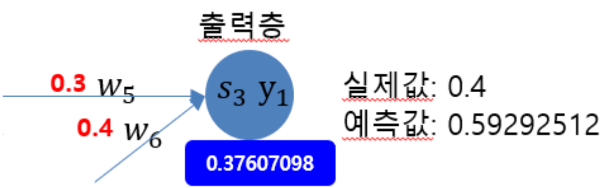


- 우리가 목적으로 하는 것은 Loss(MSE, E_{total})를 최소화
- 우리가 찾고자 하는 것은 파라미터(w)

- E를 최소화하는 w5를 구하고 싶다면?

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

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial s_3}\frac{\partial s_3}{\partial w_5}$$



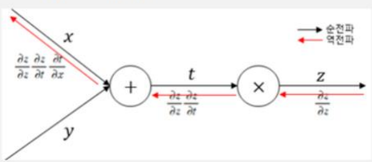
※ [보충설명] 연쇄법칙 (chain rule)

연쇄법칙 : 합성 함수의 미분에 대한 성질로 여러 함수로 구성된 합성 함수의 미분은 합성 함수를 구성하는 각 함수의 미분의 곱으로 나타낼 수 있는 성질을 말한다.

$$z = t^2$$
$$t = x + y$$
$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x}$$

이를 활용하여 미분 $\frac{\partial z}{\partial x}$을 구하면 다음과 같다.

$$\frac{\partial z}{\partial t} = 2t$$
$$\frac{\partial t}{\partial x} = 1$$
$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x} = 2t \times 1 = 2(x+y)$$



$$\frac{\partial E}{\partial w_5} = \boxed{\frac{\partial E}{\partial y_1}} \frac{\partial y_1}{\partial s_3} \frac{\partial s_3}{\partial w_5}$$

$$-\left(target_{y_1} - output_{y_1}\right) = -(0.4 - 0.59292512) = 0.19292512$$

$$\frac{\partial y_1}{\partial s_3} = y_1 \times (1 - y_1) = 0.59292512 \times (1 - 0.59292512) = 0.24136492$$

$$\frac{\partial s_3}{\partial w_5} = h_1 = 0.52871836$$

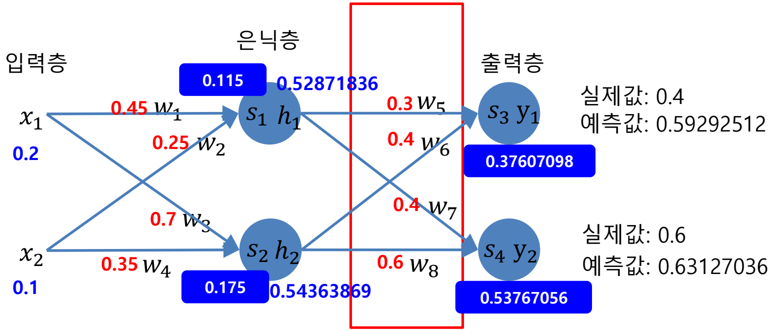
$$\frac{\partial E}{\partial w_5} = 0.19292512 \times 0.24136492 \times 0.52871836 = 0.02461996$$

$$w_5 = w_5 - \alpha \frac{\partial E}{\partial w_5} = 0.3 - 0.5 \times 0.02461996 = 0.28769002$$

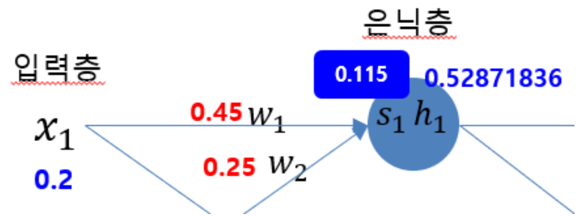
$$\frac{\partial E}{\partial w_6} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial s_3}\frac{\partial s_3}{\partial w_6}$$

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial s_4}\frac{\partial s_4}{\partial w_7}$$

$$\frac{\partial E}{\partial w_8} = \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial s_4}\frac{\partial s_4}{\partial w_8}$$



- Now, update w1
- $$\frac{\partial E}{\partial w_1} = \boxed{\frac{\partial E}{\partial h_1}} \frac{\partial h_1}{\partial s_1} \frac{\partial s_1}{\partial w_1}$$
- $$\frac{\partial E}{\partial h_1} = \frac{\partial E_{y_1}}{\partial h_1} + \frac{\partial E_{y_2}}{\partial h_1}$$



$$\frac{\partial E_{y_1}}{\partial h_1} = \frac{\partial E_{y_1}}{\partial s_3} \frac{\partial s_3}{\partial h_1} = \frac{\partial E_{y_1}}{\partial y_1} \frac{\partial y_1}{\partial s_3} \frac{\partial s_3}{\partial h_1}$$
$$= -(\mathrm{target}_{y_1} - \mathrm{output}_{y_1}) \times y_1 \times (1 - y_1) \times w_5$$
$$= -(0.4 - 0.59292512) \times 0.4 \times (1 - 0.4) \times (0.3) = 0.01389061$$

$$\frac{\partial E_{y_2}}{\partial h_1} = \frac{\partial E_{y_2}}{\partial s_4} \frac{\partial s_4}{\partial h_1} = \frac{\partial E_{y_2}}{\partial y_2} \frac{\partial y_2}{\partial s_4} \frac{\partial s_4}{\partial h_1} = 0.00300195$$

$$\frac{\partial E}{\partial h_1} = \frac{\partial E_{y_1}}{\partial h_1} + \frac{\partial E_{y_2}}{\partial h_1} = 0.01689256$$

$$\frac{\partial h_1}{\partial s_1} = h_1 \times (1 - h_1) = 0.24917526$$

$$\frac{\partial s_1}{\partial w_1} = x_1 = 0.2$$

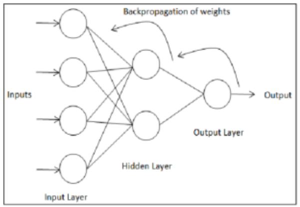
$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial h_1}\frac{\partial h_1}{\partial s_1}\frac{\partial s_1}{\partial w_1} = 0.01689256 \times 0.24917526 \times 0.2 = 0.00084184$$

w_1 *update*

$$w_1' = w_1 - \alpha \frac{\partial E}{\partial w_1} = 0.45 - 0.5 \times 0.00084184 = 0.44957908$$

Vanishing gradient

- Backpropagation is based on differentiation of activation function



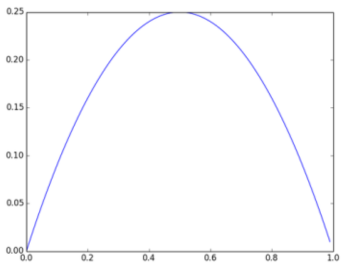
- If we use sigmoid or tanh as activation function

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial z_3}\frac{\partial z_3}{\partial w_5}$$

$$w_5 = w_5 - \alpha\frac{\partial E}{\partial w_5} = 0.3 - 0.5 \times 0.9630084 = -0.1815042$$

Sigmoid derivative(미분계수):

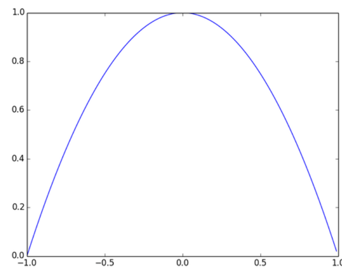
$$\sigma(x) = \frac{1}{1+e^{-x}}$$



tanh derivative(미분계수):

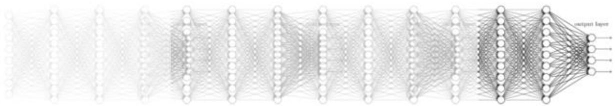
$$\tanh(x) = \frac{1-e^{-x}}{1+e^{-x}}$$

$$= 2\sigma(2x) - 1$$

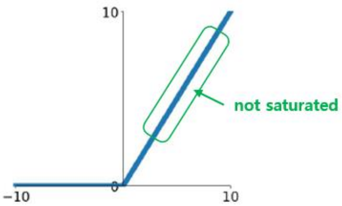


Solution

Vanishing gradient (NN winter2: 1986-2006)



- Using ReLu



$$ReLu(x) = \max(0, x)$$