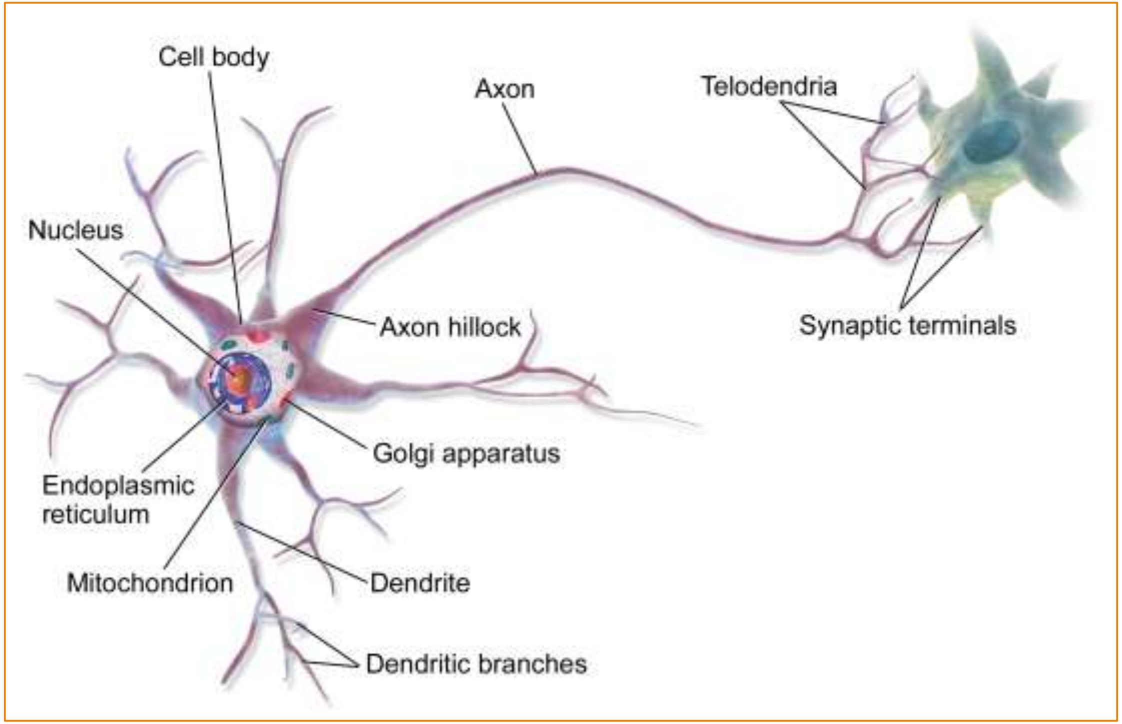
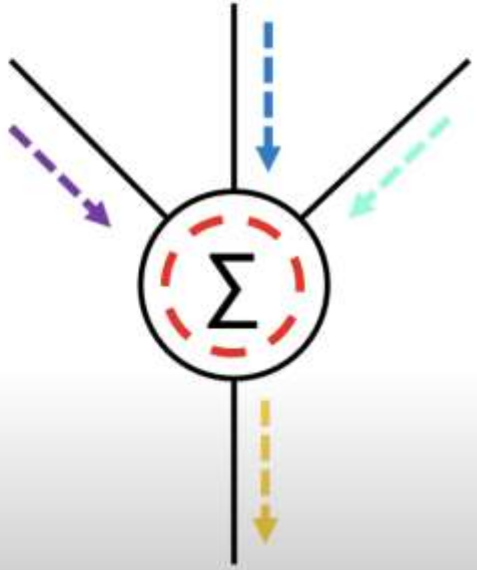
Simple Neural Networks and Neural Language Models

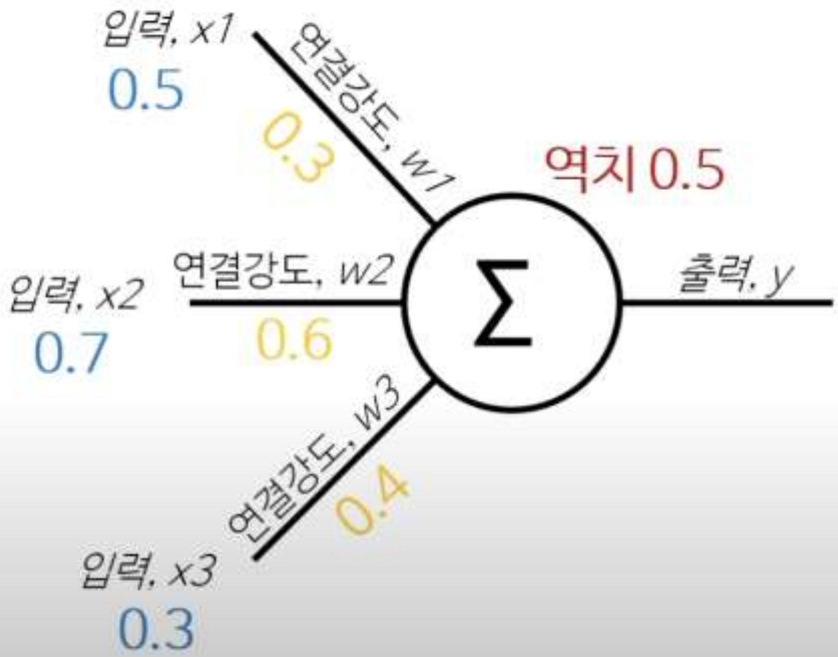
Units in Neural Networks

This is in your brain

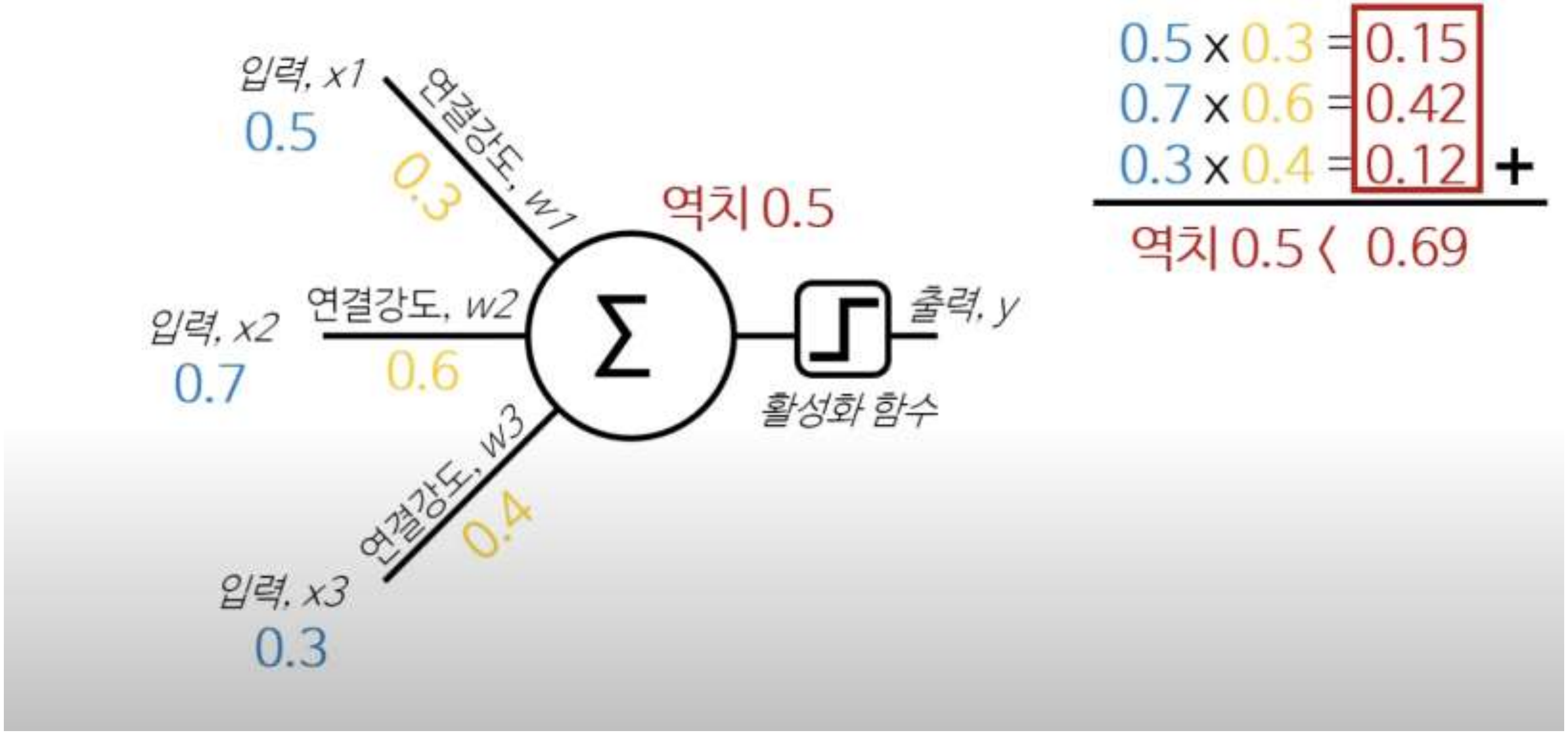


This is in your brain

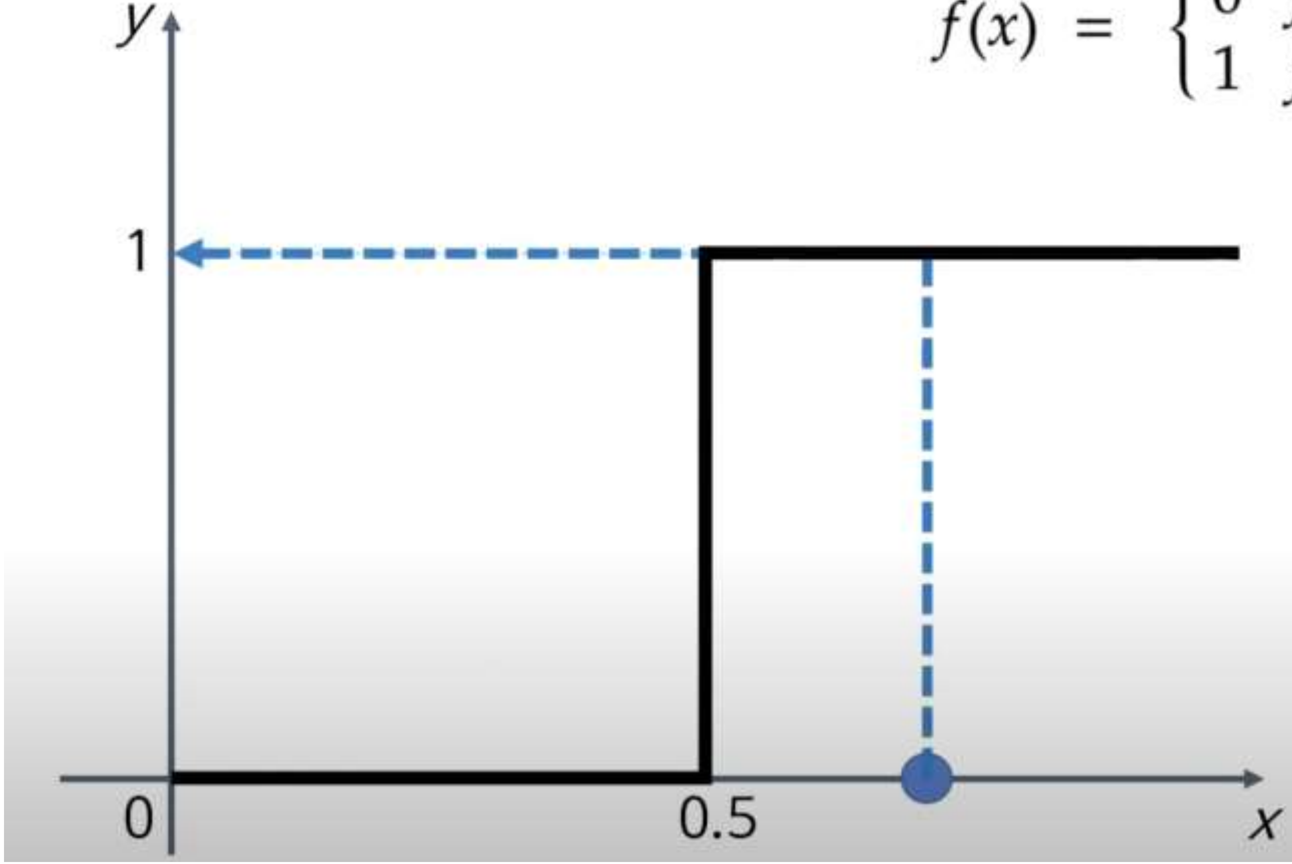


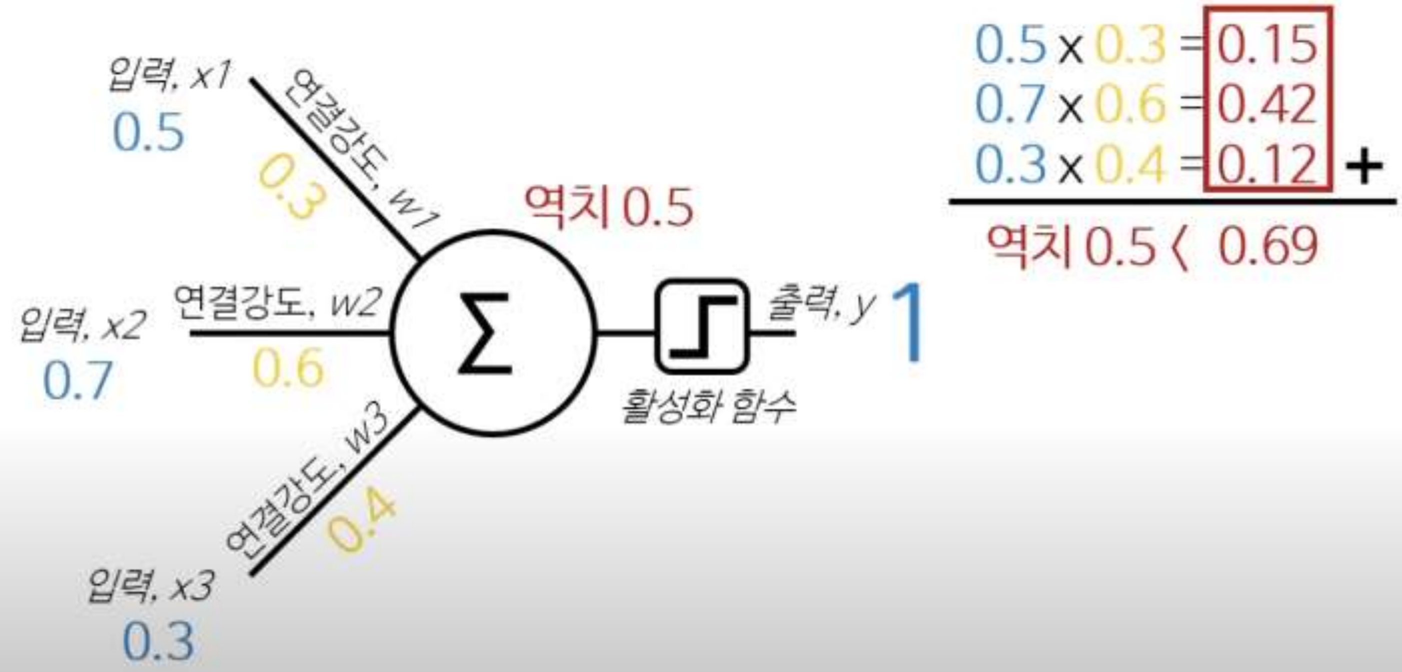


$$
\begin{array}{r}
0.5 \times 0.3 = \boxed{0.15} \\
0.7 \times 0.6 = \boxed{0.42} \\
0.3 \times 0.4 = \boxed{0.12} \; + \\
\hline
\text{역치 } 0.5 < 0.69
\end{array}
$$


입력, x1
0.5
연결강도, w1
0.3
입력, x2
0.7
연결강도, w2
0.6
입력, x3
0.3
연결강도, w3
0.4
Σ
역치 0.5
출력, y
활성화 함수
0.5 x 0.3 = 0.15
0.7 x 0.6 = 0.42
0.3 x 0.4 = 0.12 +
역치 0.5 〈 0.69

$$f(x) = \begin{cases} 0 & for\ x < 0.5 \\ 1 & for\ x \geqslant 0.5 \end{cases}$$





입력, x1
0.5
연결강도, w1
0.3
입력, x2
0.7
연결강도, w2
0.6
입력, x3
0.3
연결강도, w3
0.4
역치 0.5
Σ
출력, y
1
활성화 함수
0.5 x 0.3 = 0.15
0.7 x 0.6 = 0.42
0.3 x 0.4 = 0.12
+
역치 0.5 〈 0.69

$$y_k = \varphi\left(\sum_{j=0}^{m} w_{kj}x_j\right)$$

$$y_k = \varphi\left(\sum_{j=0}^{m} \boxed{w_{kj}x_j}\right)$$

0.5 x 0.3 = 0.15
0.7 x 0.6 = 0.42
0.3 x 0.4 = 0.12

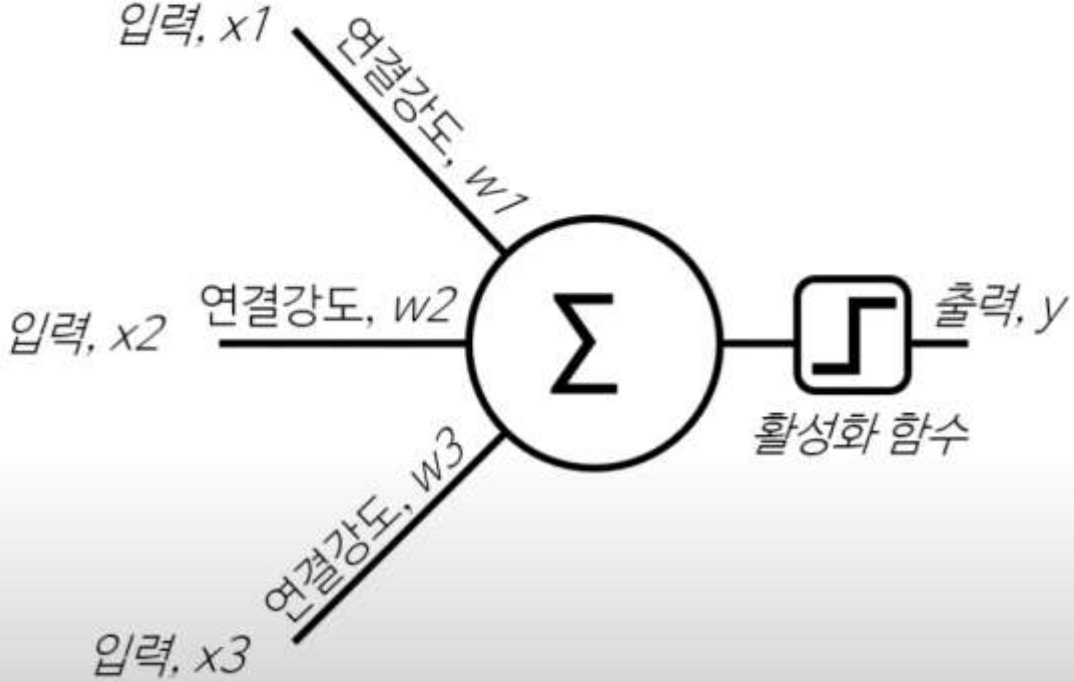
$$y_k = \varphi\left(\boxed{\sum_{j=0}^{m}} w_{kj} x_j\right)$$

$$\begin{array}{r} 0.5 \times 0.3 = \boxed{0.15} \\ 0.7 \times 0.6 = \boxed{0.42} \\ 0.3 \times 0.4 = \boxed{0.12} \; + \\ \hline 0.69 \end{array}$$

$$y_k = \boxed{\varphi}\left(\sum_{j=0}^{m} w_{kj}x_j\right)$$

$$\begin{array}{r} 0.5 \times 0.3 = 0.15 \\ 0.7 \times 0.6 = 0.42 \\ 0.3 \times 0.4 = 0.12 \; + \\ \hline f(0.69) \end{array}$$

$$f(x) = \begin{cases} 0 & for\ x < 0.5 \\ 1 & for\ x \geqslant 0.5 \end{cases}$$


입력, x1
연결강도, w1
입력, x2
연결강도, w2
입력, x3
연결강도, w3
Σ
출력, y
활성화 함수

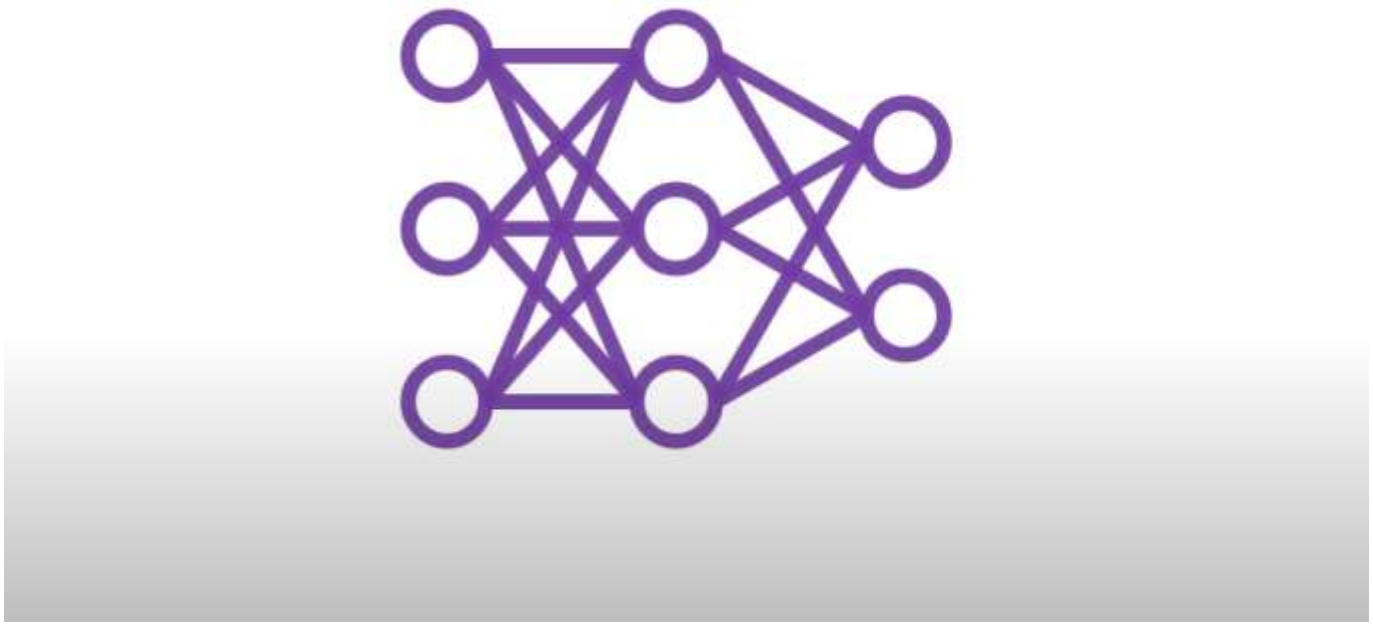
퍼셉트론 (Perceptron)이란?
:1943년 신경생리학자인 McCulloch 와 계산신경과학자인 Pitts가 제안한 McCulloch-Pitts Neuron을 바탕으로 미국의 심리학자인 Rosenblatt이 1958년에 구현해낸 인공신경망

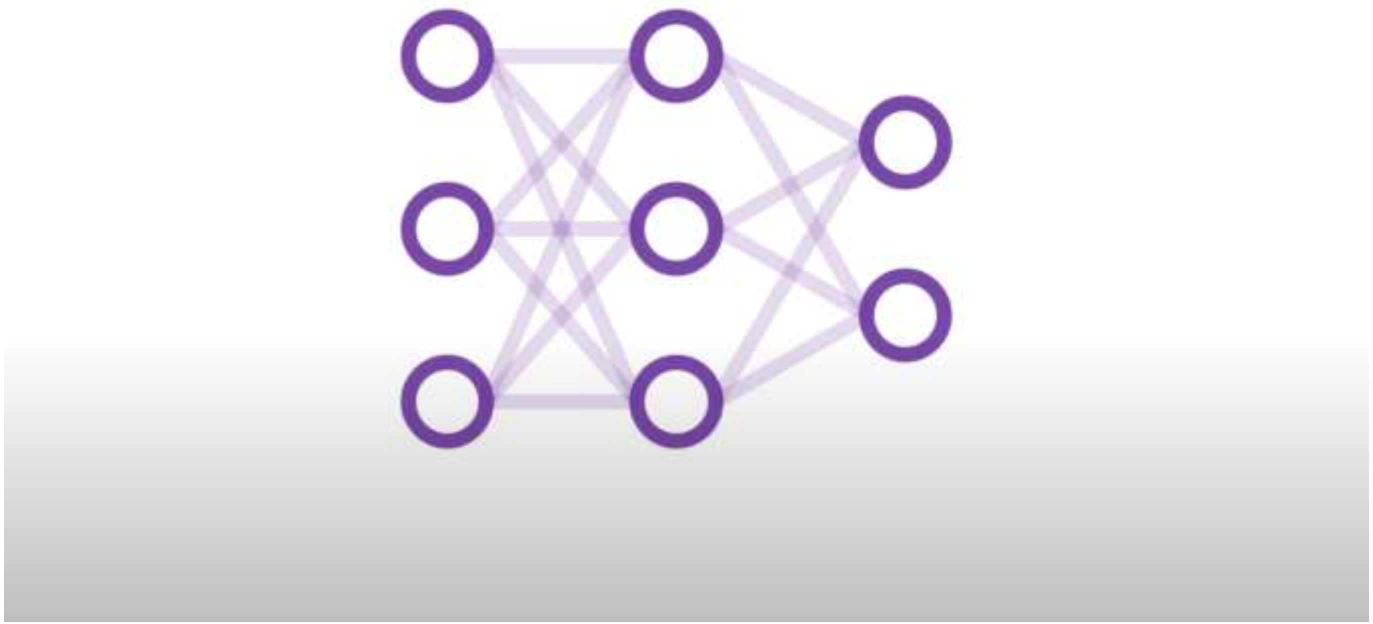
McCulloch (right) and Pitts (left)

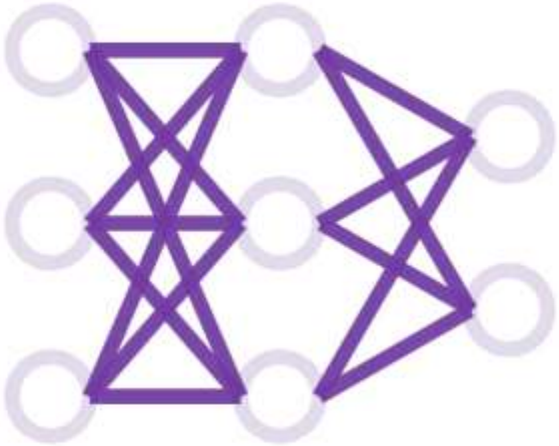
https://www.historyofinformation.com/detail.php?entryid=782

Frank Rosenblatt

https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon







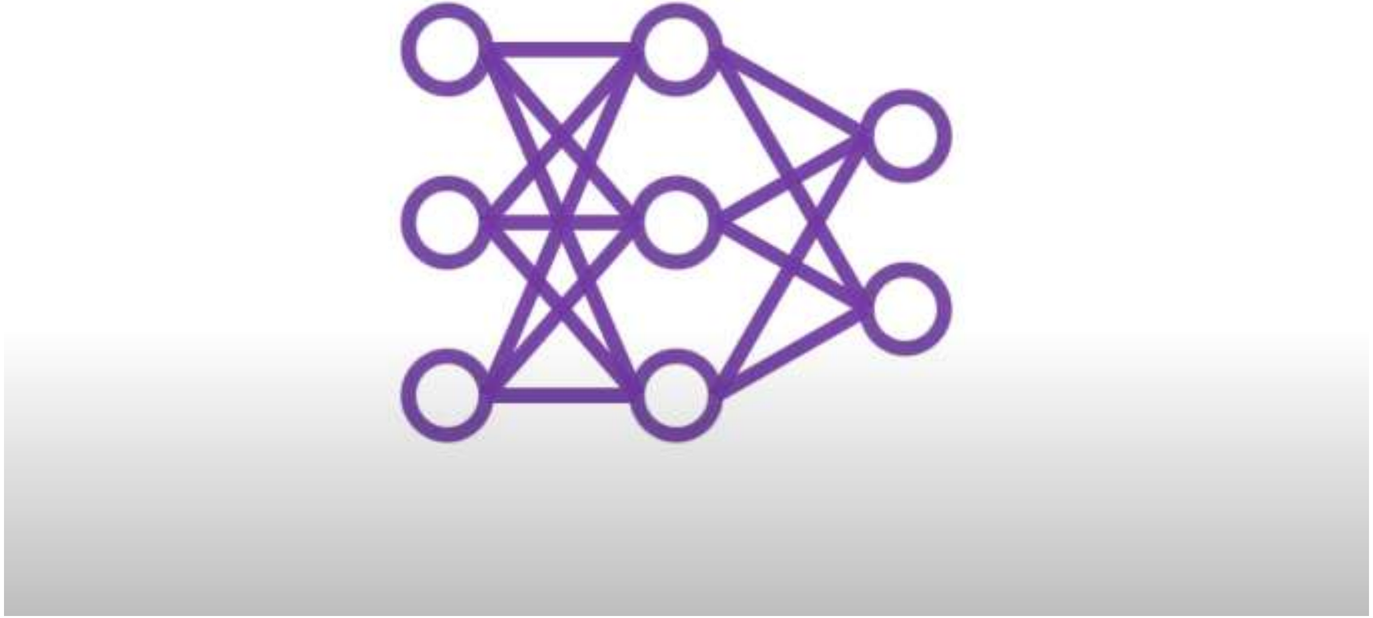


Photo by Pauline Loroy on Unsplash

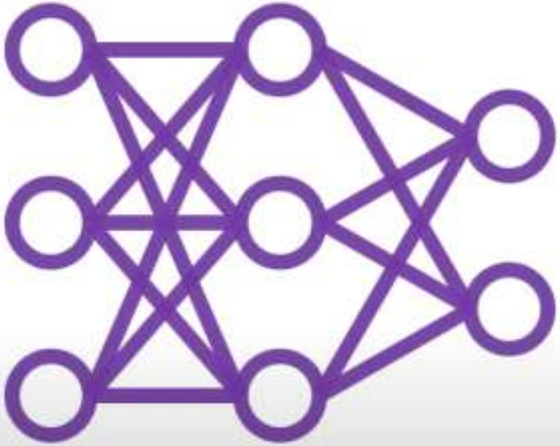


Photo by Pauline Loroy on Unsplash

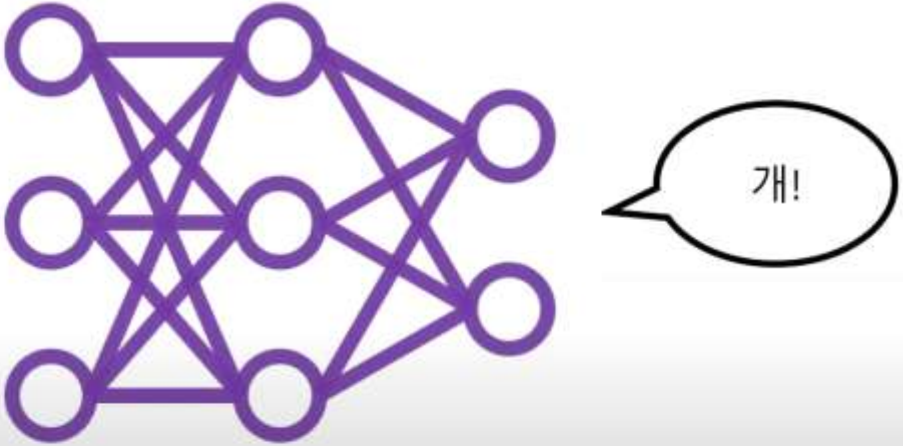
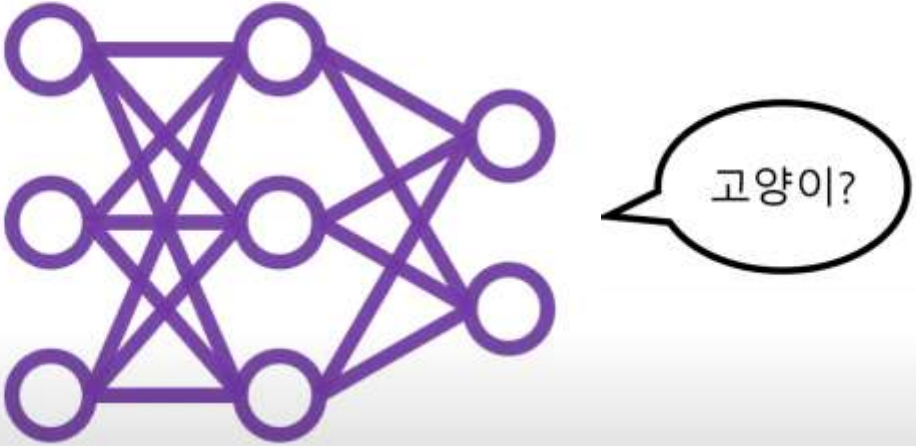
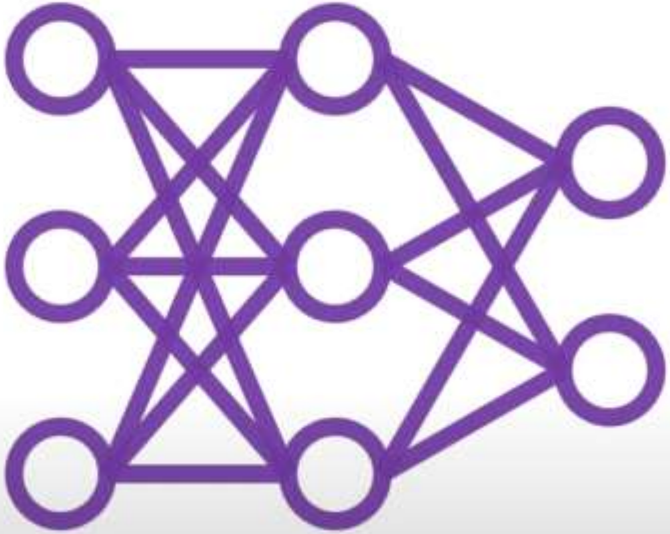
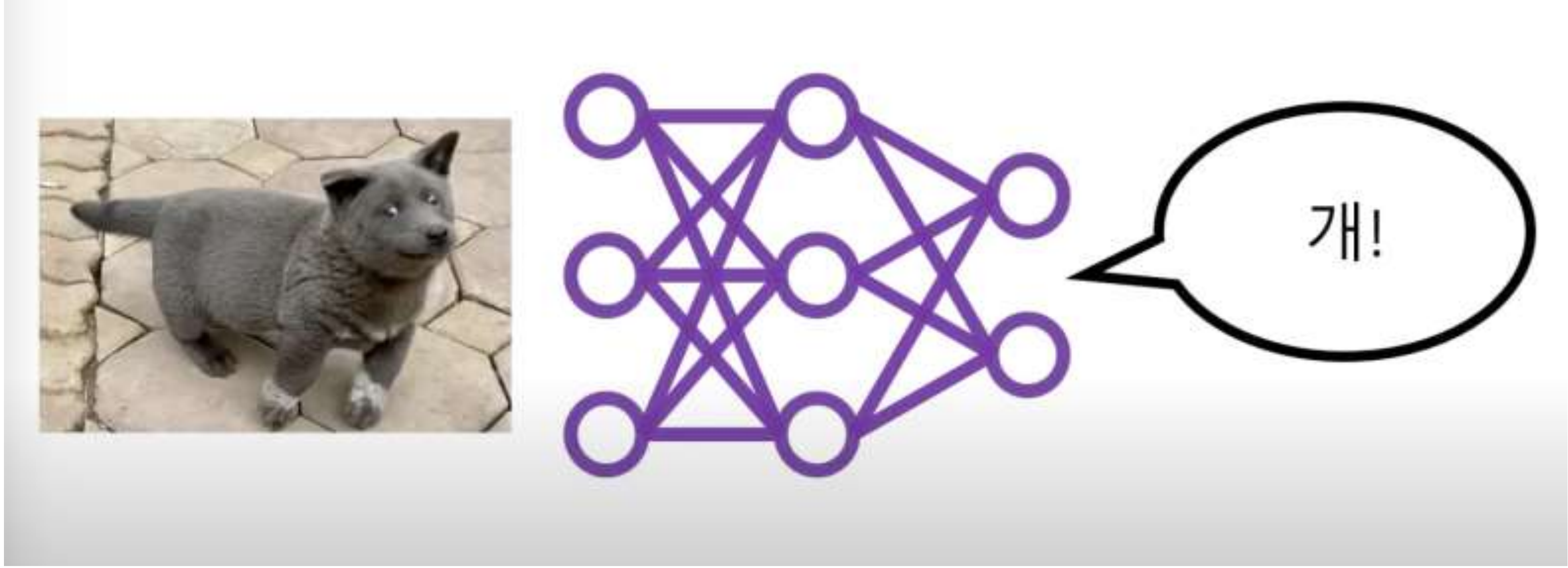
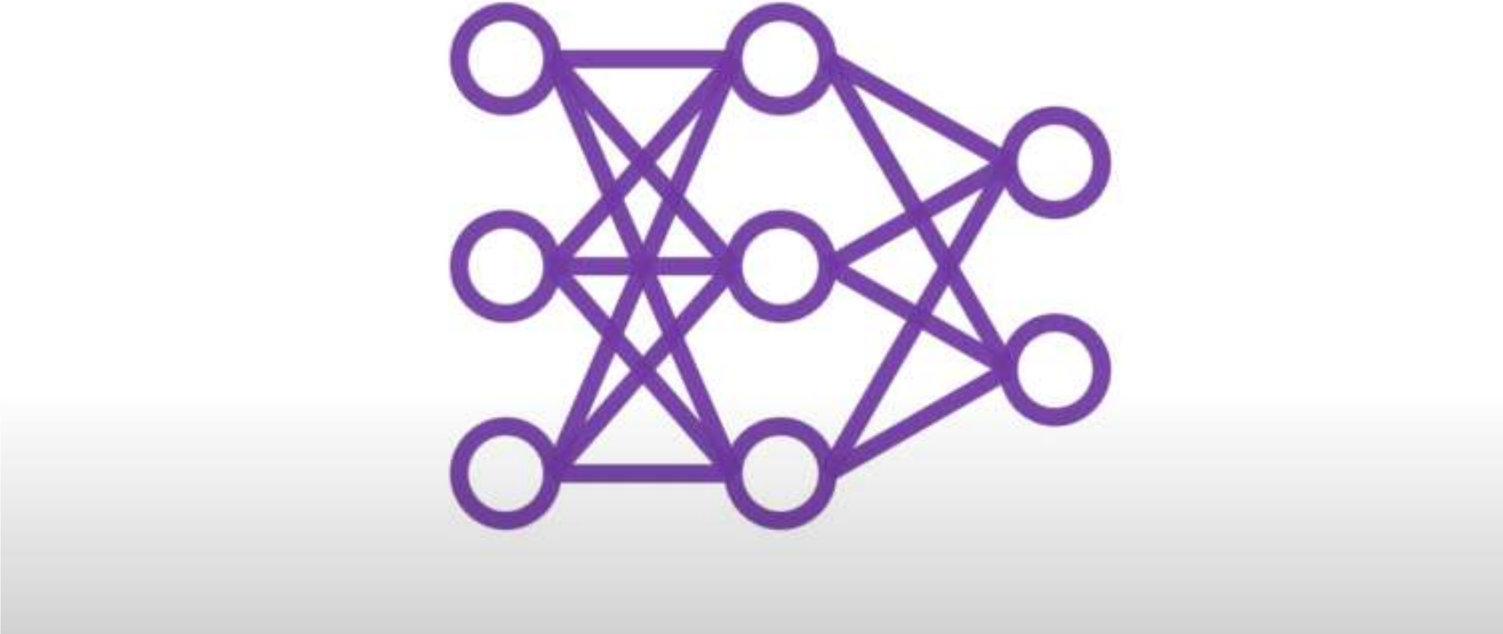


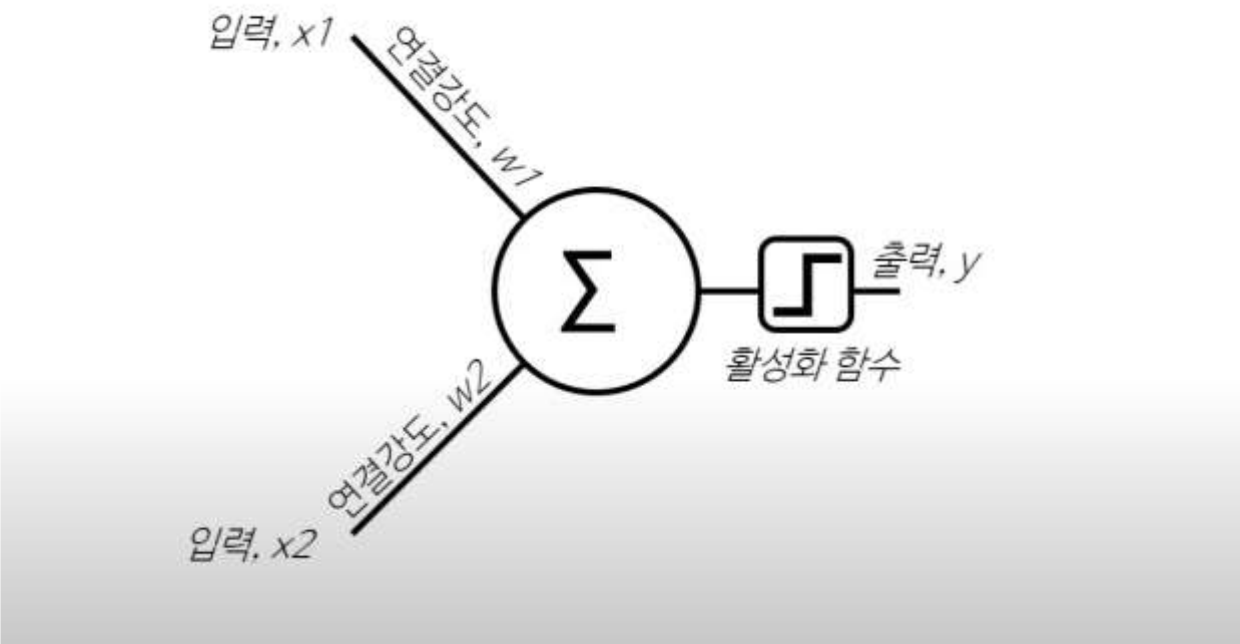
Photo by Pauline Loroy on Unsplash

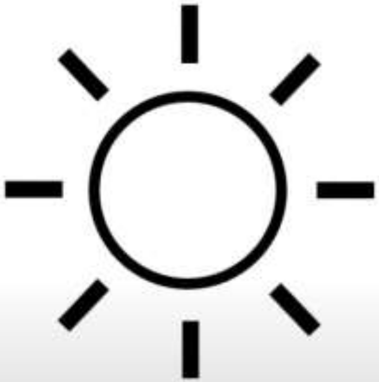





개!

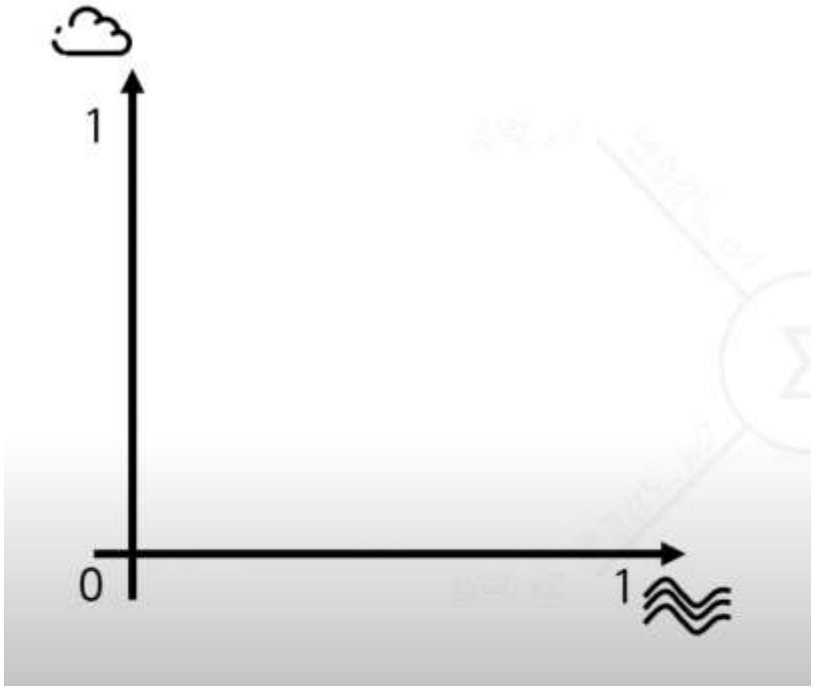



입력, x1
연결강도, w1
Σ
출력, y
활성화 함수
연결강도, w2
입력, x2

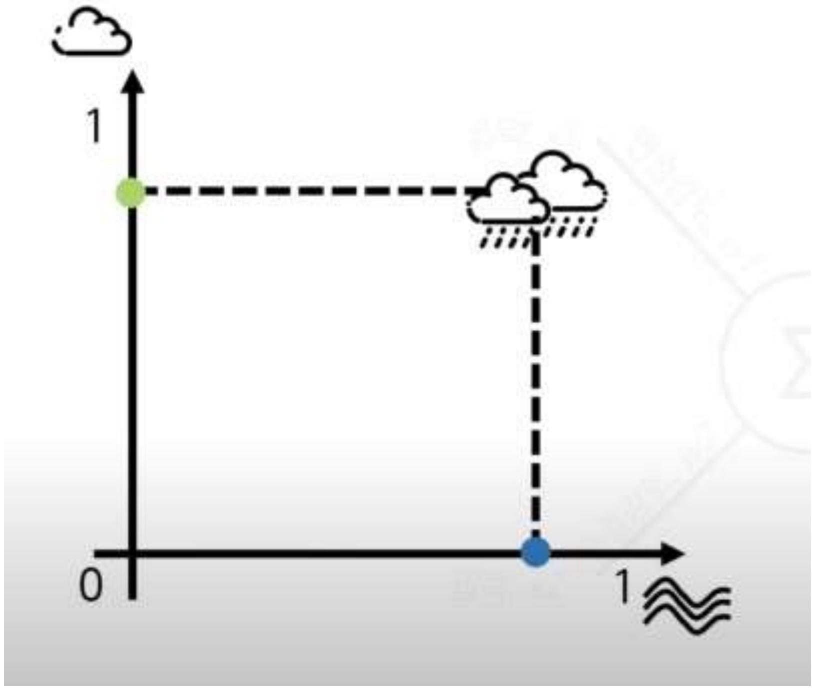


아침에 **구름**이 없고 **바람**이
약하면 그 날은 맑은 날이 될
확률이 높음

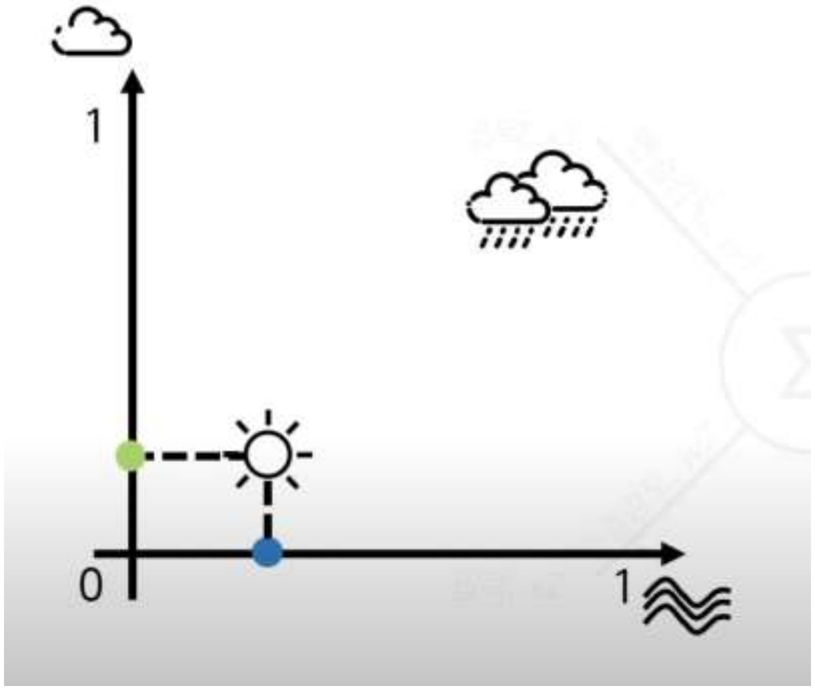
아침에 **구름**이 많고 **바람**이
강하면 그 날은 맑은 날이 될
확률이 높음

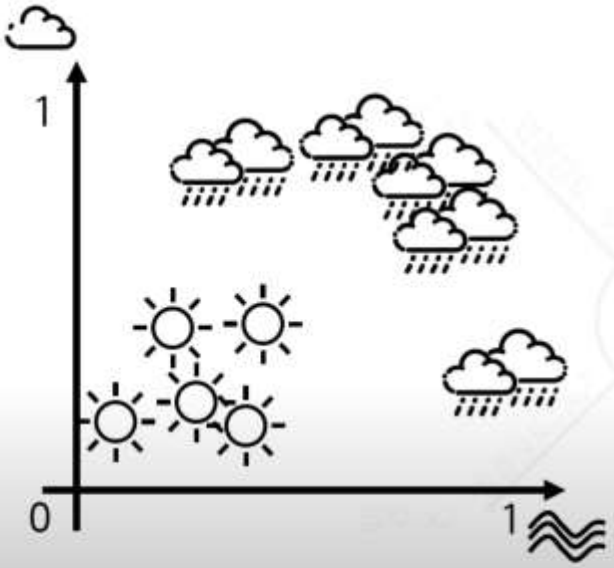


1
0
1

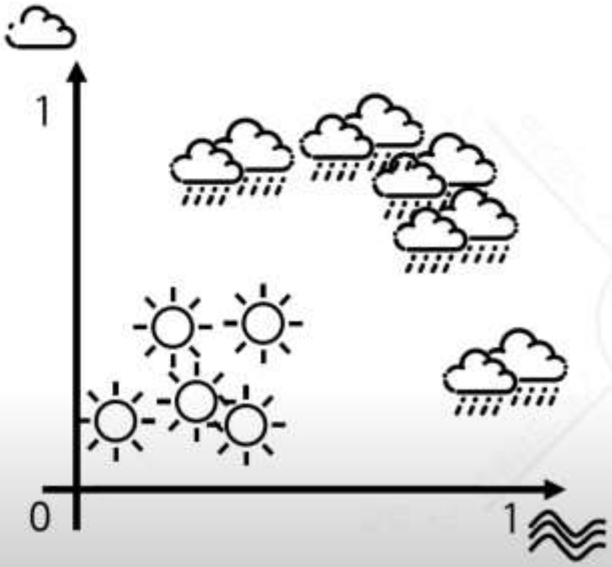


1
0
1


1
0
1



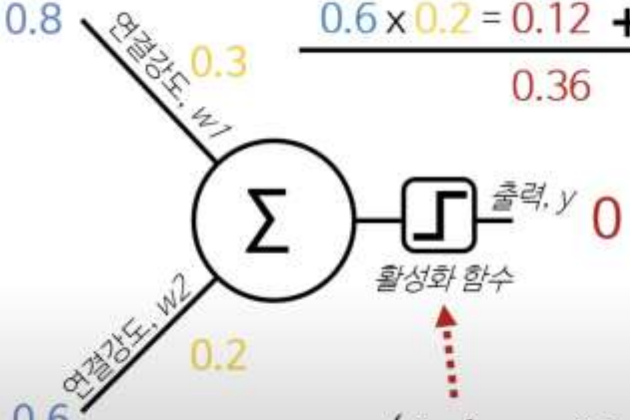
0.8	0.6	
0.6	0.9	
0.1	0.2	
0.3	0.1	
0.6	0.6	
0.4	0.3	
0.1	0.2	



x1	x2	y
0.8	0.6	1
0.6	0.9	1
0.1	0.2	0
0.3	0.1	0
0.6	0.6	1
0.4	0.3	0
0.1	0.2	0

x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

$0.8 \times 0.3 = 0.24$
$0.6 \times 0.2 = 0.12$ +
0.36



$$f(x) = \begin{cases} 0 & for\ x < 0.5 \\ 1 & for\ x \geqslant 0.5 \end{cases}$$

y
1
1
0
0
1
0
0

y'
0

퍼셉트론의 학습방법:

새 연결강도 = 현 연결강도 + 현 입력값 x 오차 x 학습률

y	y’
1	0
1	
0	
0	
1	
0	
0	

퍼셉트론의 학습방법:

새 연결강도 = 현 연결강도 + 현 입력값 x 오차 x 학습률

0

y	y’
1	0
1	
0	
0	
1	
0	
0	

퍼셉트론의 학습방법:

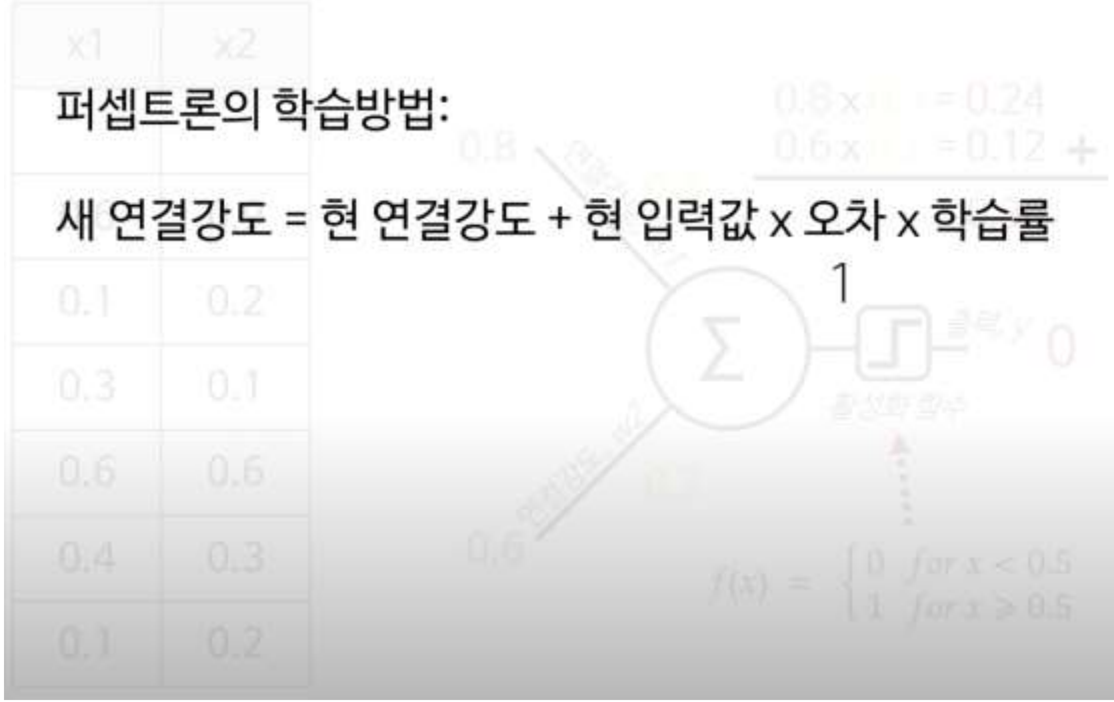
새 연결강도 = 현 연결강도 ~~+ 현 입력값 x 오차 x 학습률~~

0

y	y'
1	0
1	
0	
0	
1	
0	
0	

퍼셉트론의 학습방법:

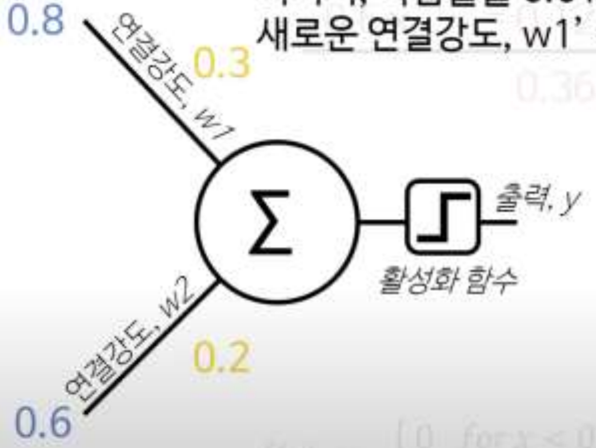
새 연결강도 = 현 연결강도 + 현 입력값 x 오차 x 학습률



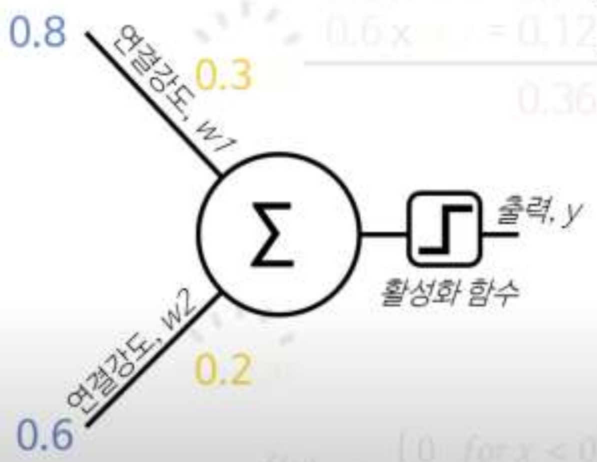
y	y’
1	0
1	
0	
0	
1	
0	
0	

x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

새로운 연결강도, w1’ = w1 + x1 x 오차 x 학습률
여기서, 학습률을 0.01로 가정한다면,
새로운 연결강도, w1’ = 0.3 + 0.8 x 1 x 0.01 = 0.308

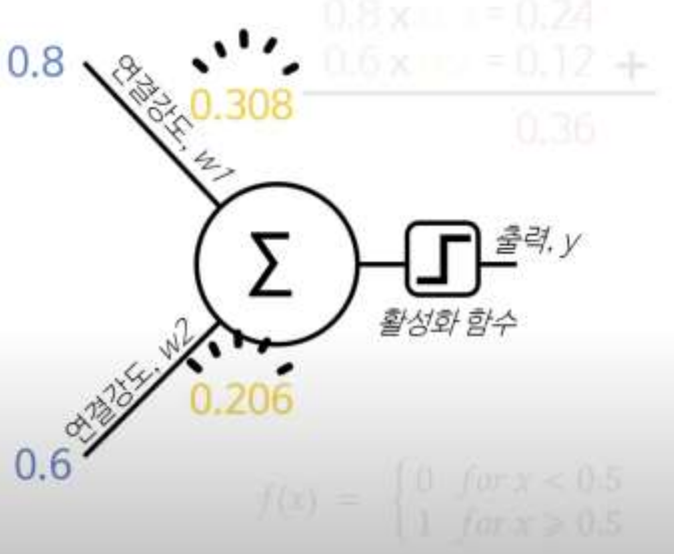


x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

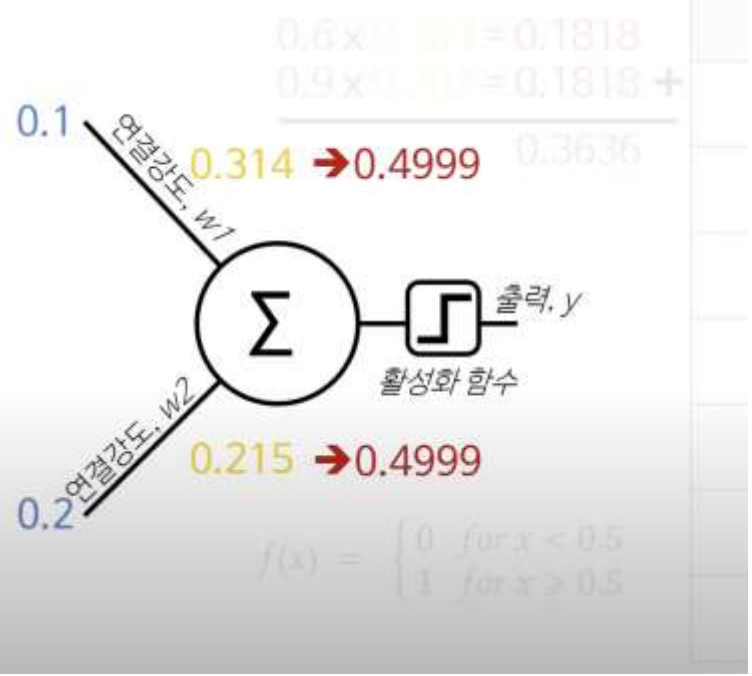


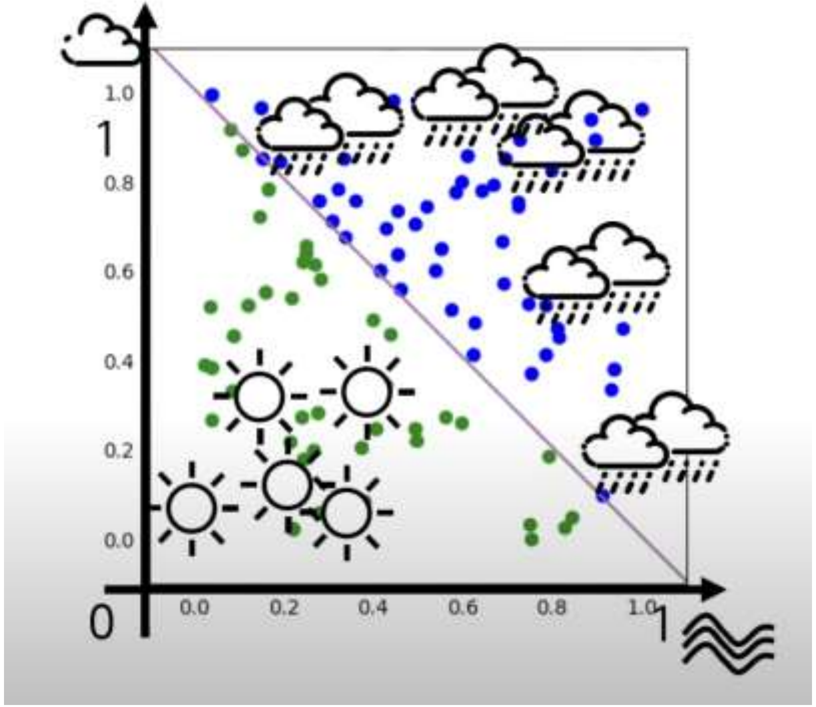
새로운 연결강도, w2' = w2 + x2 x 오차 x 학습률
= 0.2 + 0.6 x 1 x 0.01
= 0.206

x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

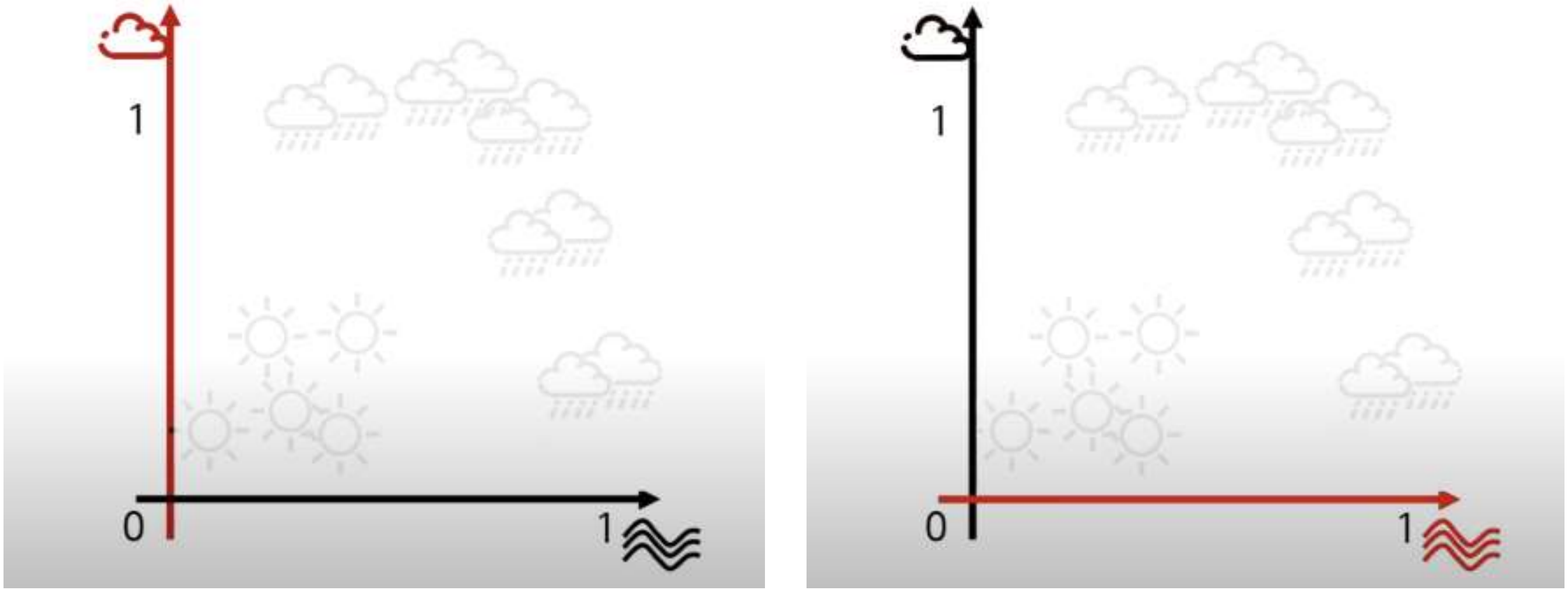


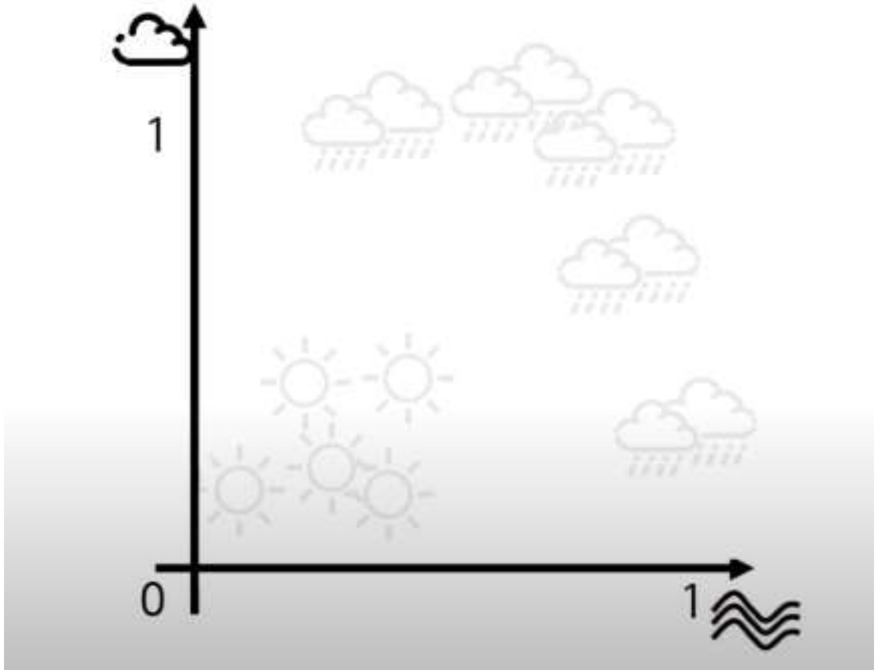
x1	x2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

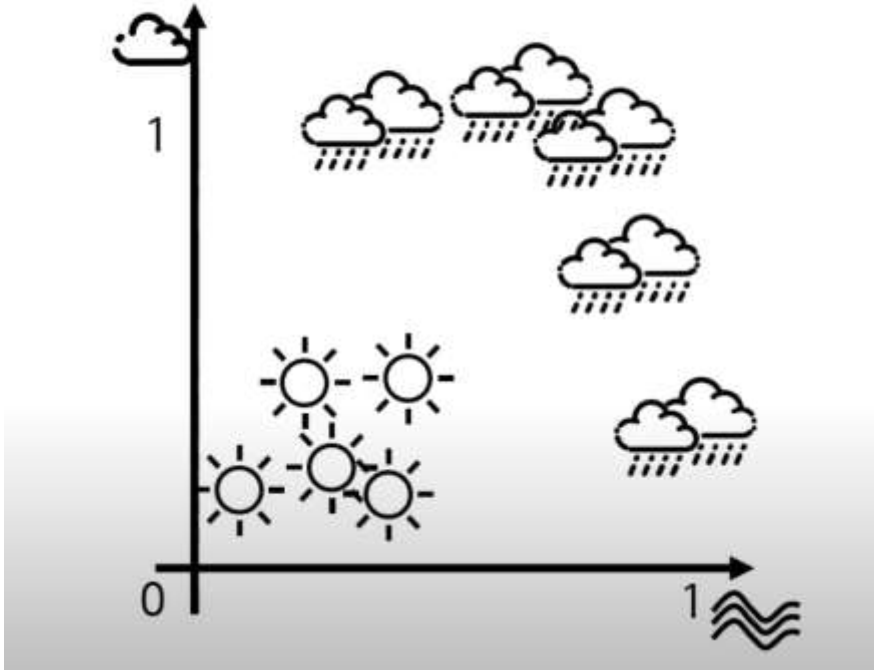


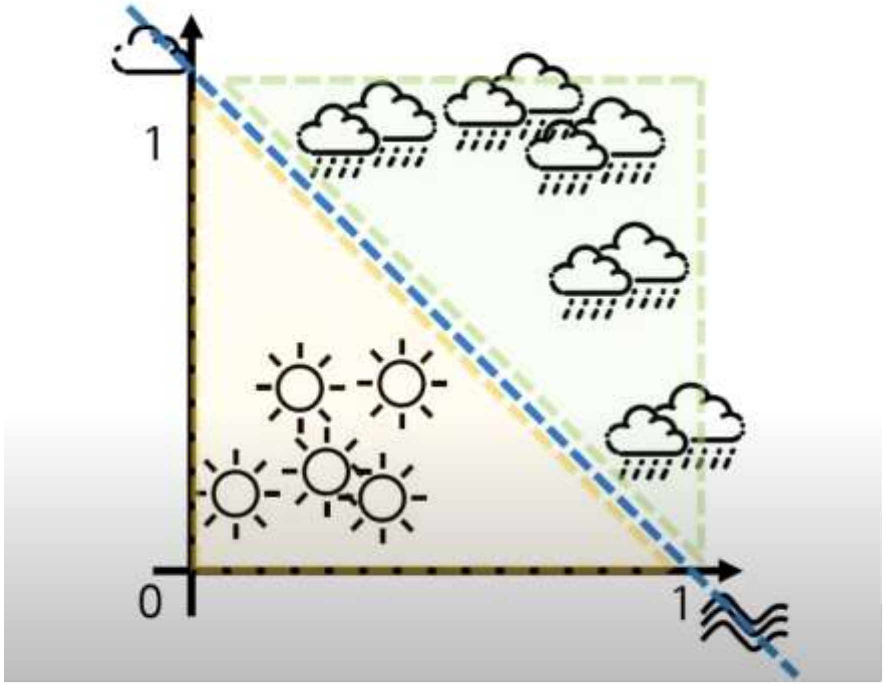


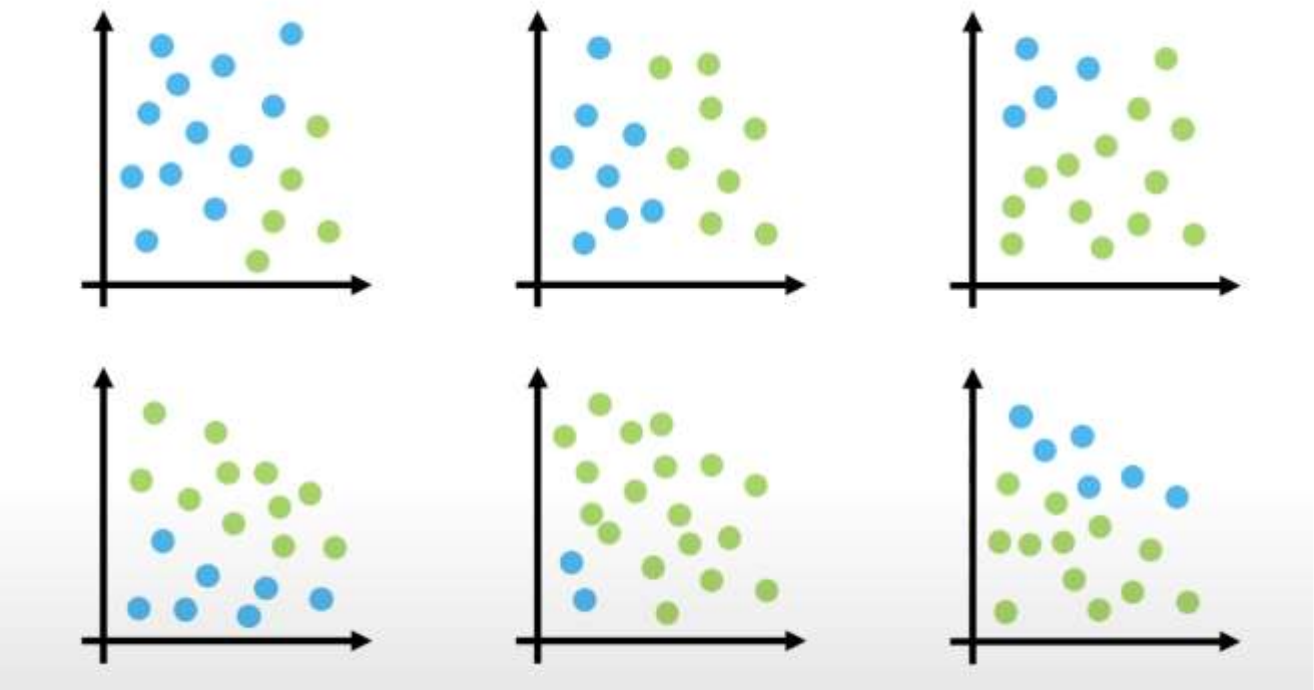
1.0
1
0.8
0.6
0.4
0.2
0.0
0
0.0
0.2
0.4
0.6
0.8
1.0
1

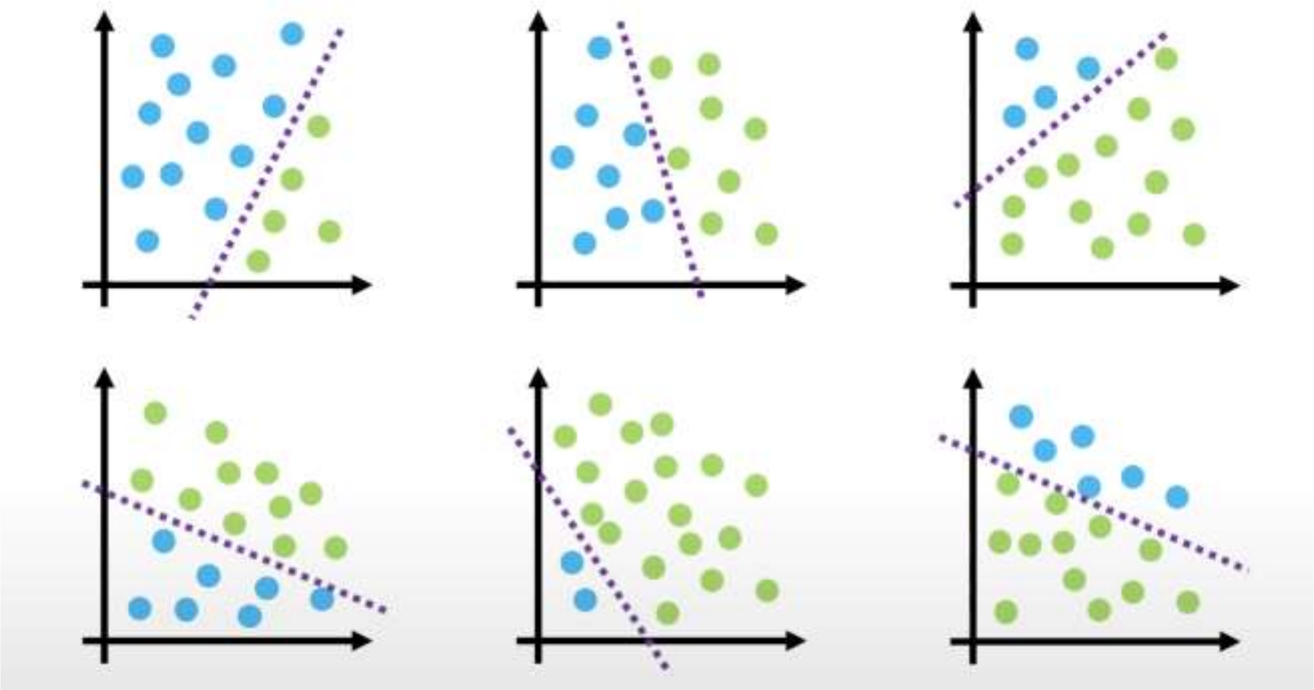

1
0
1
1
0
1

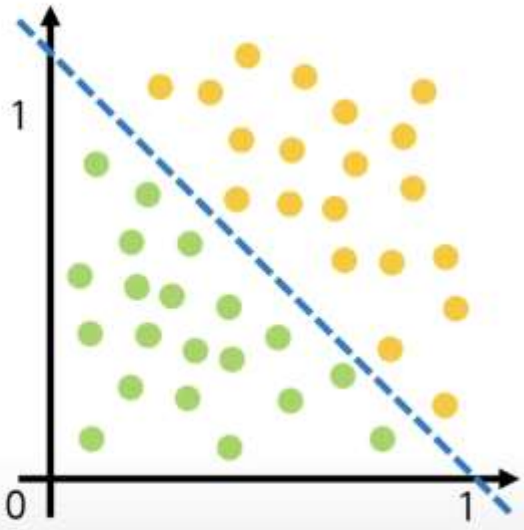

1
0
1

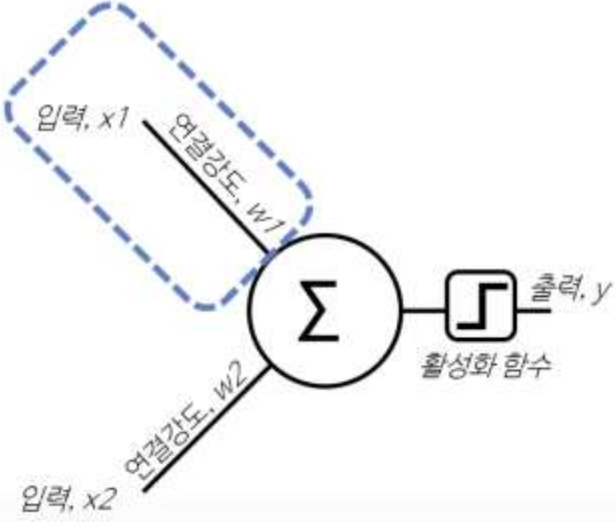

1
0
1

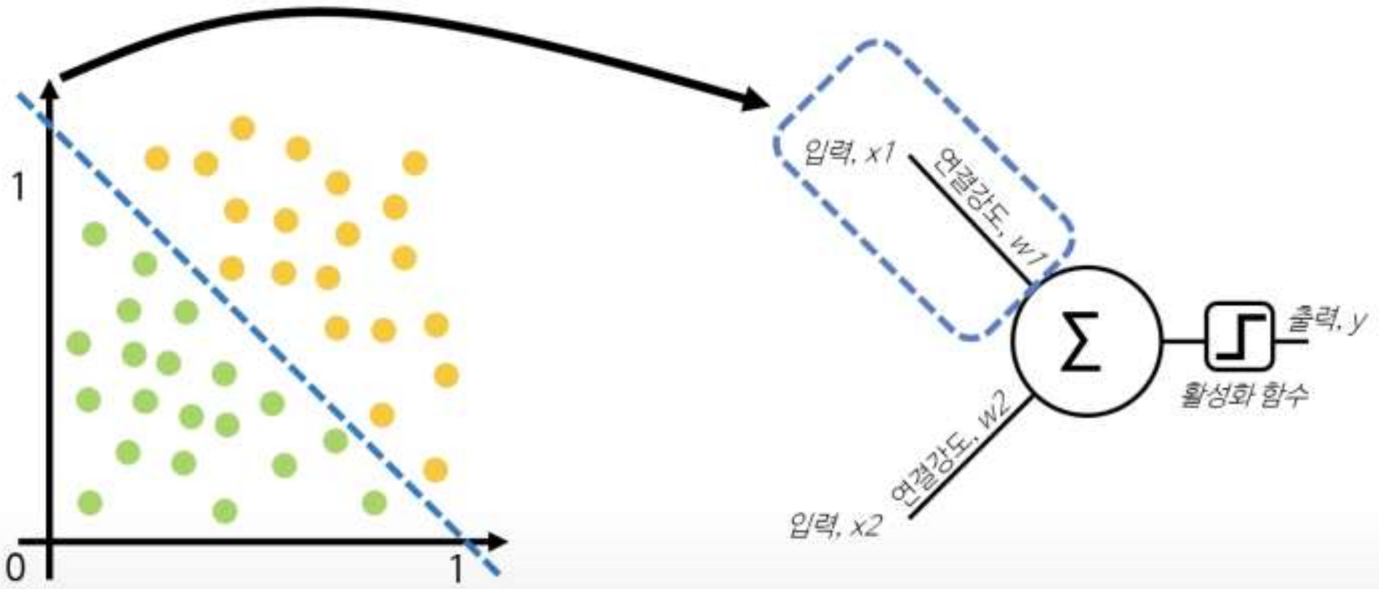

1
0
1



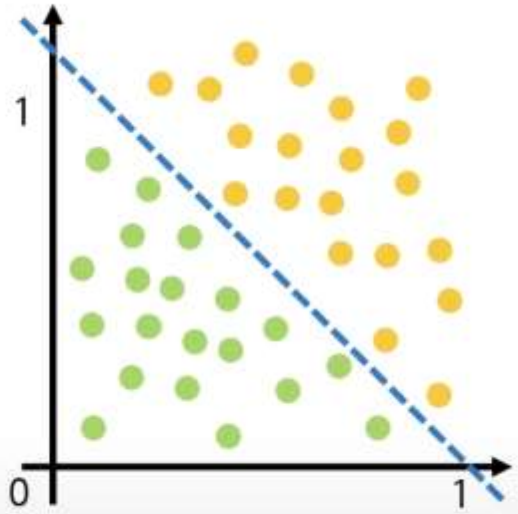


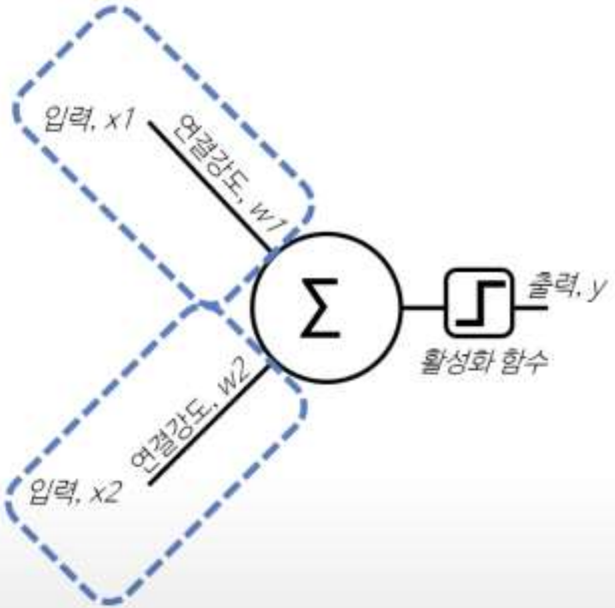

1
0
1

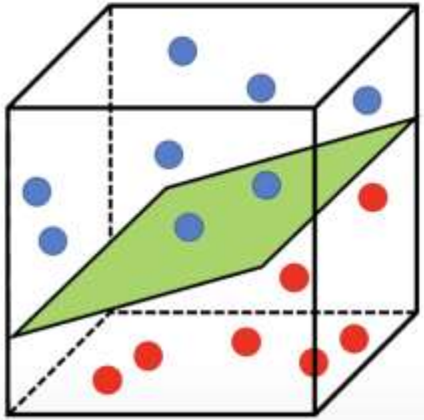

입력, x1
연결강도, w1
Σ
출력, y
활성화 함수
연결강도, w2
입력, x2

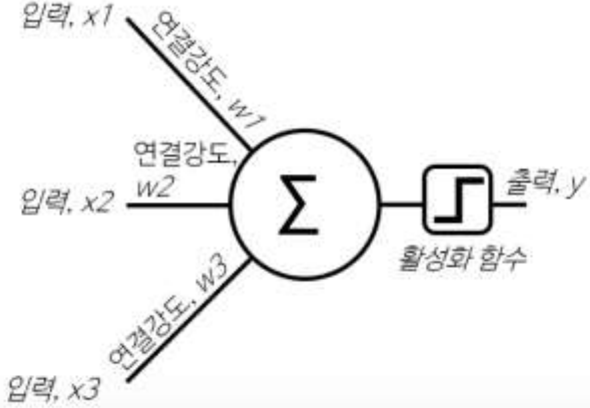


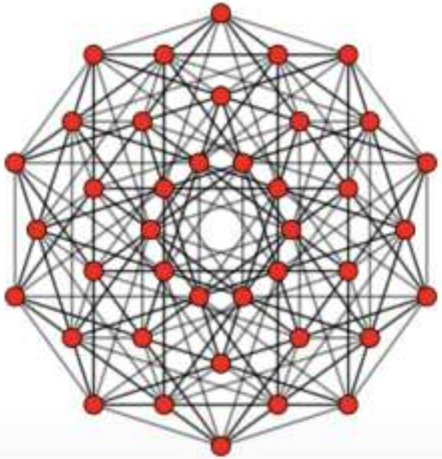
1
0
1
입력, x1
연결강도, w1
Σ
출력, y
활성화 함수
연결강도, w2
입력, x2


1
0
1

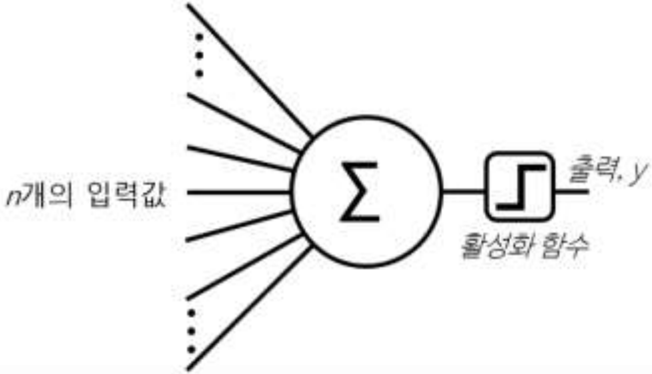

입력, x1
연결강도, w1
입력, x2
연결강도, w2
Σ
출력, y
활성화 함수

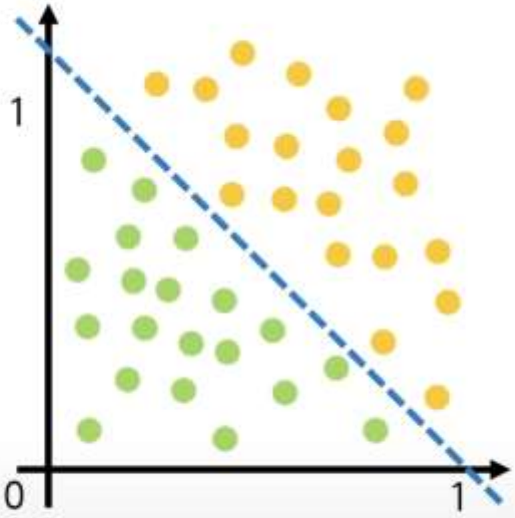



입력, x1
연결강도, w1
연결강도, w2
입력, x2
연결강도, w3
입력, x3
Σ
출력, y
활성화 함수

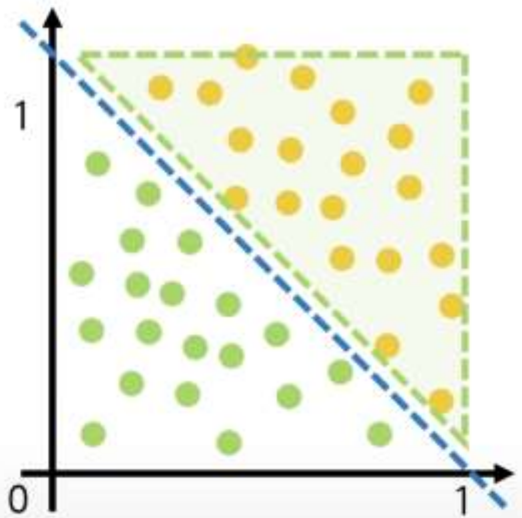


Five-dimensional space. (2022, August 17). In *Wikipedia*
https://en.wikipedia.org/wiki/Five-dimensional_space

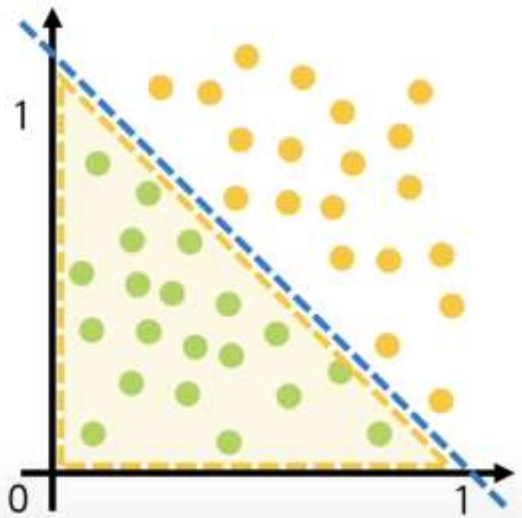




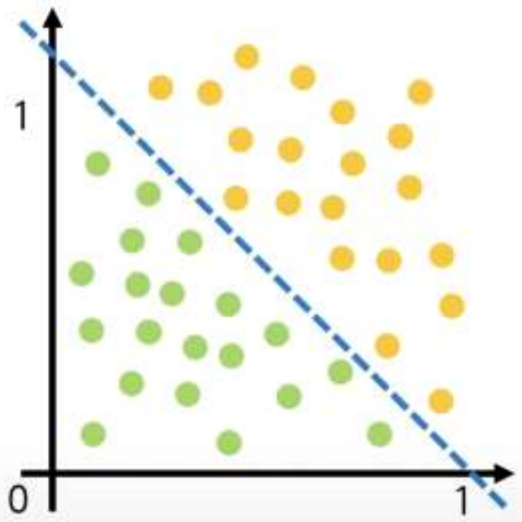
$$w_1x_1 + w_2x_2 = 0.5$$



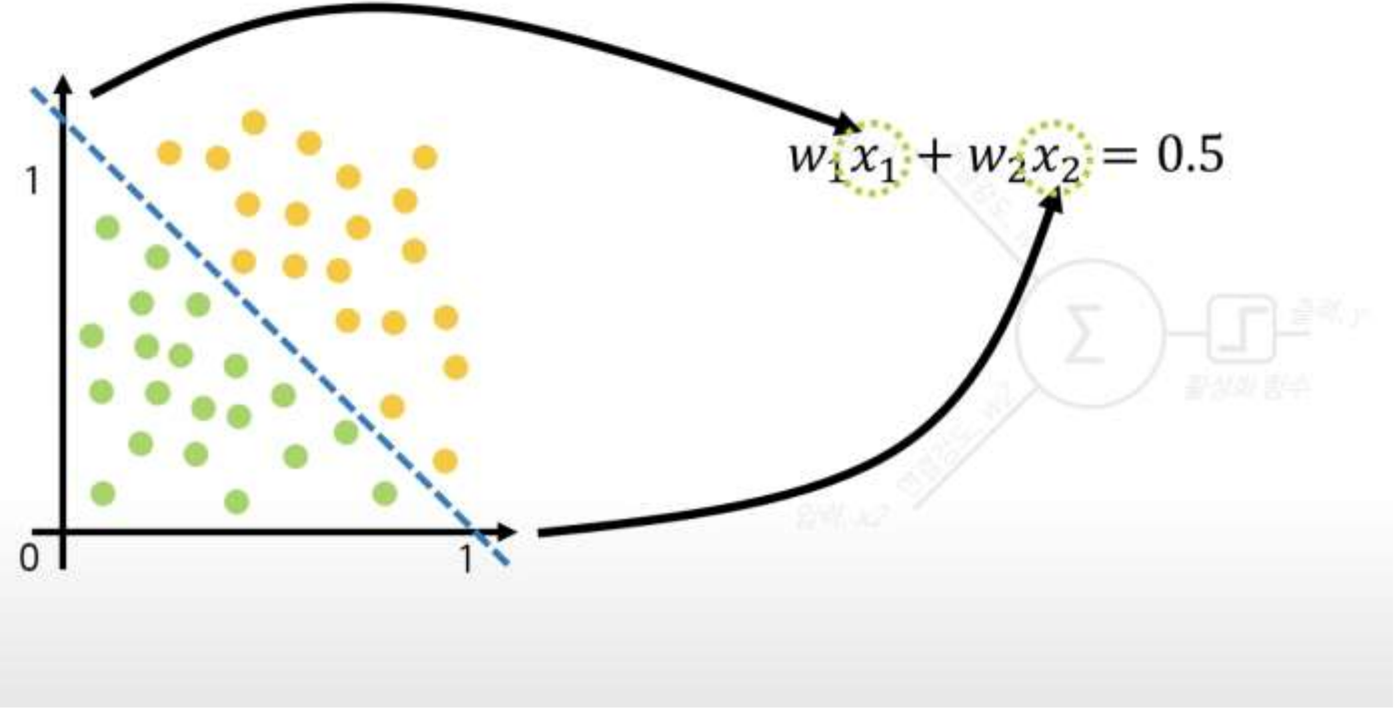
$$w_1x_1 + w_2x_2 > 0.5$$



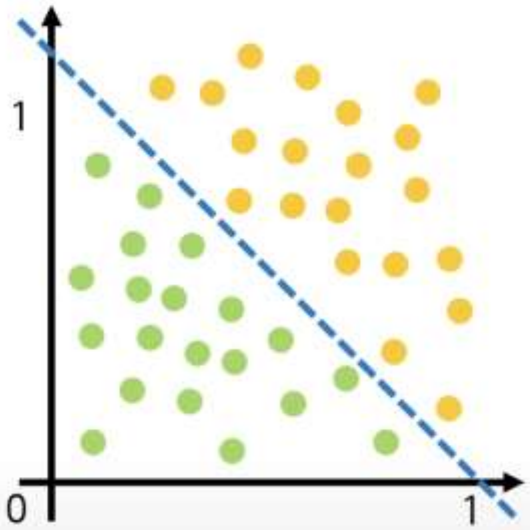
$$w_1x_1 + w_2x_2 < 0.5$$



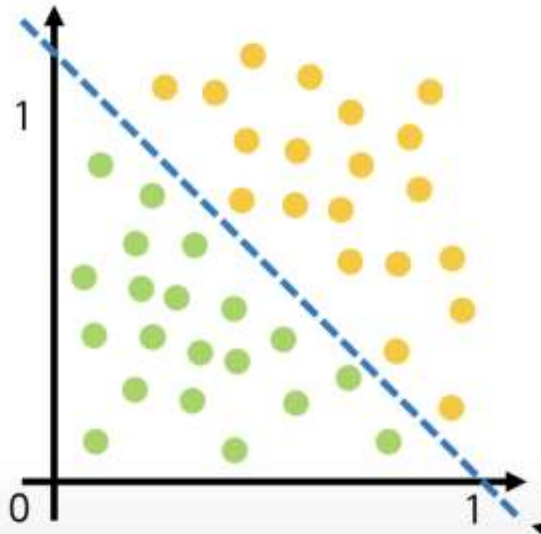
$$w_1 x_1 + w_2 x_2 = 0.5$$



1
0
1
$w_1 x_1 + w_2 x_2 = 0.5$



$$w_1 x_1 + w_2 x_2 = 0.5$$

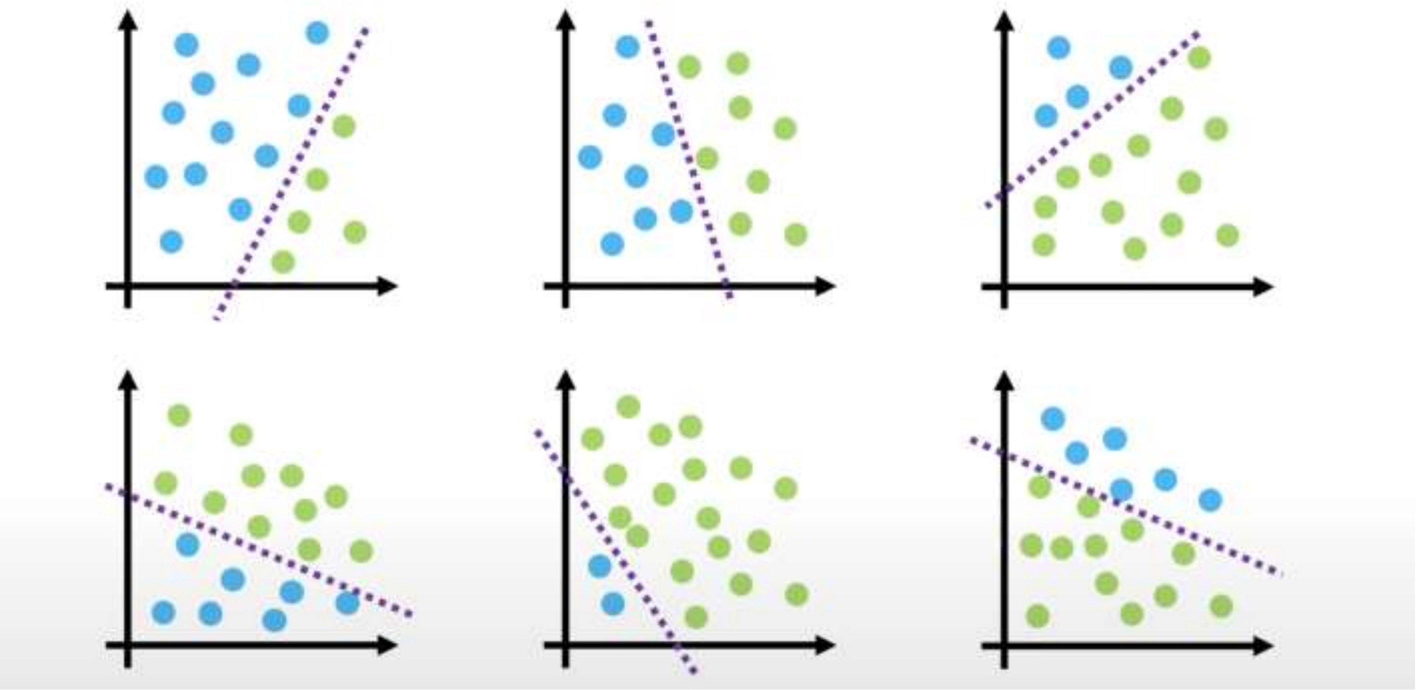


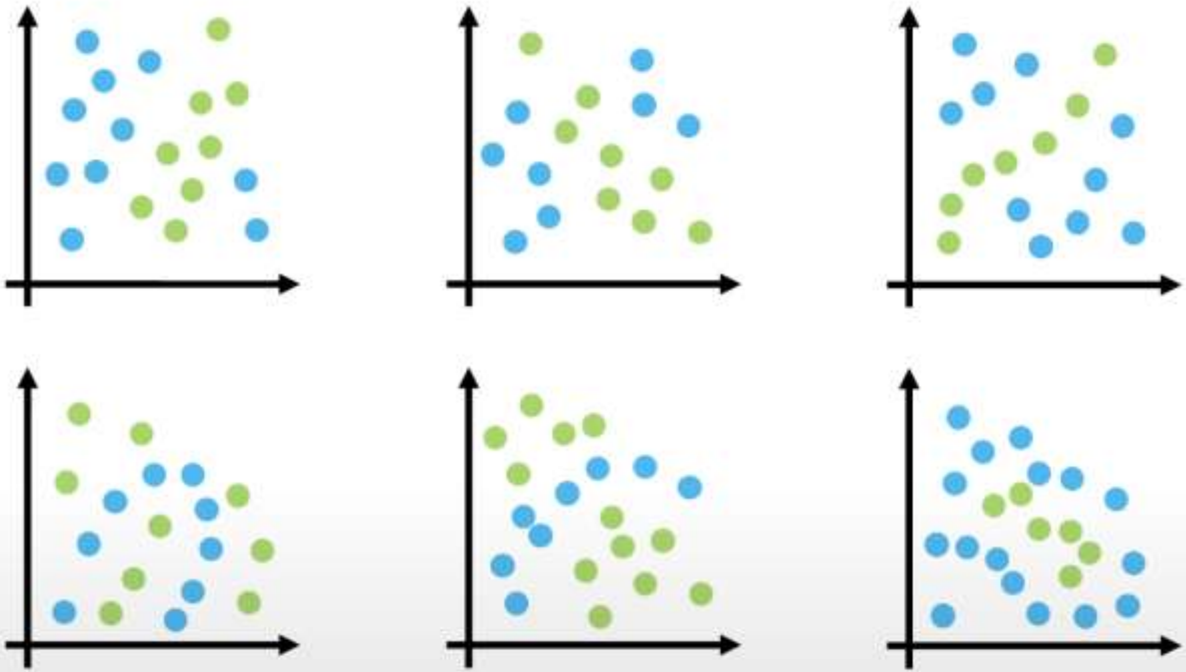
$$w_1x_1 + w_2x_2 = 0.5$$

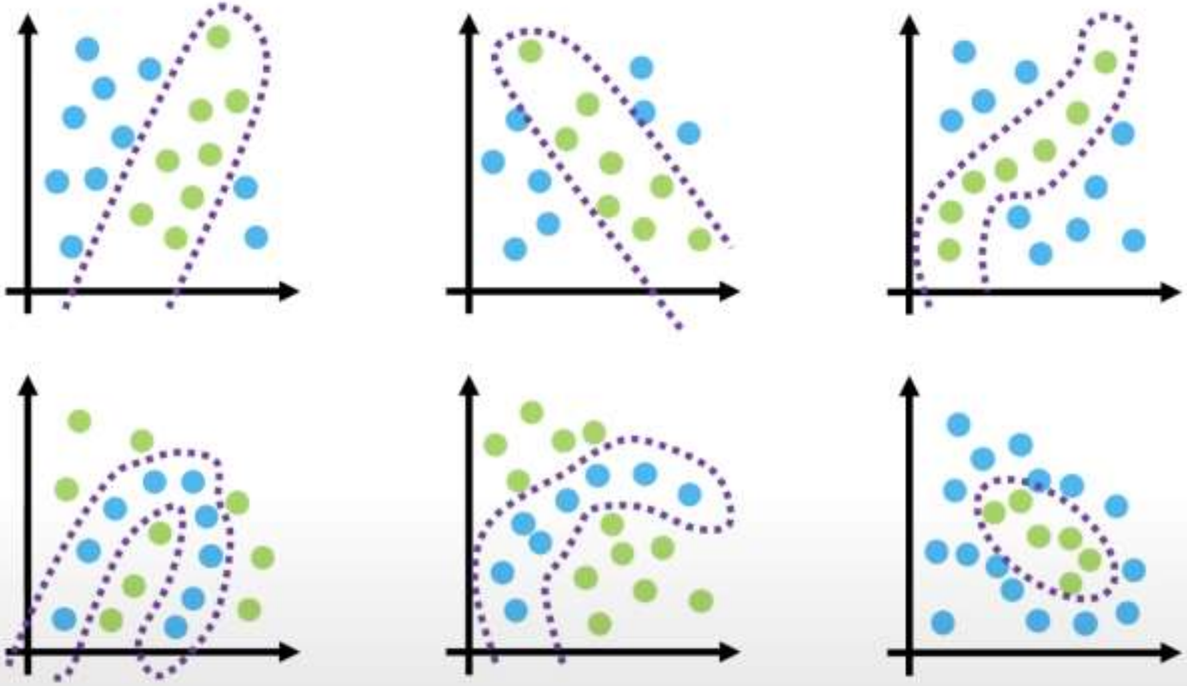
$$0.499999x_1 + 0.499999x_2 = 0.5$$

$$x_1 + x_2 = 1$$

$$x_1 = -x_2 + 1$$

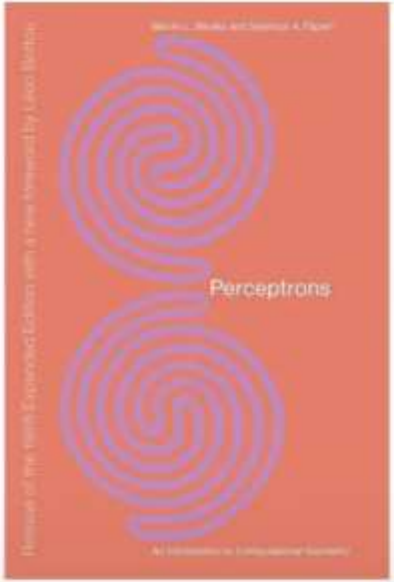




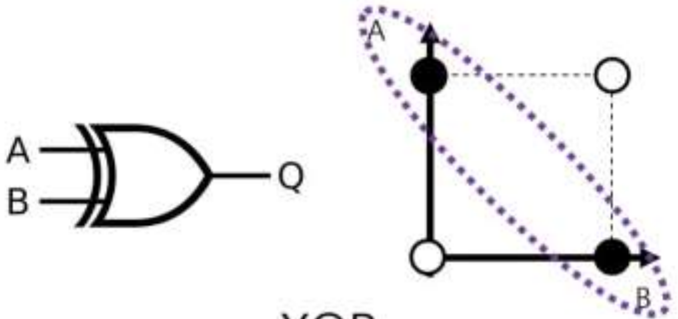



NO!

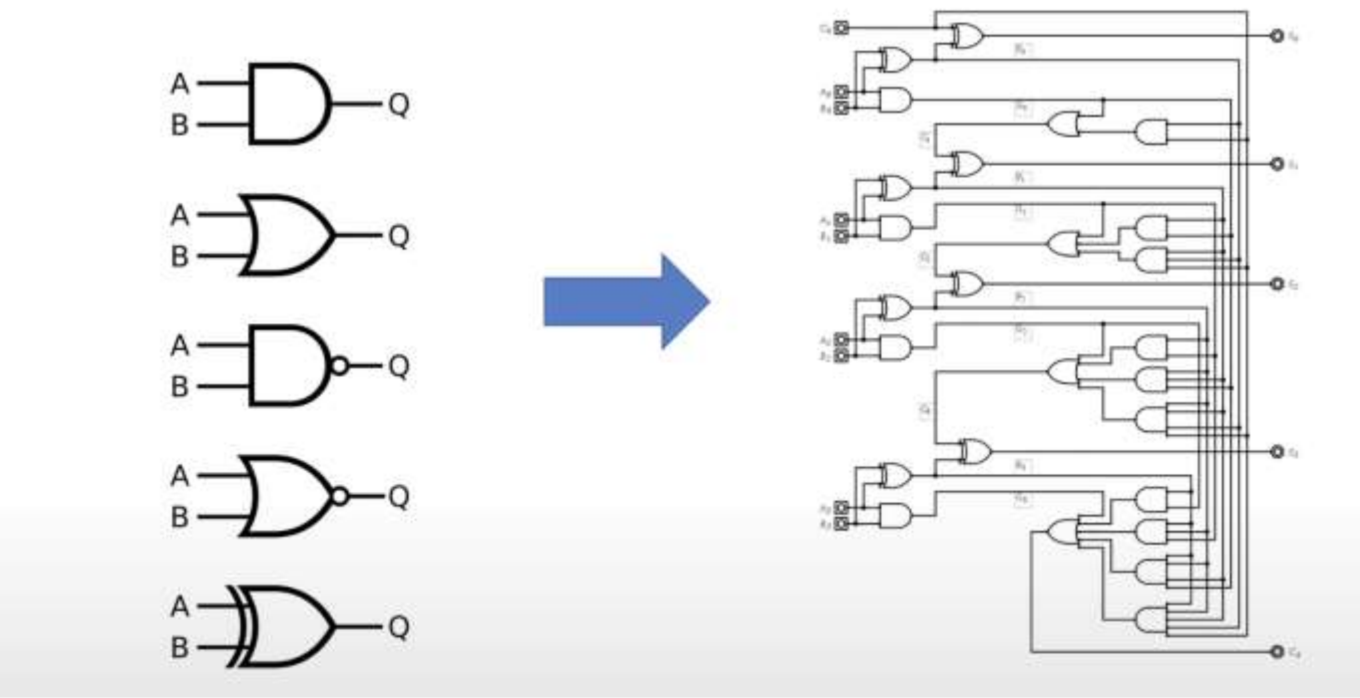
https://news.mit.edu/2016/seymc
papert-pioneer-of-constructionist
learning-dies-0801



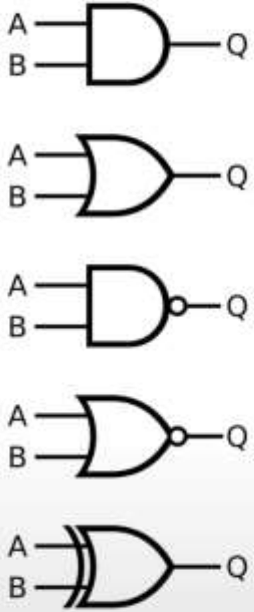
mit.edu/2016/marvin-minsky-obituary-0125

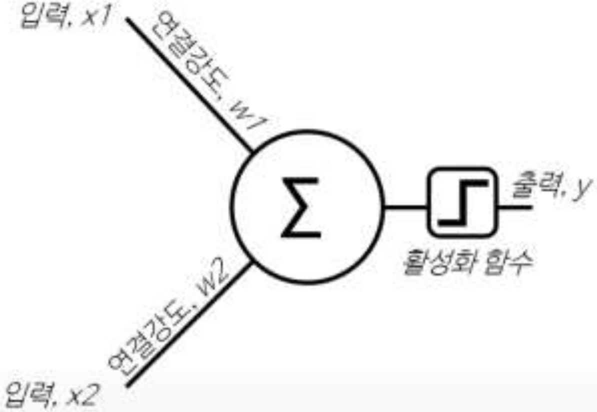


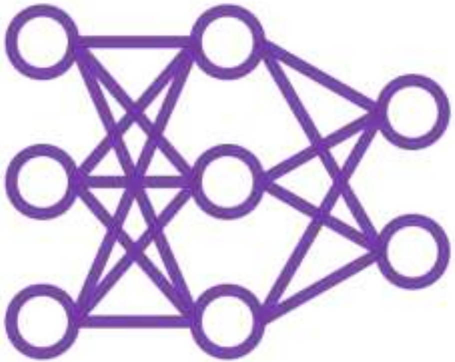
XOR

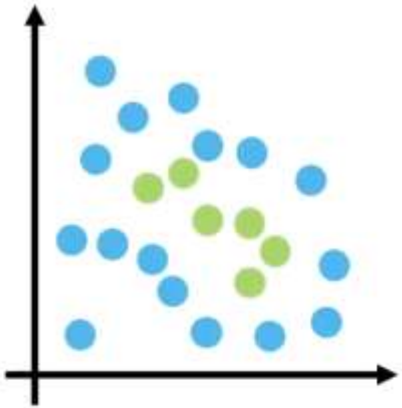


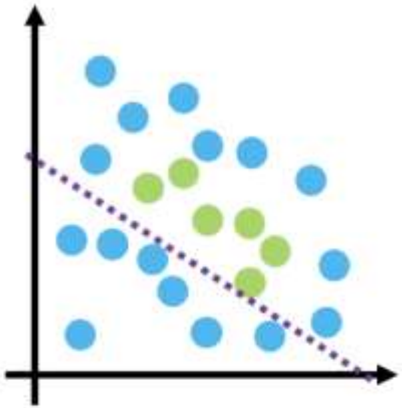
A
B
Q
A
B
Q
A
B
Q
A
B
Q
A
B
Q

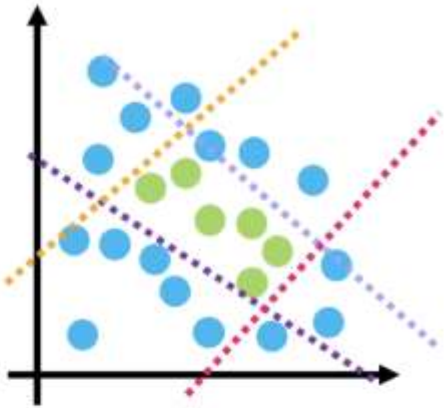

A
B
Q

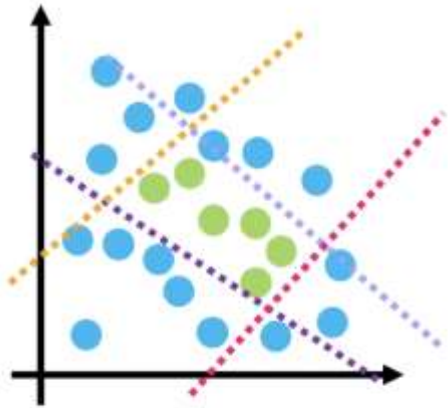

입력, x1
연결강도, w1
입력, x2
연결강도, w2
Σ
출력, y
활성화 함수

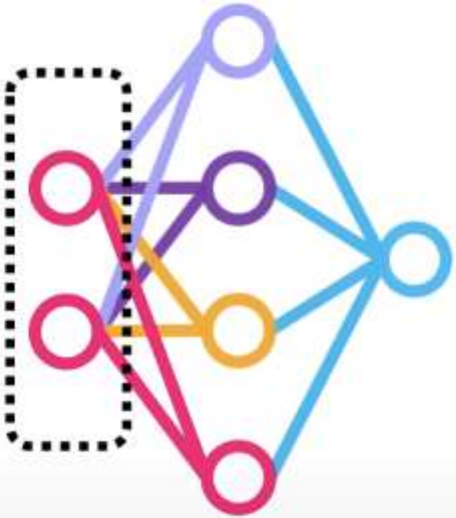


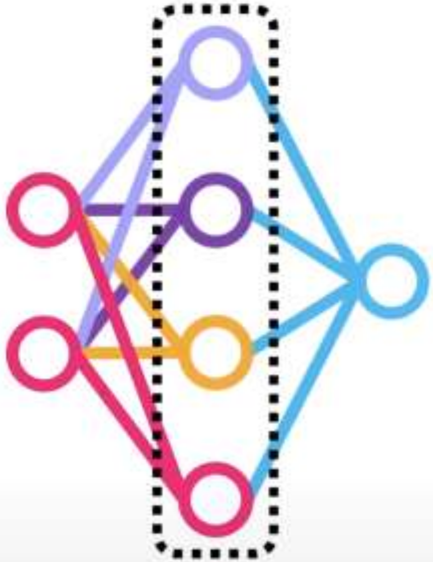


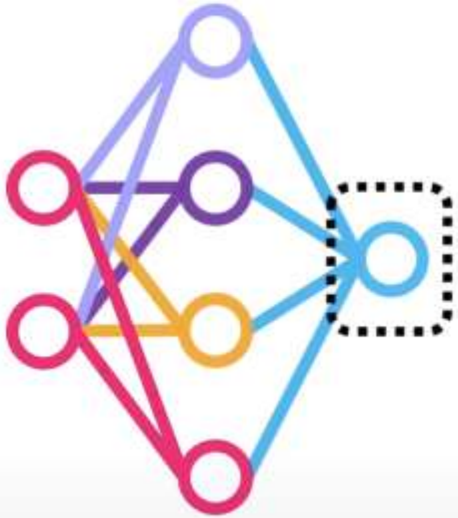


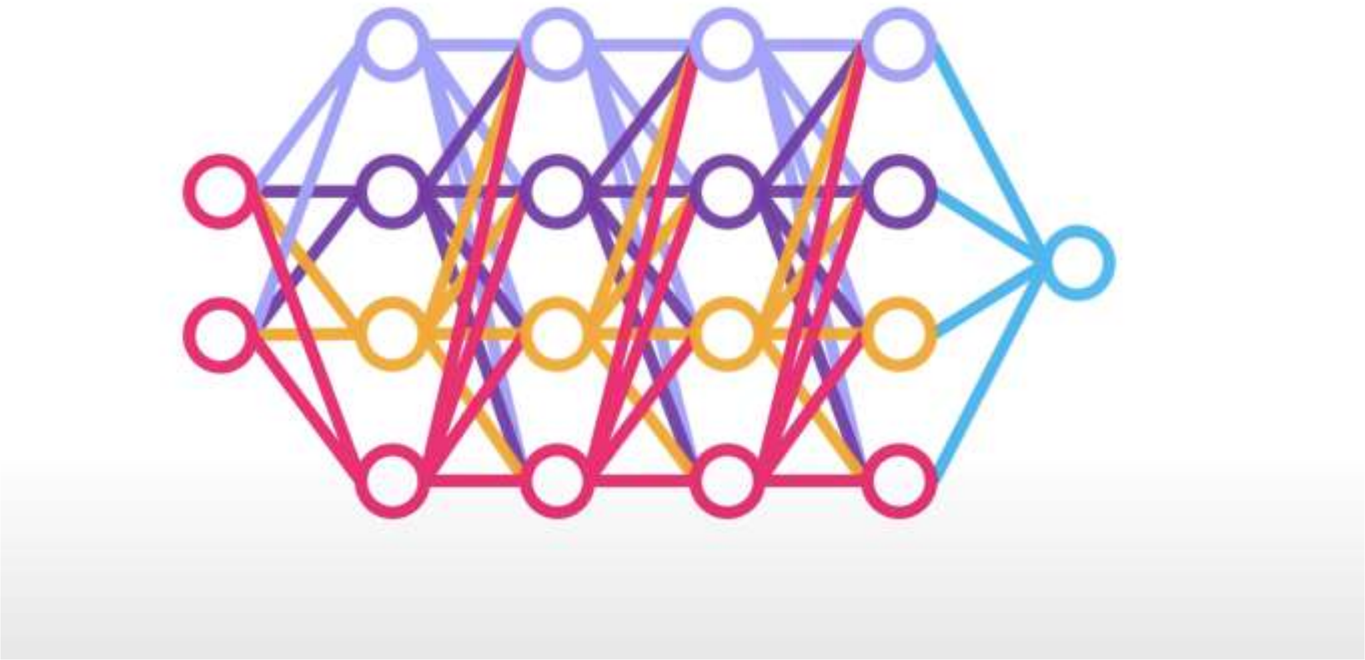


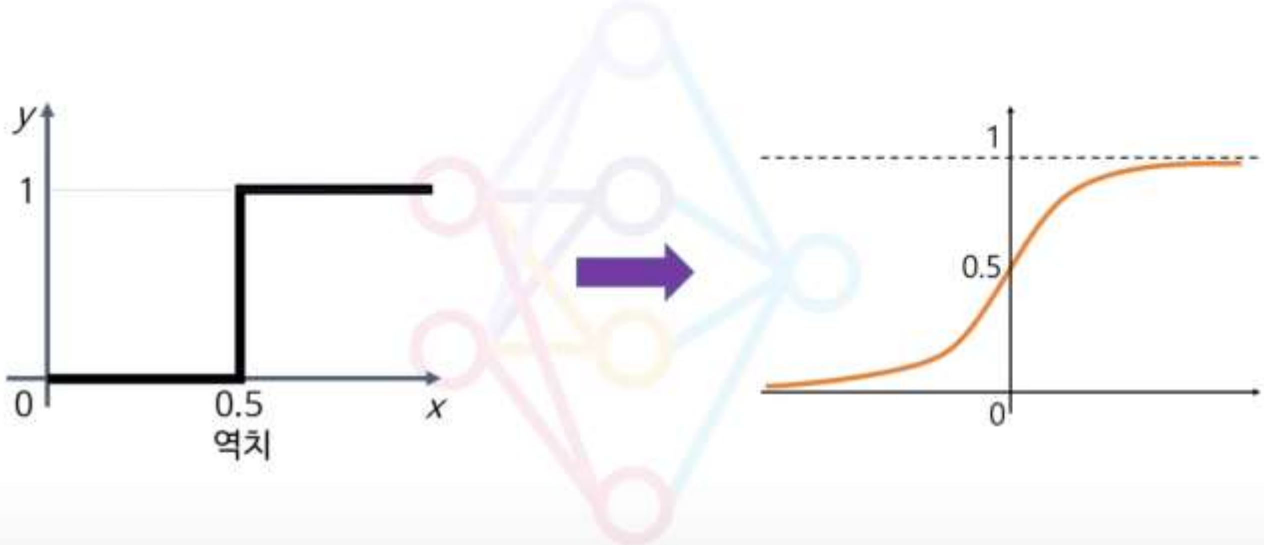




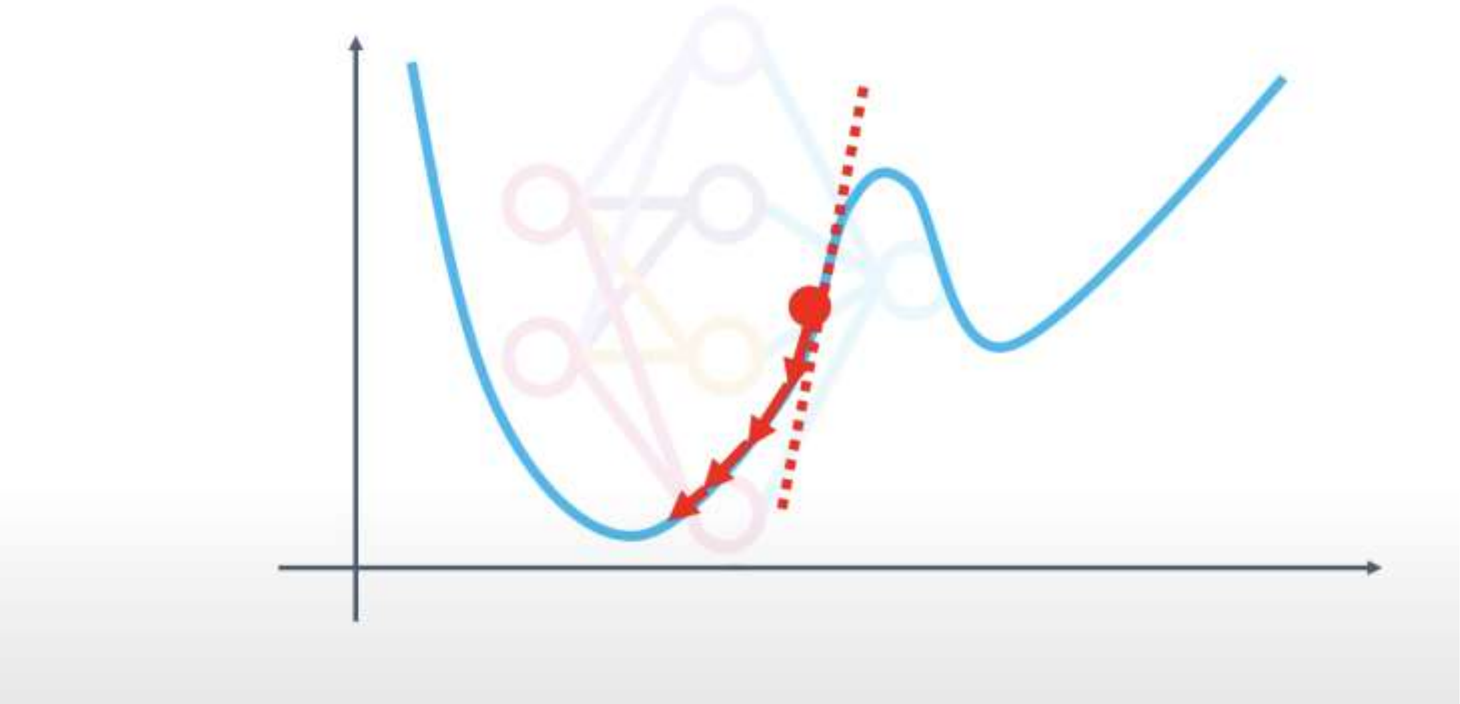


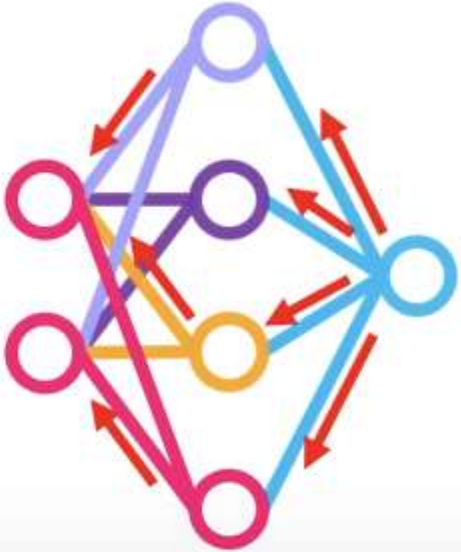


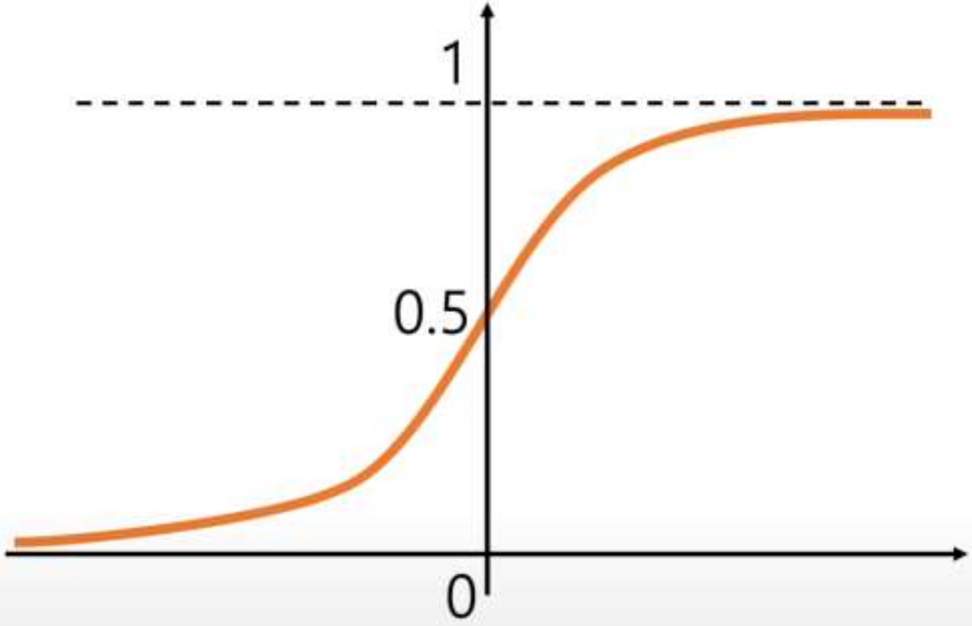


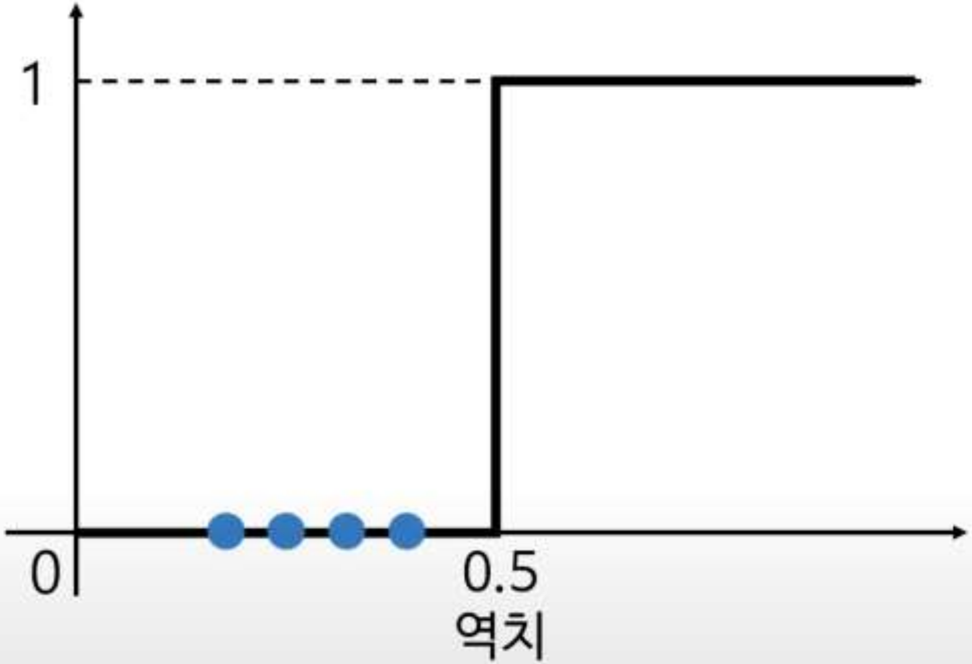


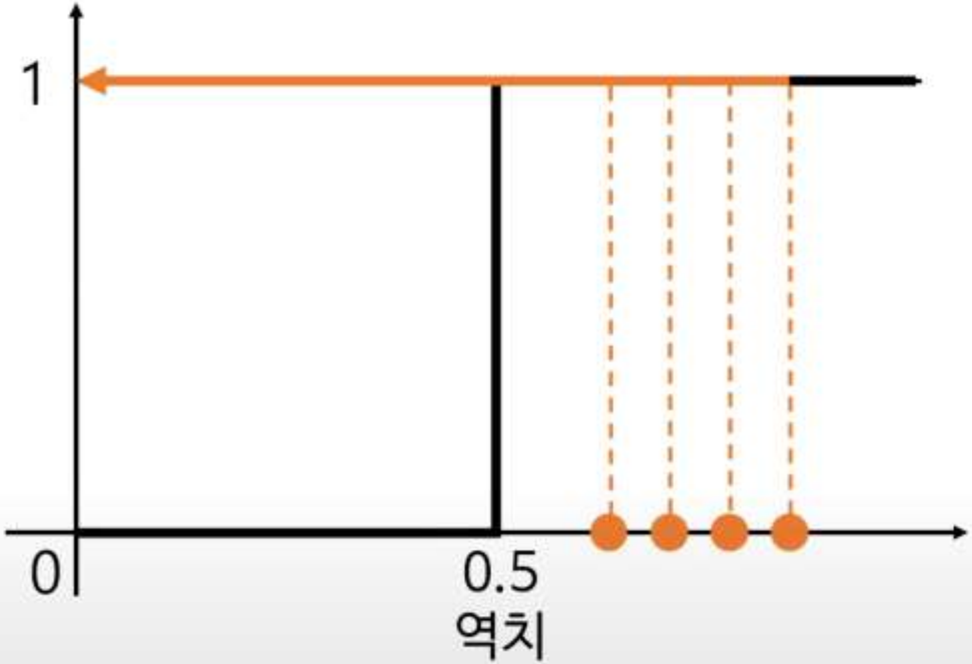
y
1
0
0.5
역치
x
1
0.5
0

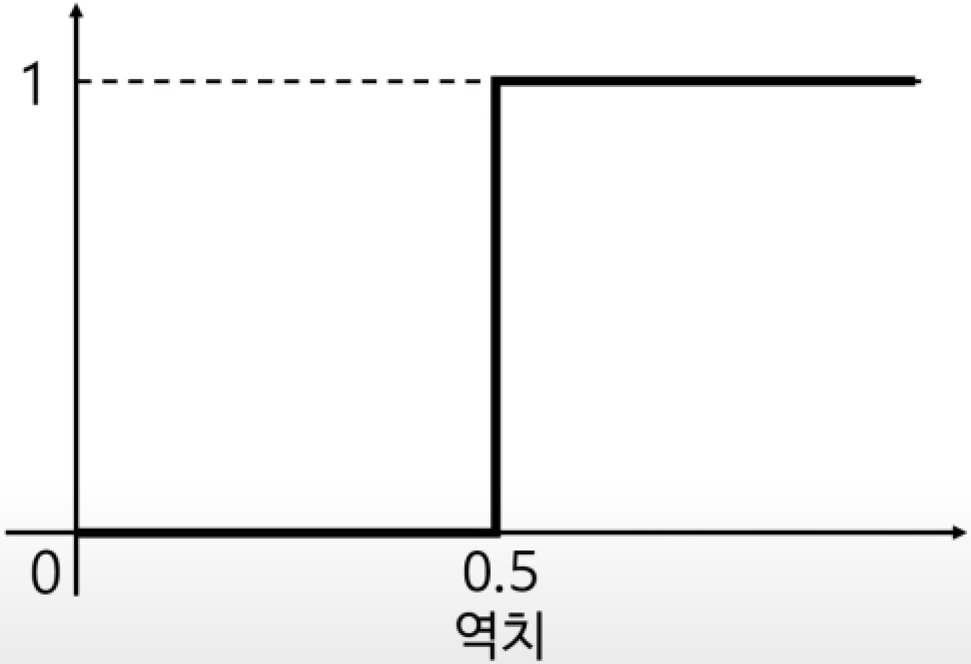




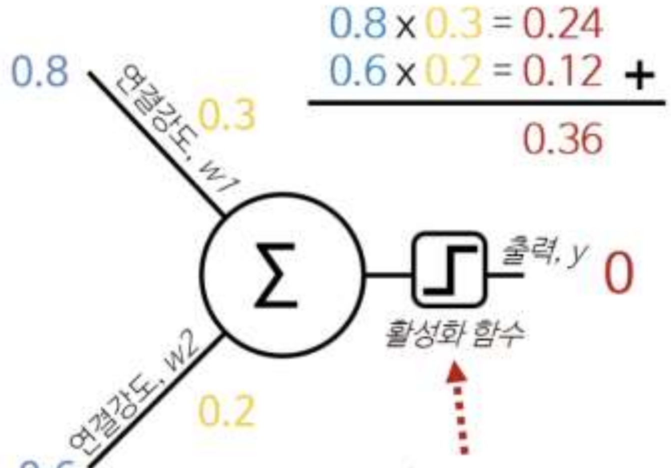

1
0.5
0


1
0
0.5
역치


1
0
0.5
역치


1
0
0.5
역치

x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

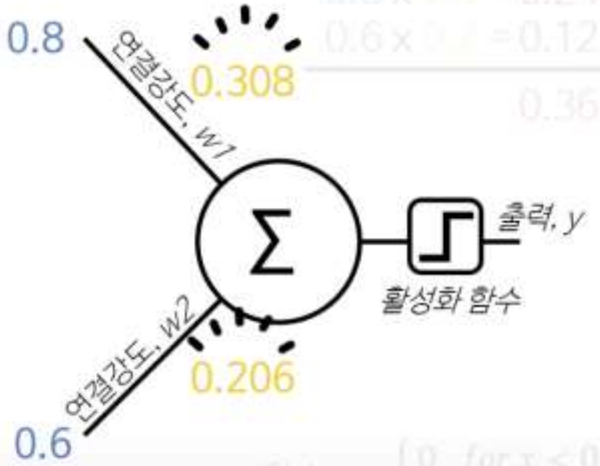


$$f(x) = \begin{cases} 0 & for\ x < 0.5 \\ 1 & for\ x \geqslant 0.5 \end{cases}$$

y
1
1
0
0
1
0
0

y’
0

x1	x2
0.6	0.9
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2



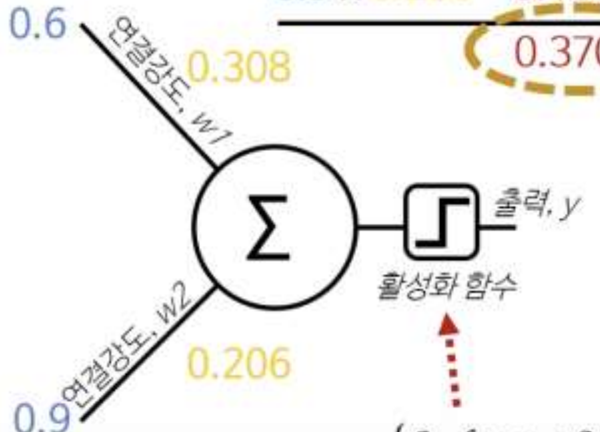
y	y'
1	0
1	
0	
0	
1	
0	
0	

x1	x2
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

$0.6 \times 0.308 = 0.1848$

$0.9 \times 0.206 = 0.1854$ +

0.3702

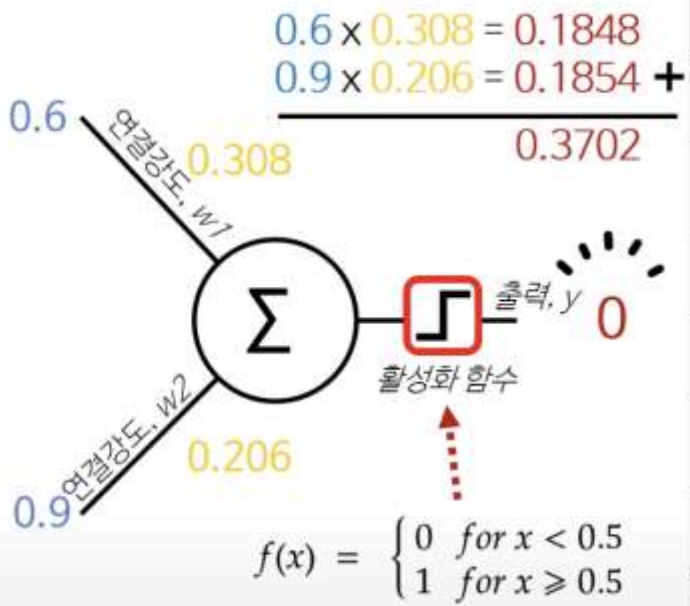


$$f(x) = \begin{cases} 0 & for\ x < 0.5 \\ 1 & for\ x \geqslant 0.5 \end{cases}$$

y
1
1
0
0
1
0
0

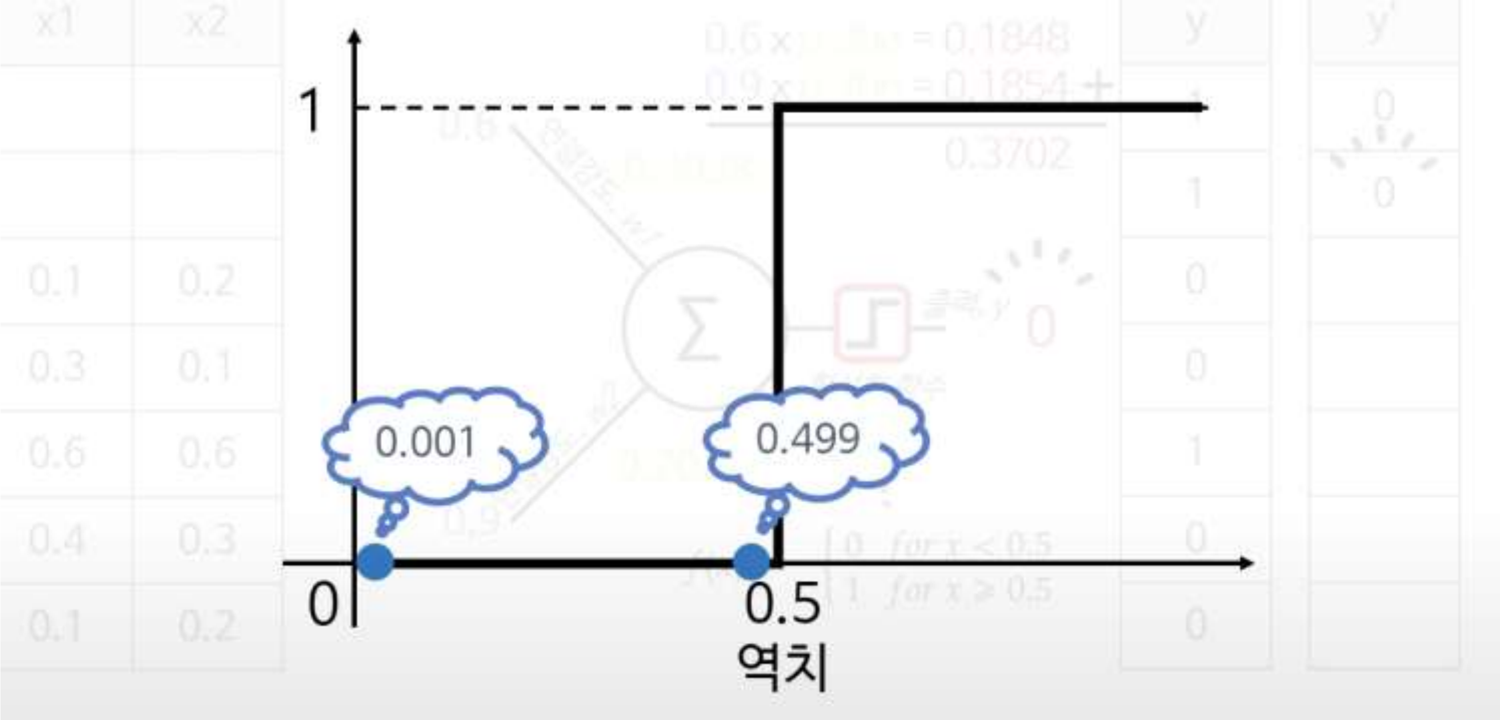
y’
0

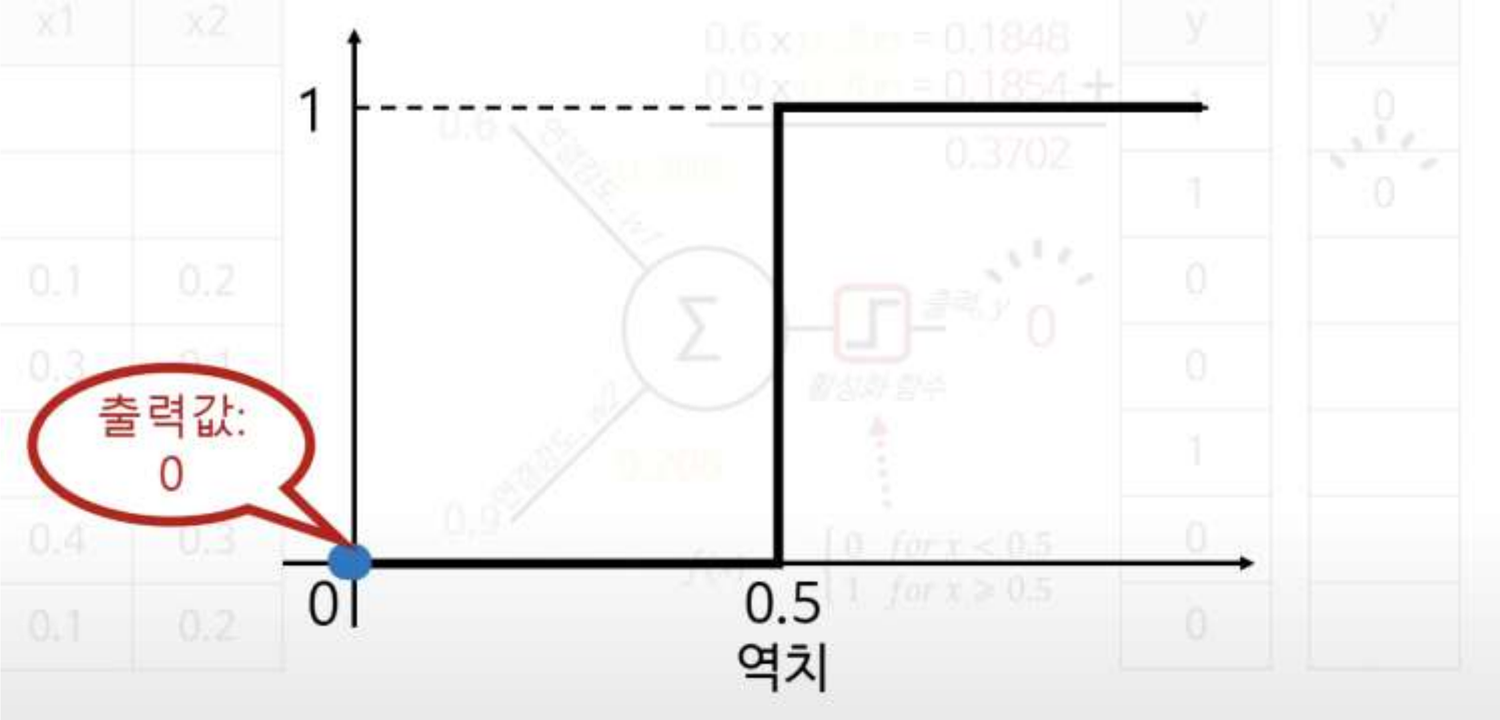
x1	x2
0.1	0.2
0.3	0.1
0.6	0.6
0.4	0.3
0.1	0.2

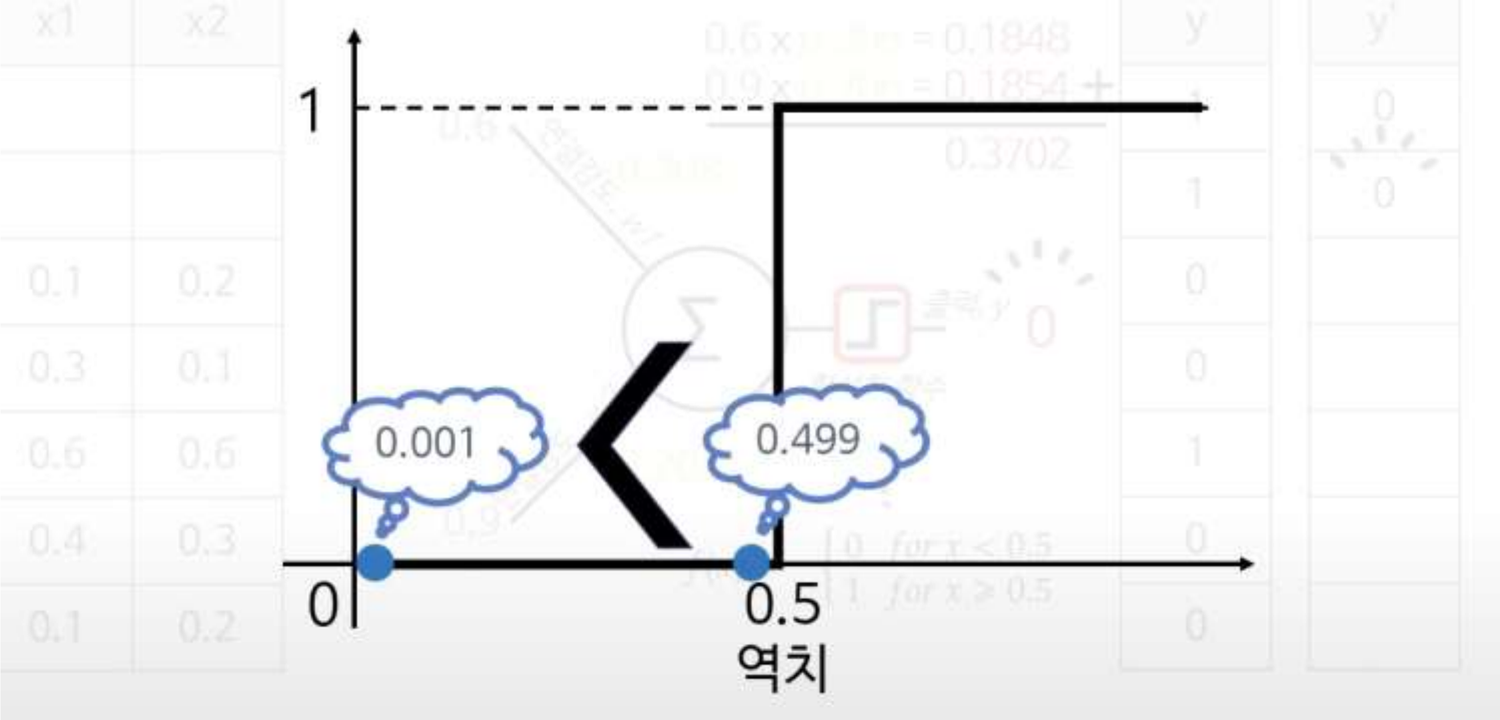


y
1
1
0
0
1
0
0

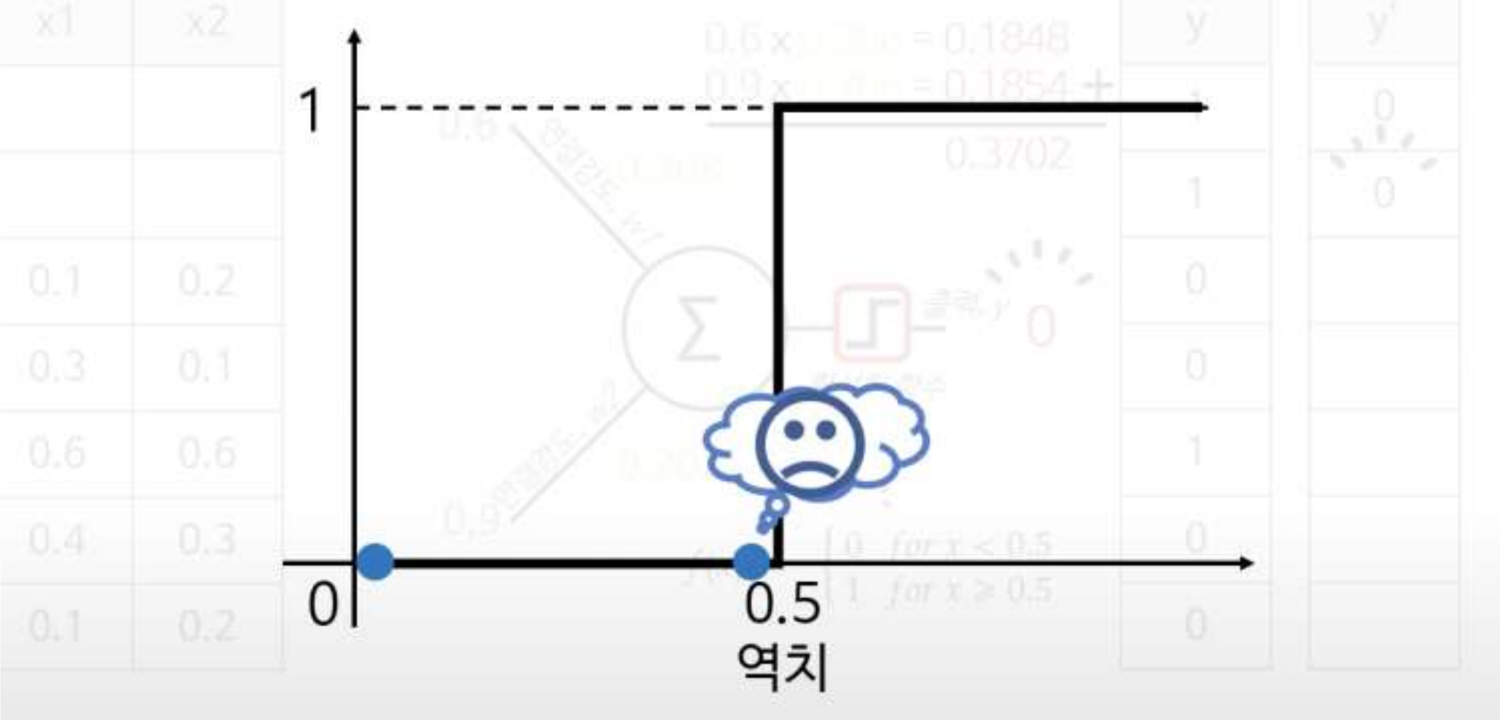
y’
0
0

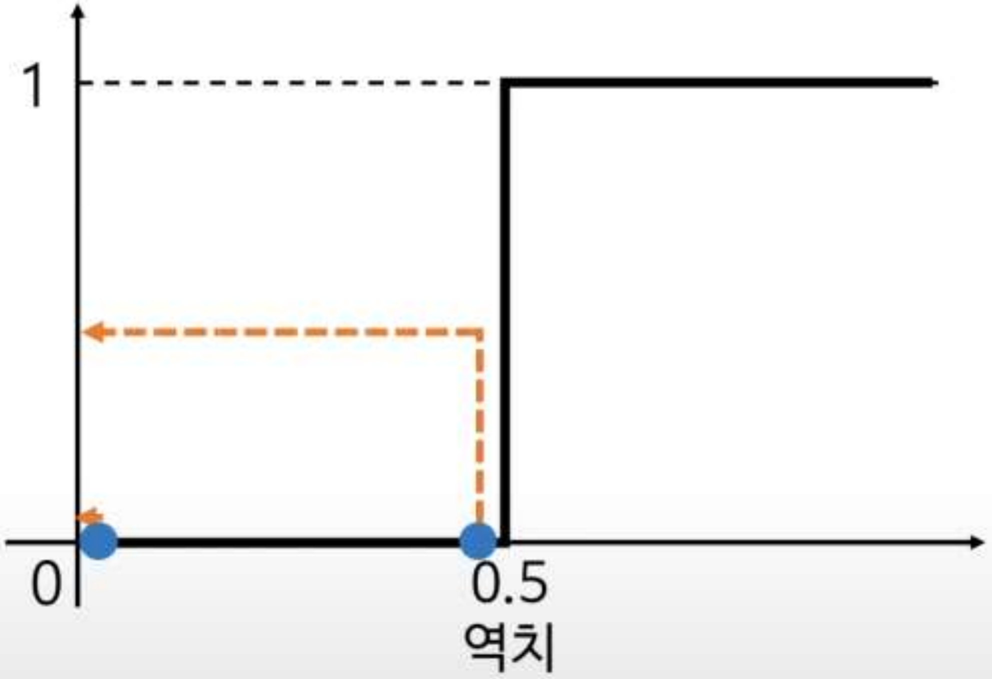

1
0
0.001
0.499
0.5
역치

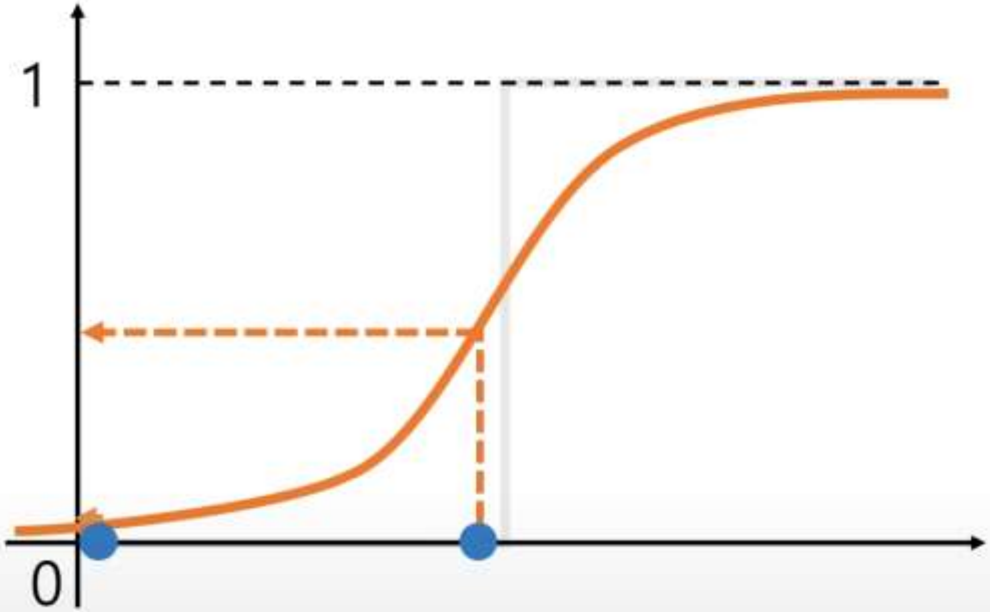

1
출력값:
0
0
0.5
역치

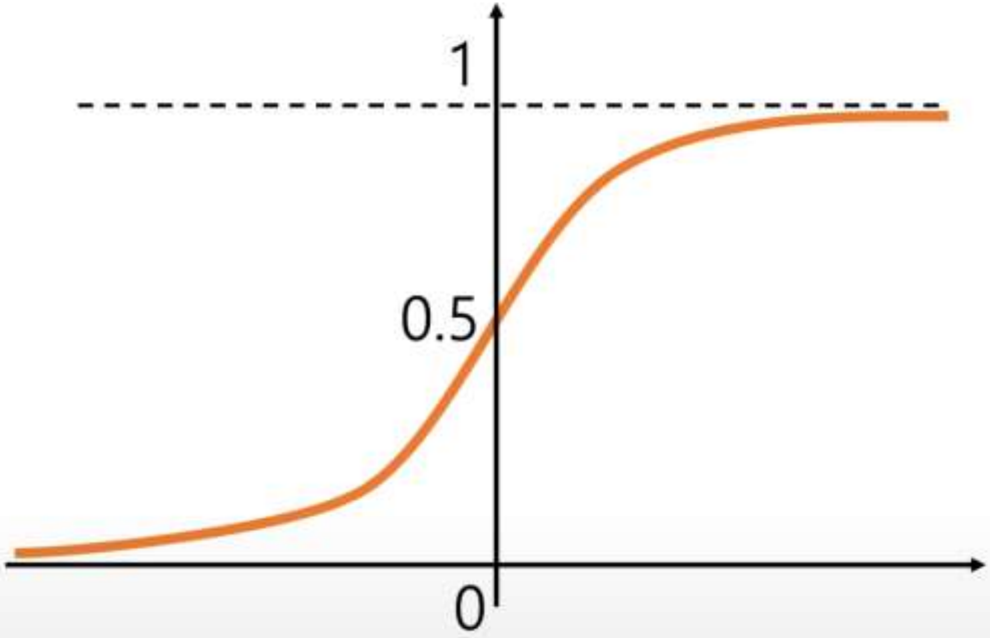


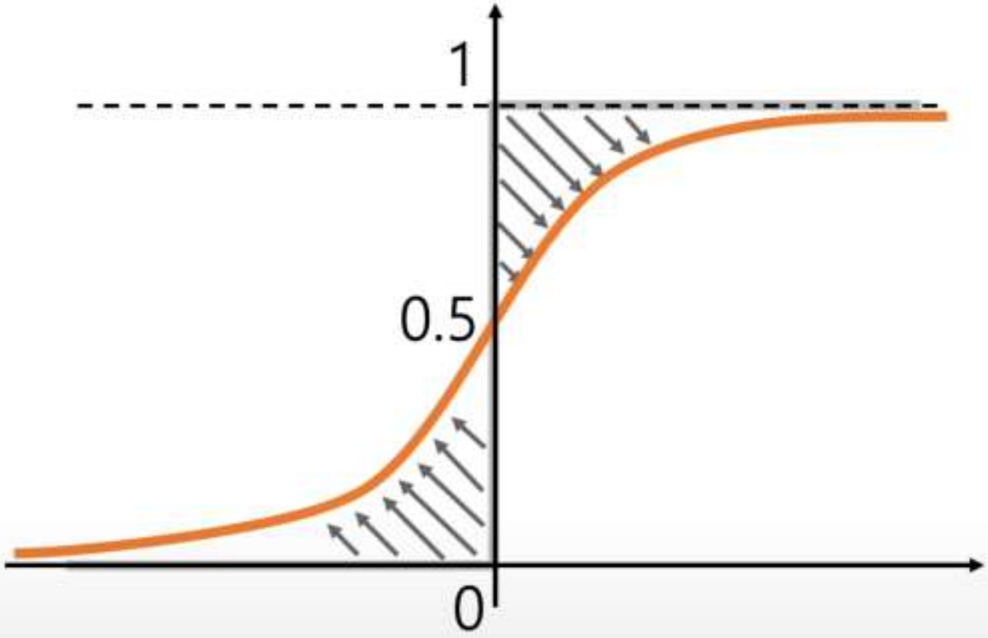
1
0
0.001
0.499
0.5
역치

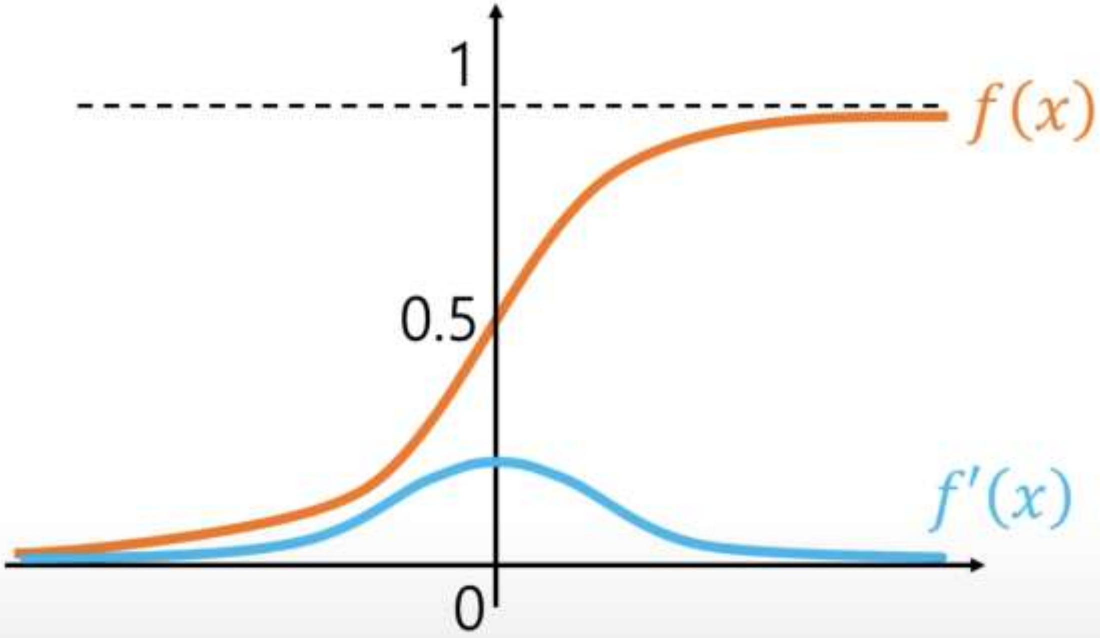

1
0
0.5
역치

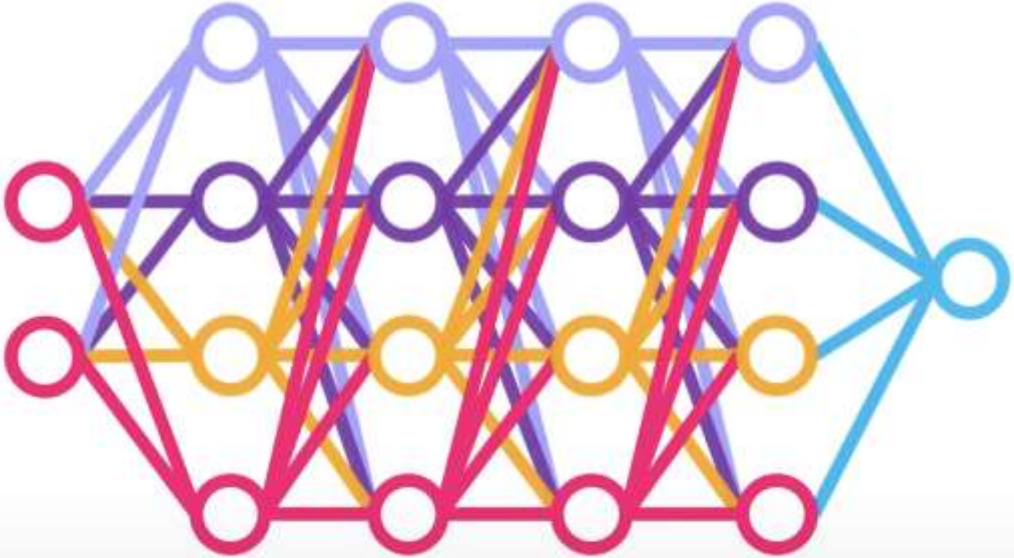

1
0
0.5
역치

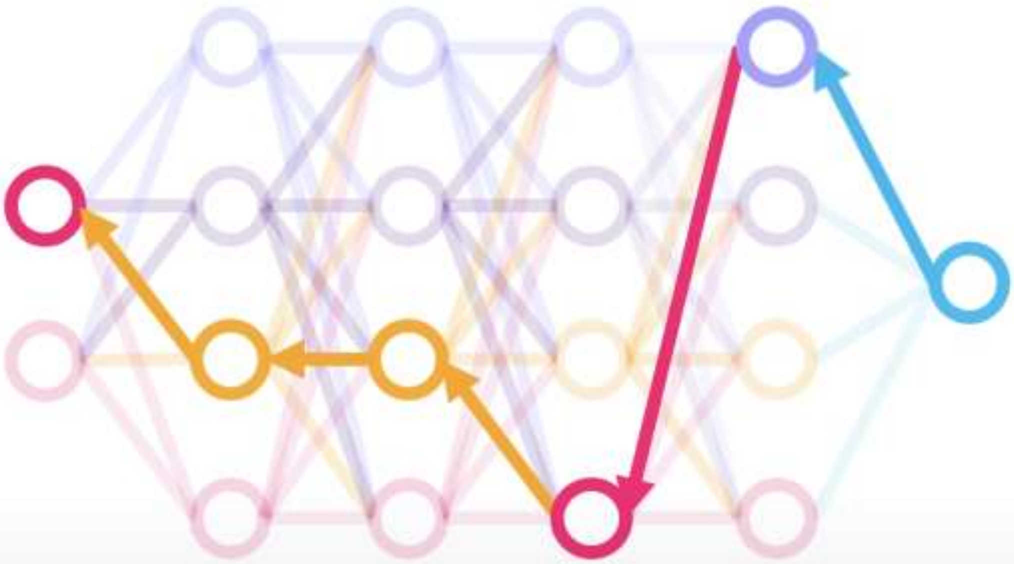

1
0

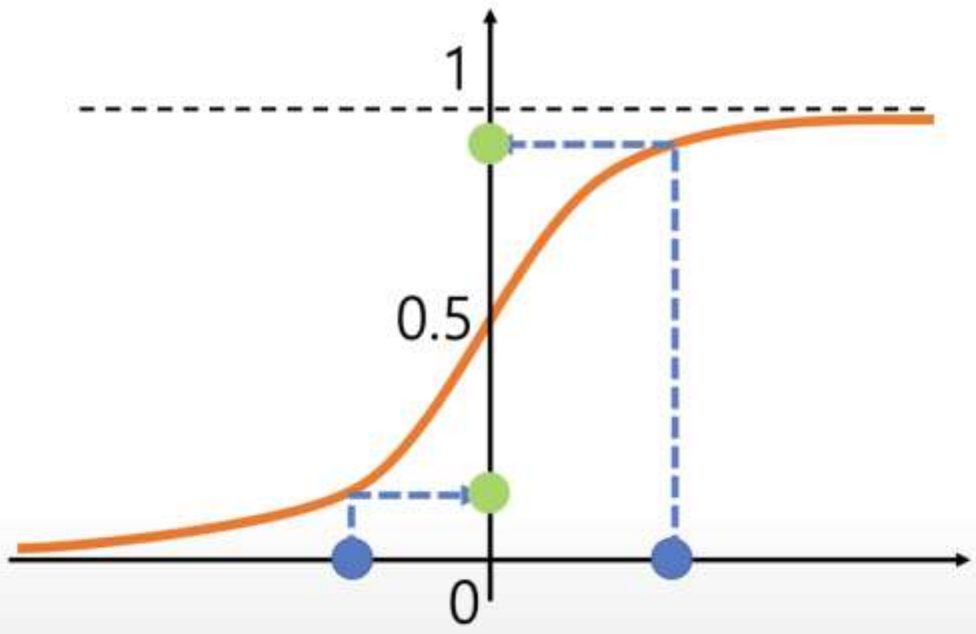

1
0.5
0

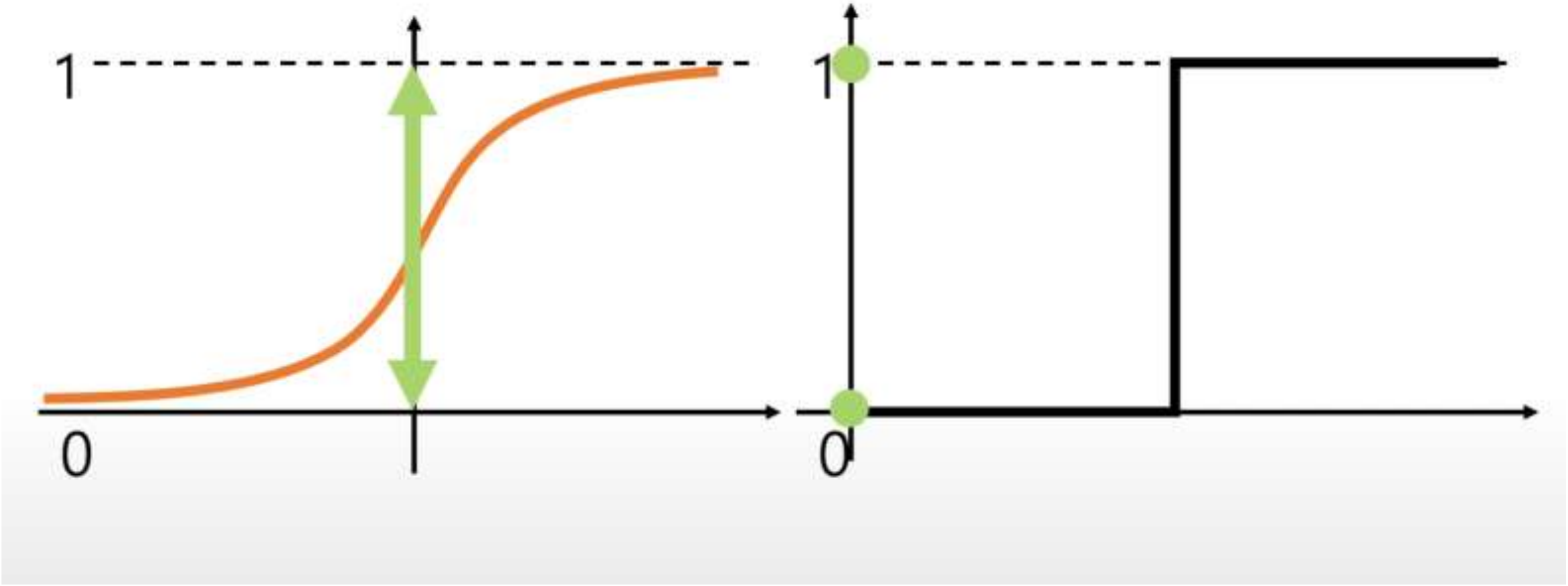

1
0.5
0

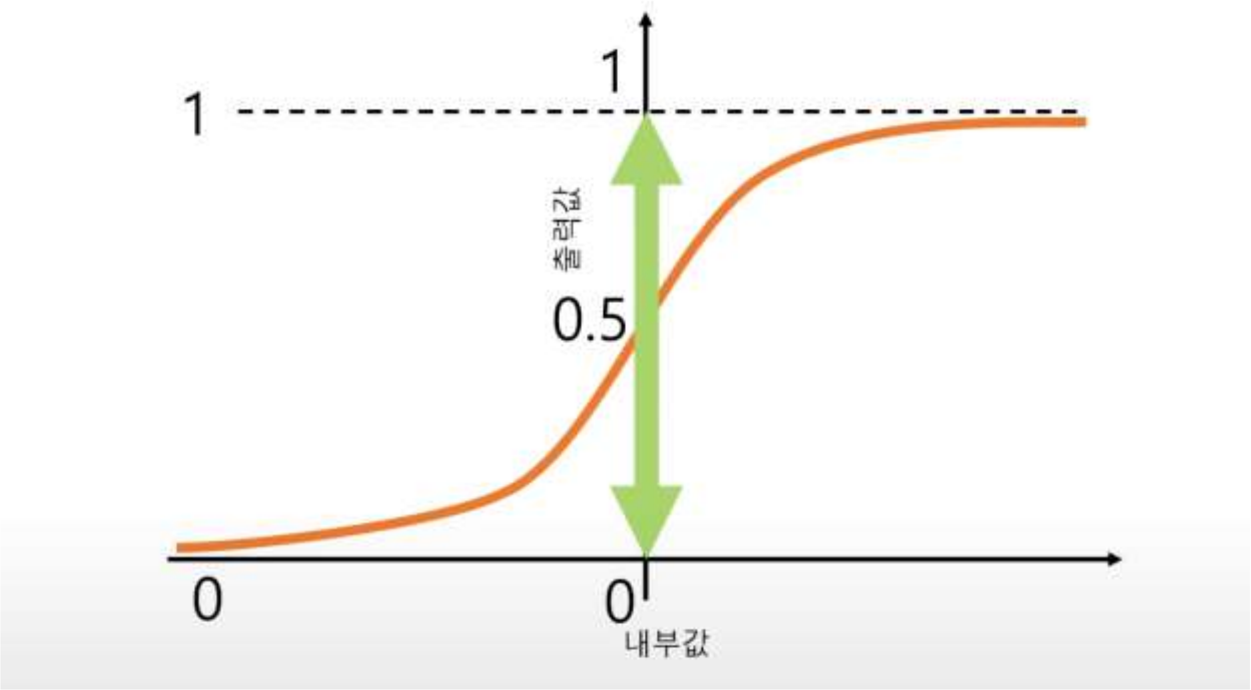

1
0.5
0
$f(x)$
$f'(x)$

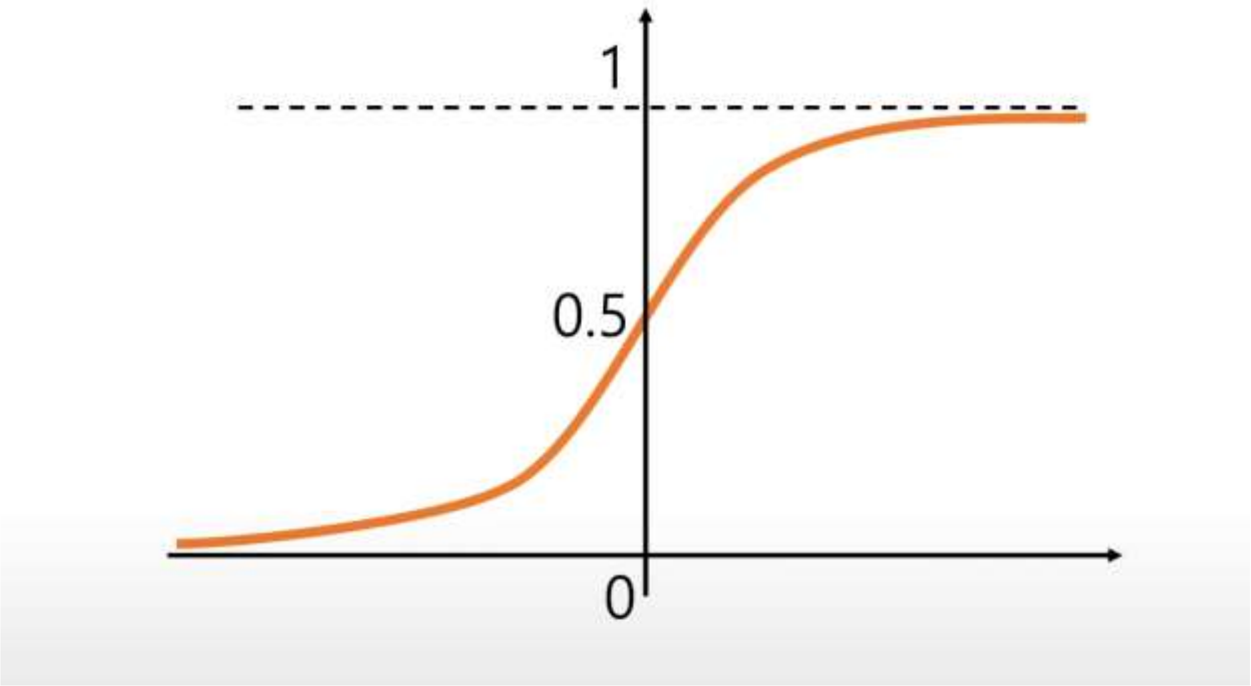


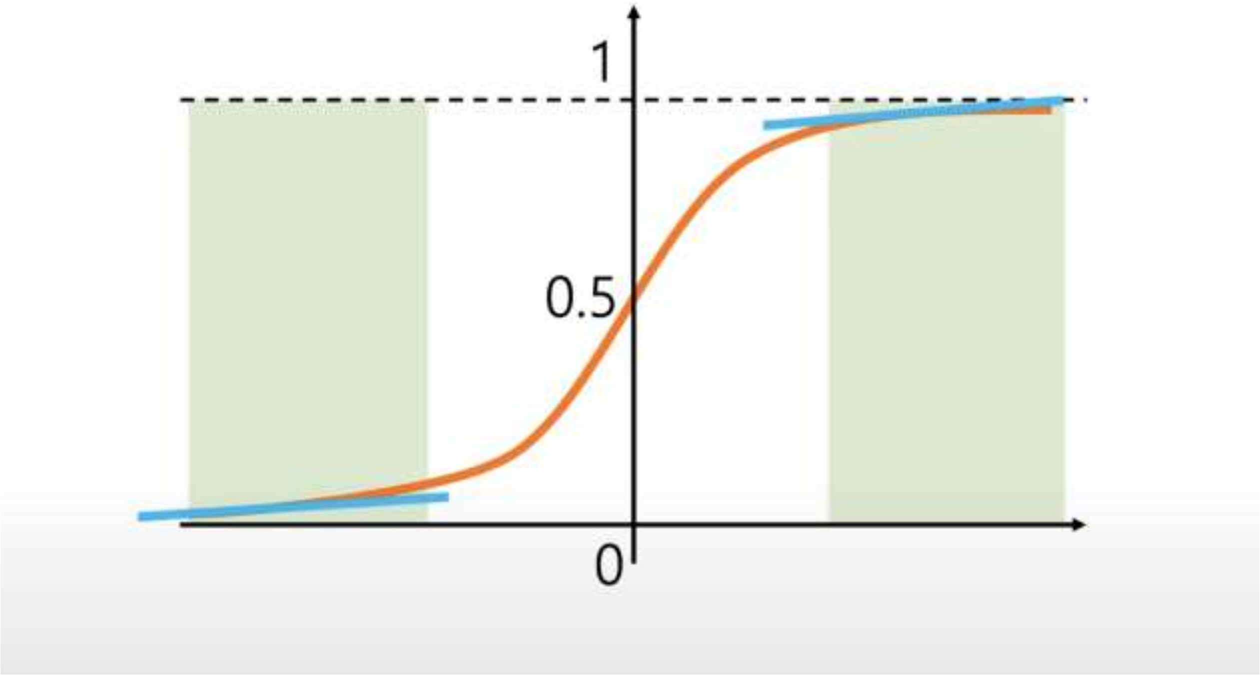



1
0.5
0

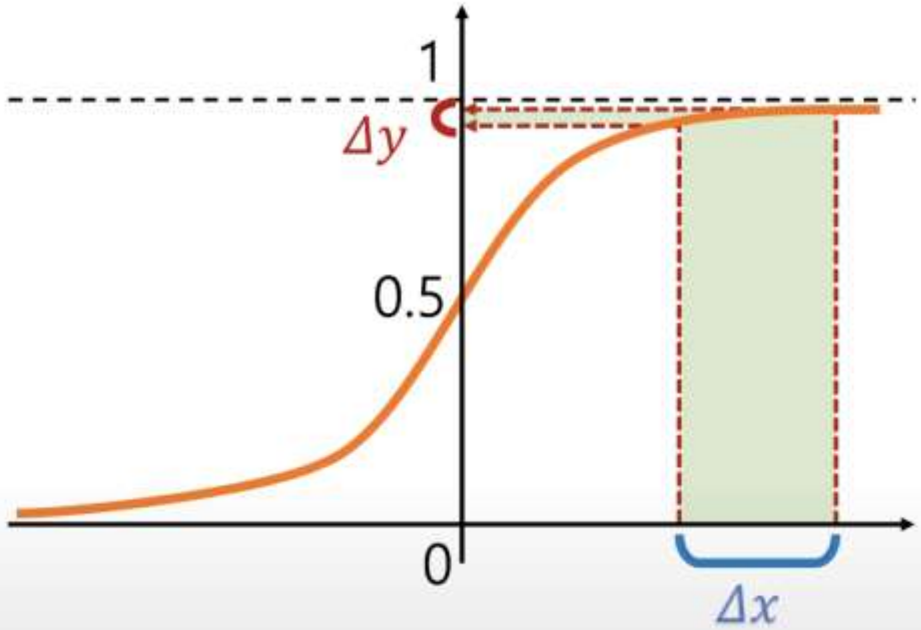

1
0
1
0

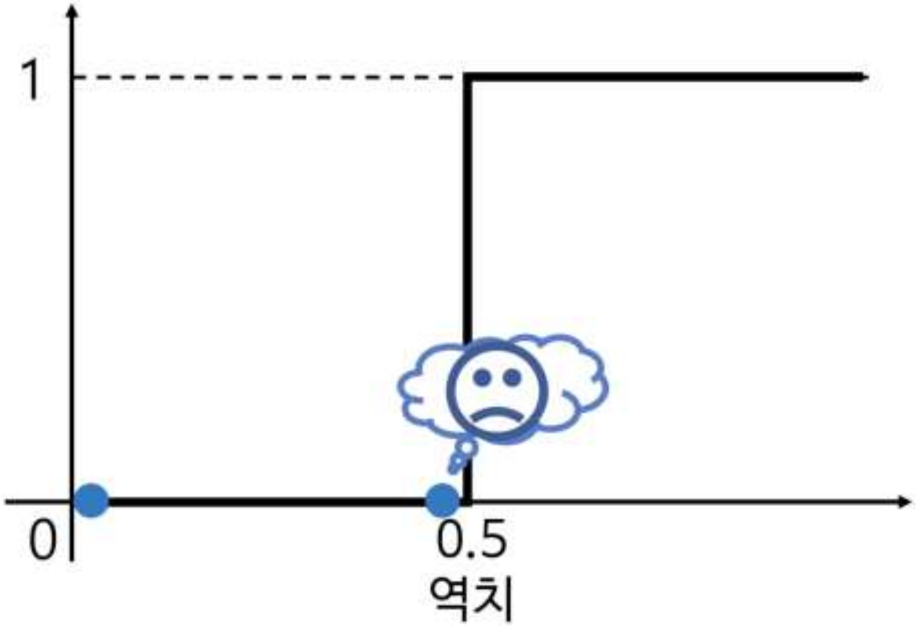

1
1
출력값
0.5
0
0
내부값

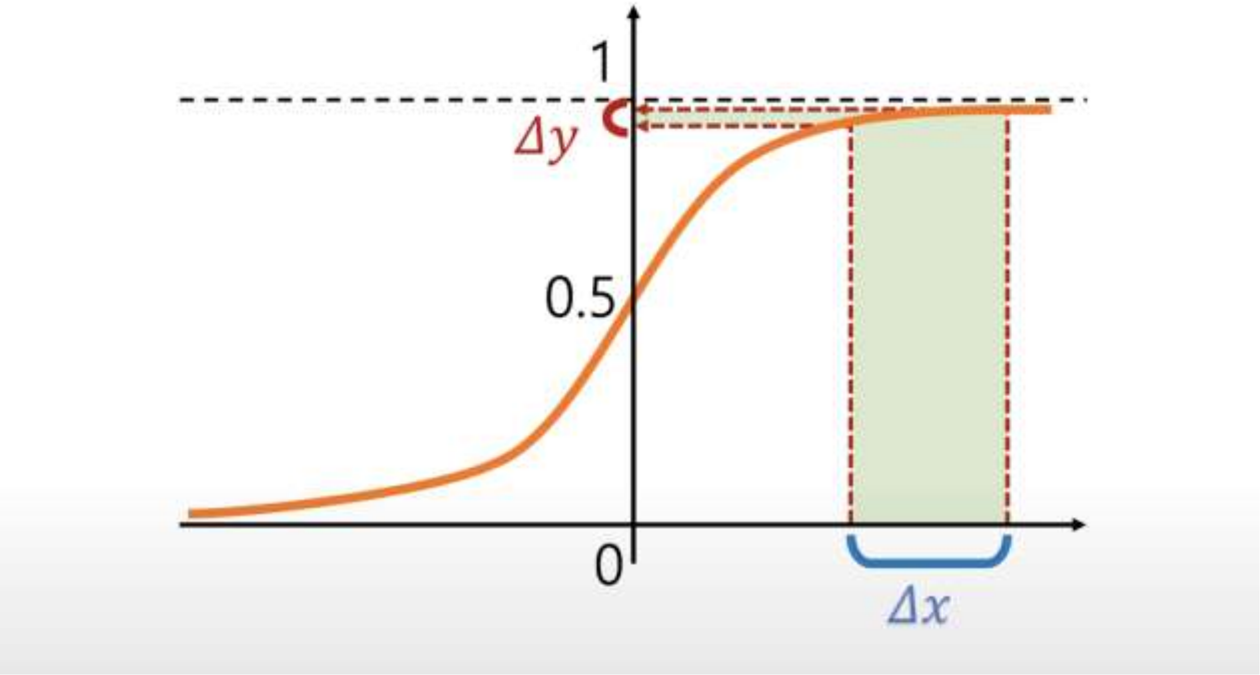

1
0.5
0

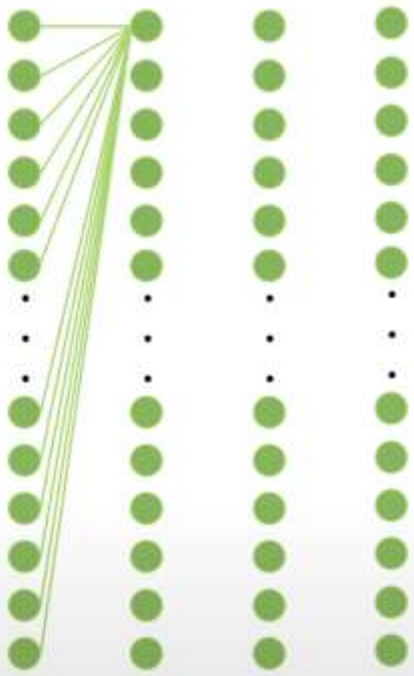


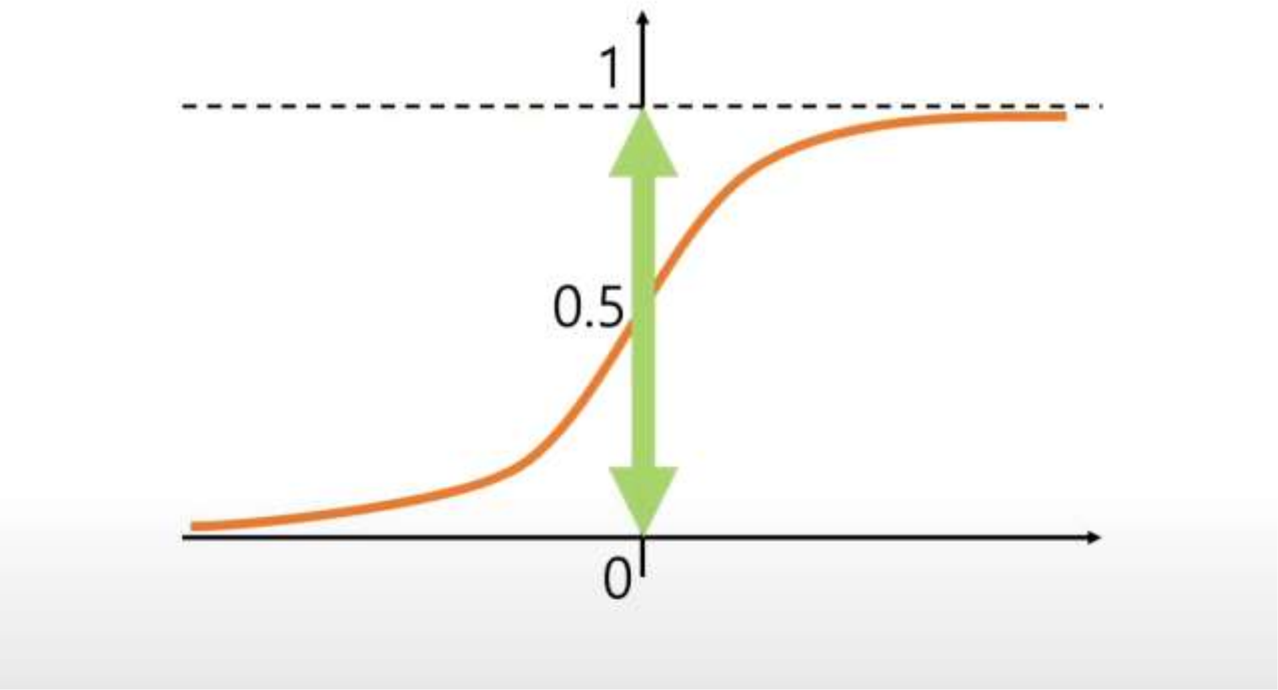
1
0.5
0

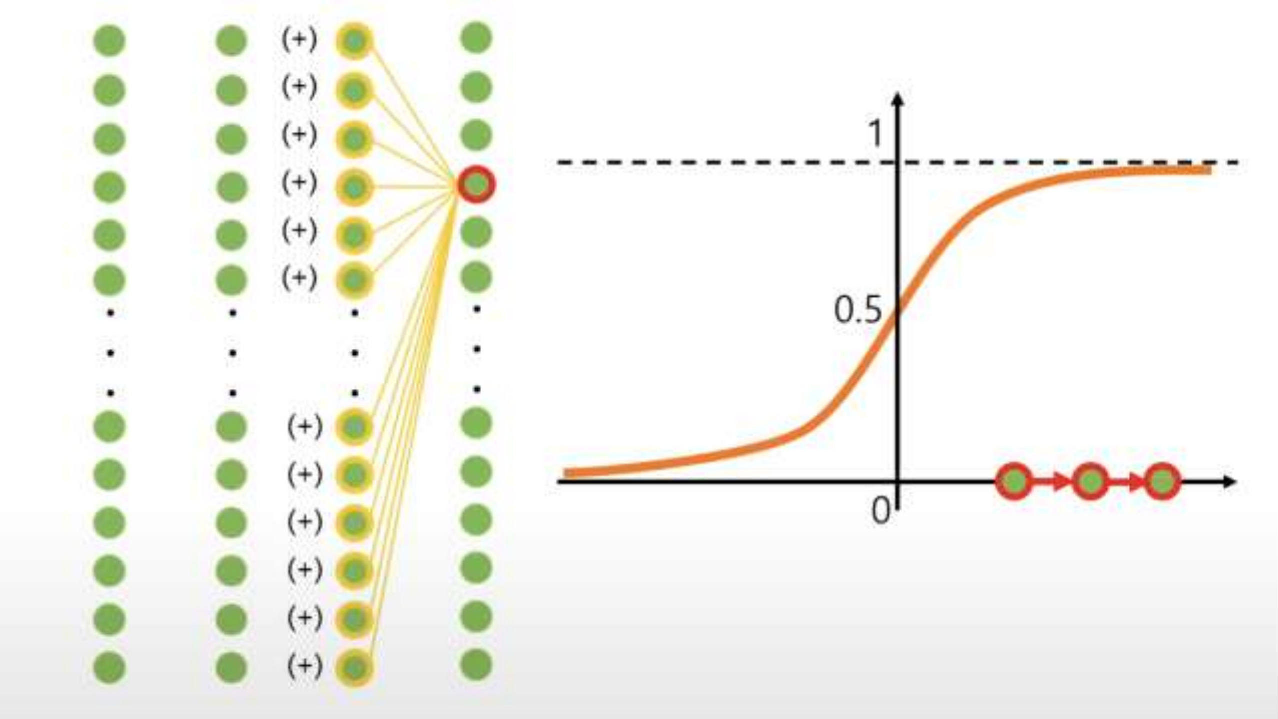

1
0.5
0
Δy
Δx

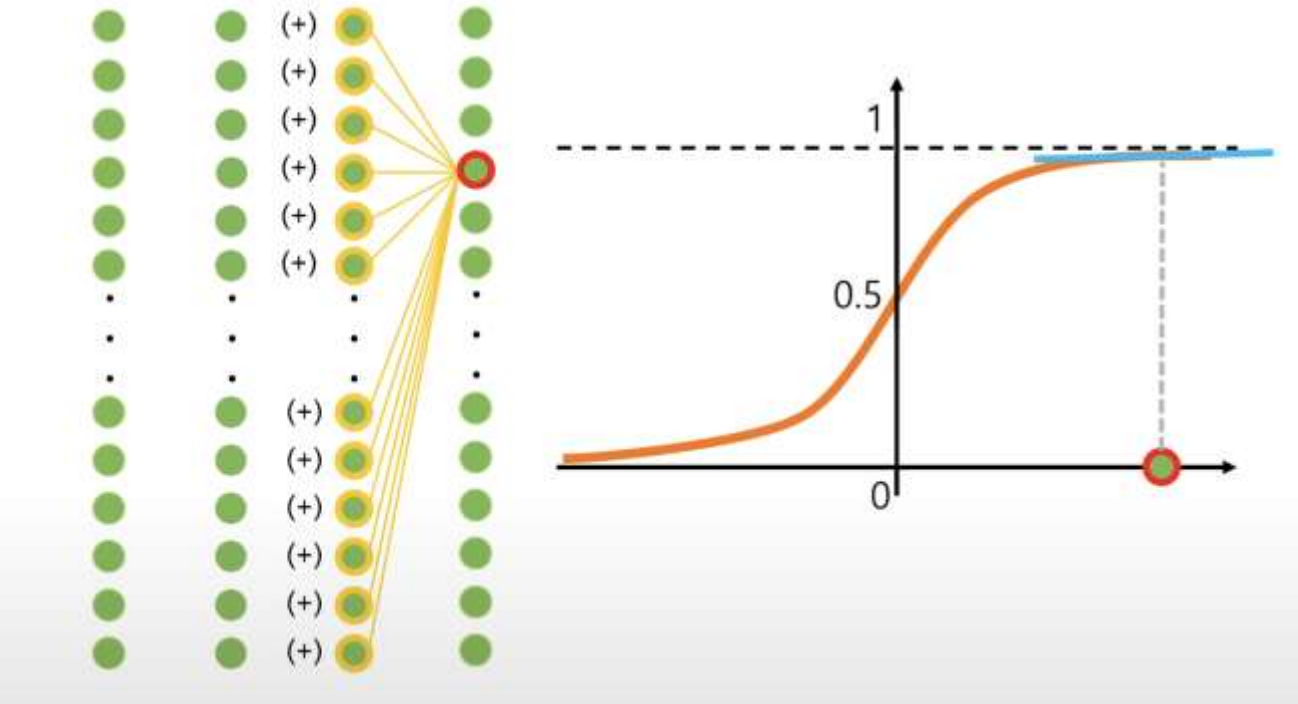

1
0
0.5
역치


1
0.5
0
Δy
Δx

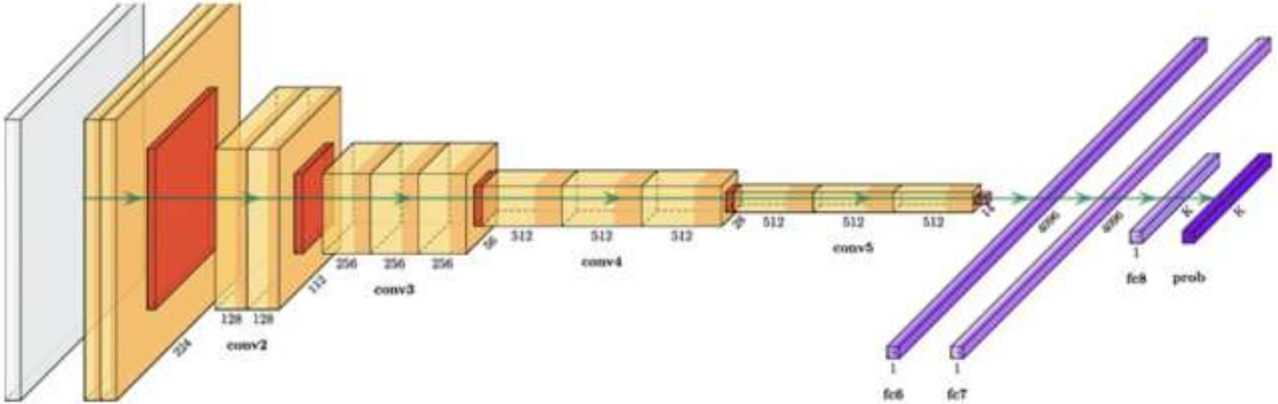



1
0.5
0

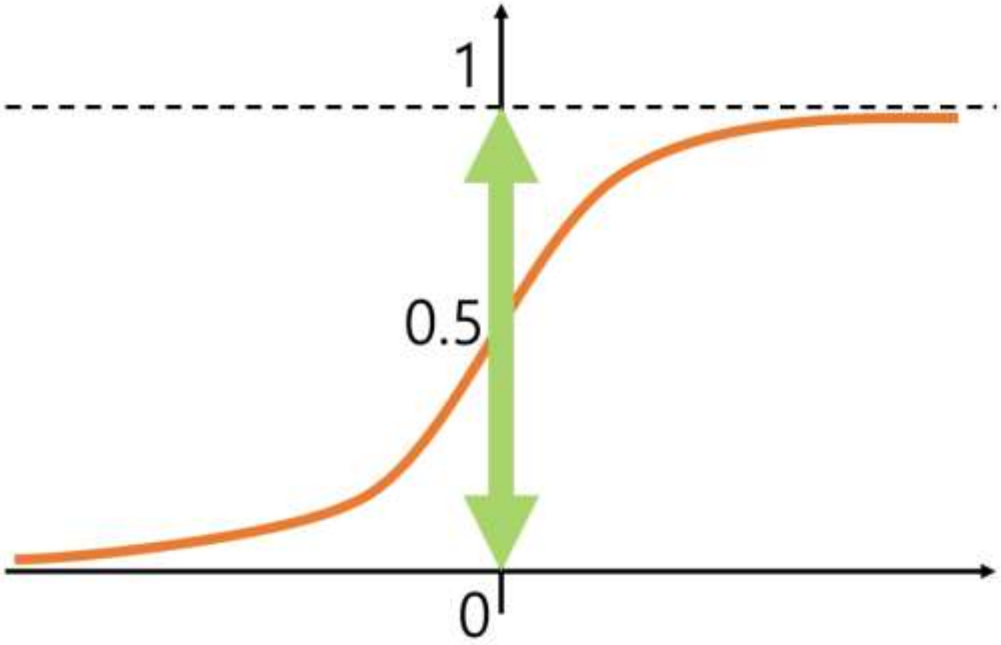

(+)
1
0.5
0



(+)
1
0.5
0



Blauch, Nicholas & Behrmann, Marlene & Plaut, David. (2020). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. Cognition. 208. 104341. 10.1016/j.cognition.2020.104341.


1
0.5
0

퍼셉트론의 학습방법:

새 연결강도 = 현 연결강도 + 현 입력값 x 오차 x 학습률

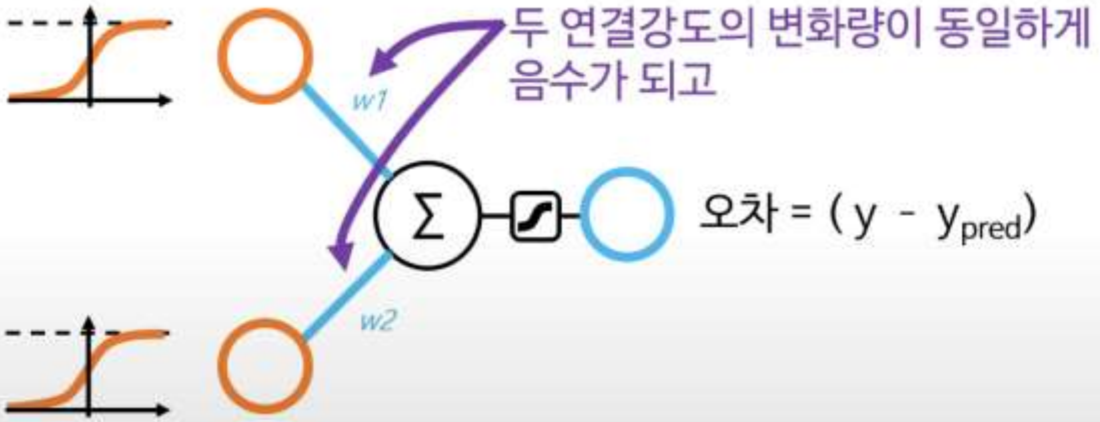
다층신경망의 학습방법:

새 연결강도의 변화량 $\propto$ [현 입력값 x 오차]

다층신경망의 학습방법:

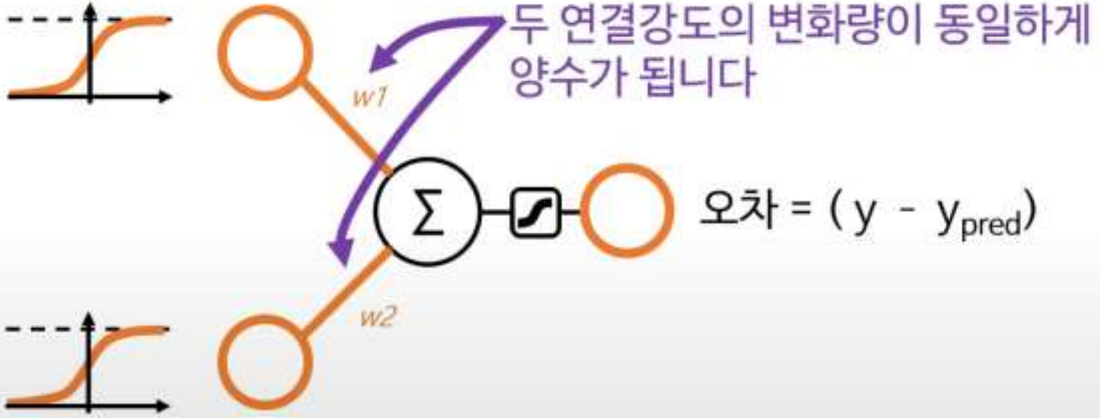
(−) (+) (−)

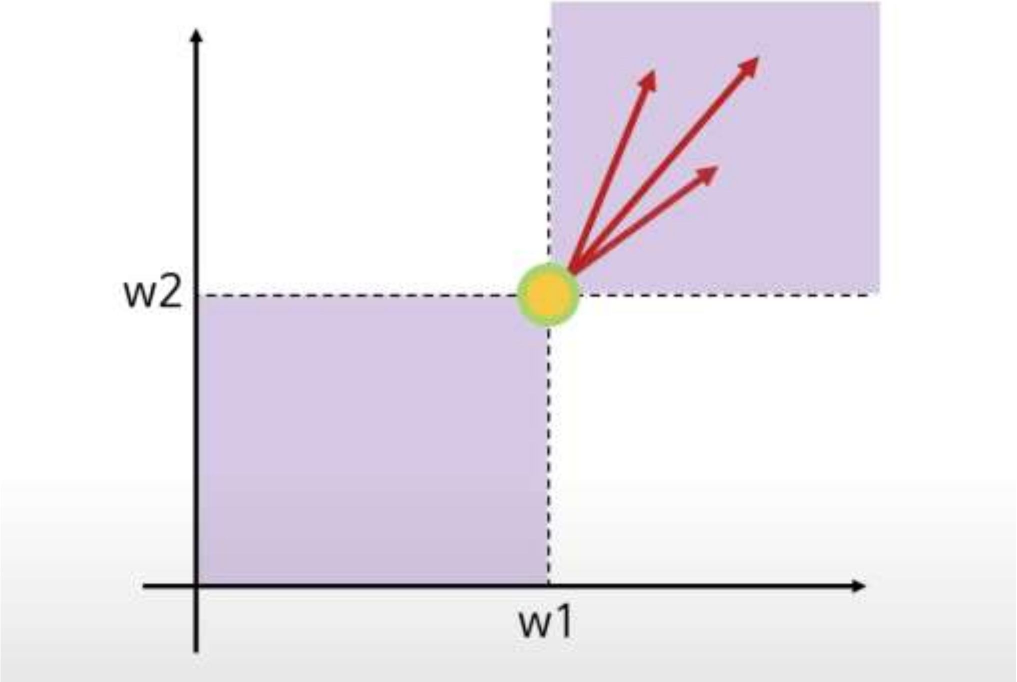
새 연결강도의 변화량 ∝ [현 입력값 x 오차]

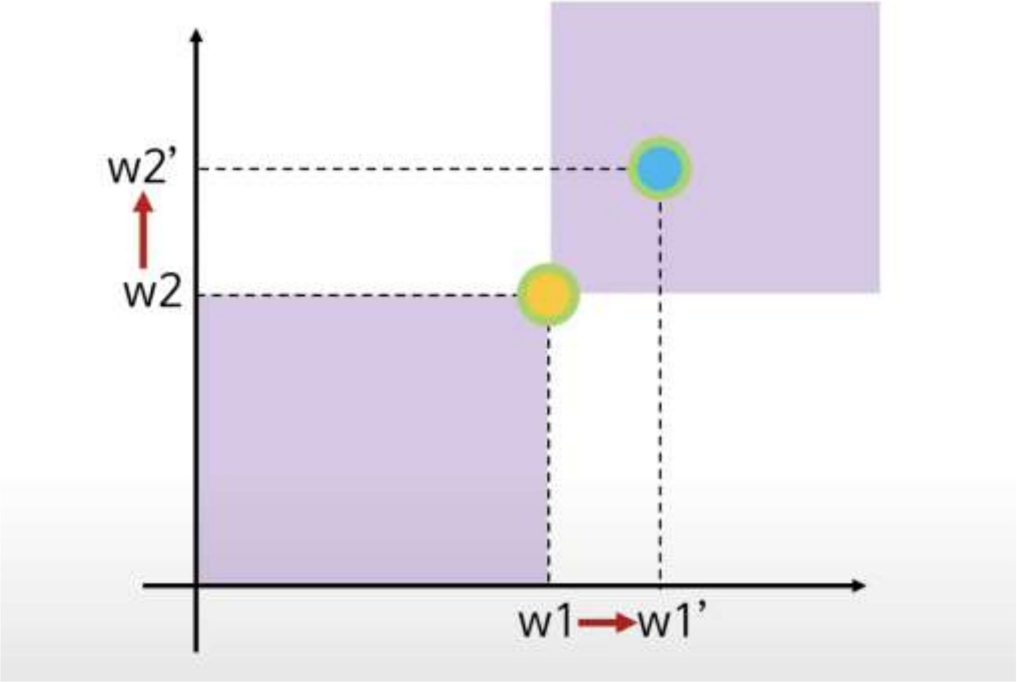


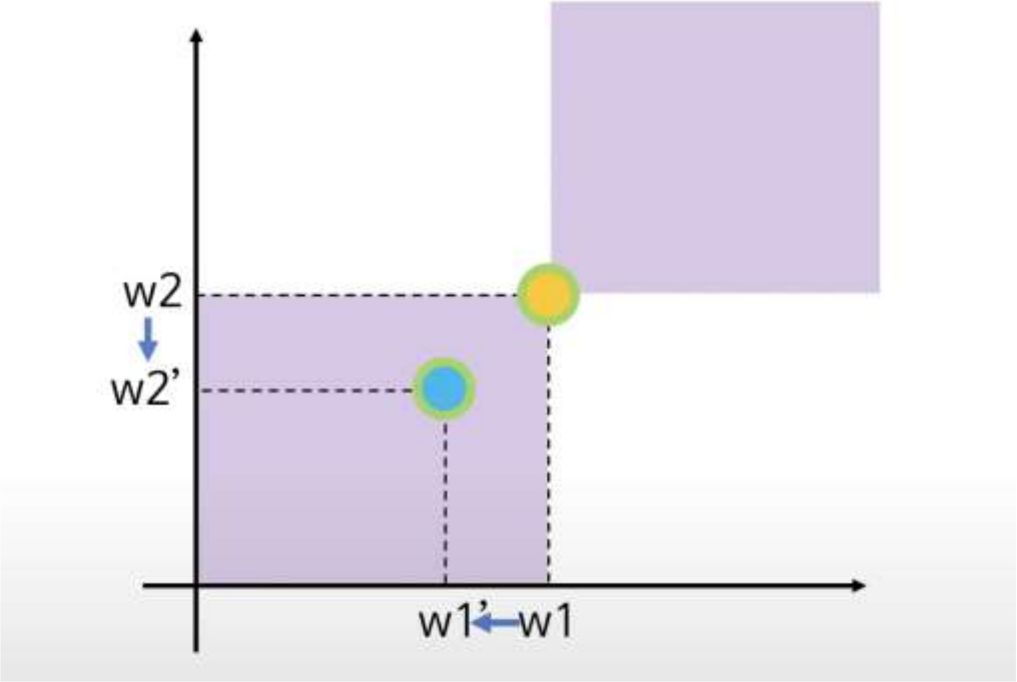
다층신경망의 학습방법:

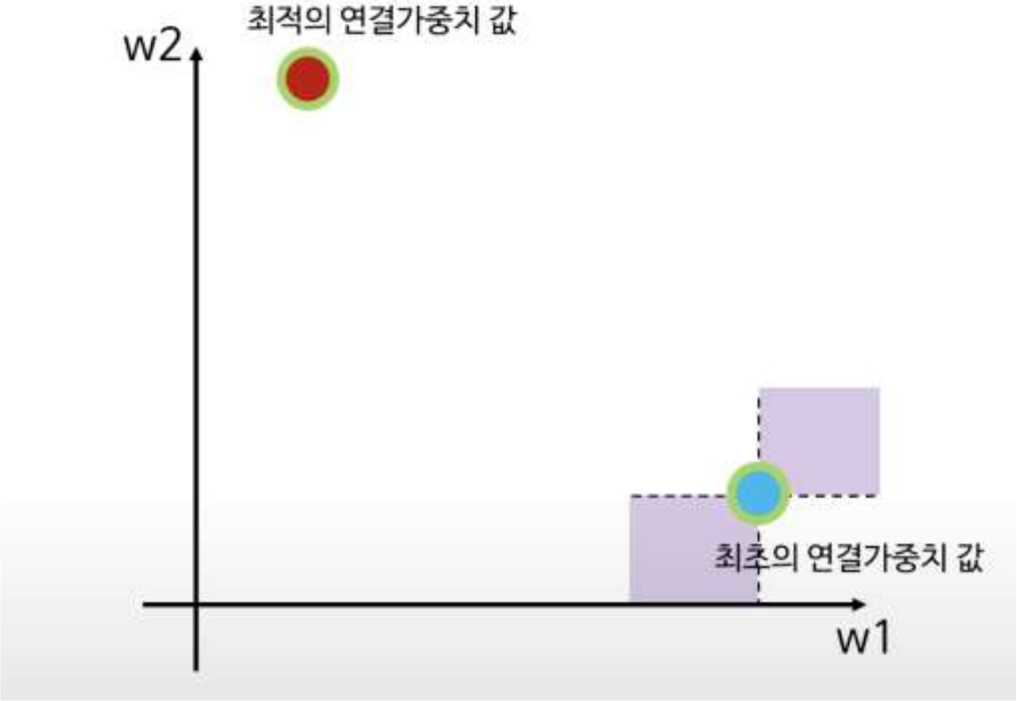
$$\underset{(+)}{\text{새 연결강도의 변화량}} \propto [\underset{(+)}{\text{현 입력값}} \times \underset{(+)}{\text{오차}}]$$

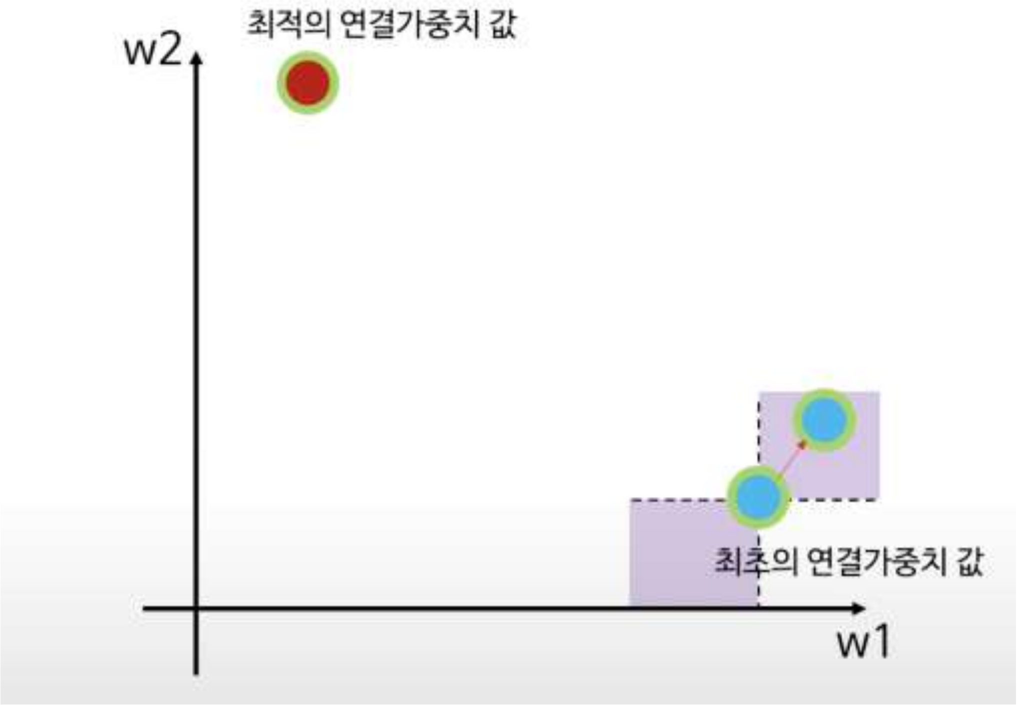



w2
w1

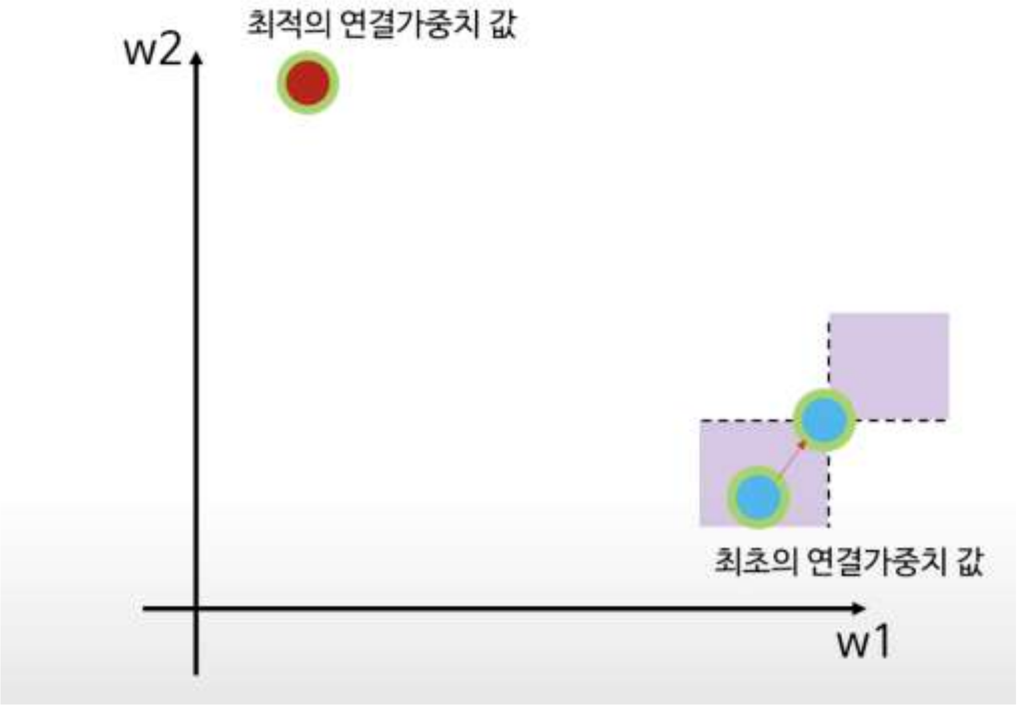

w2'
w2
w1
w1'

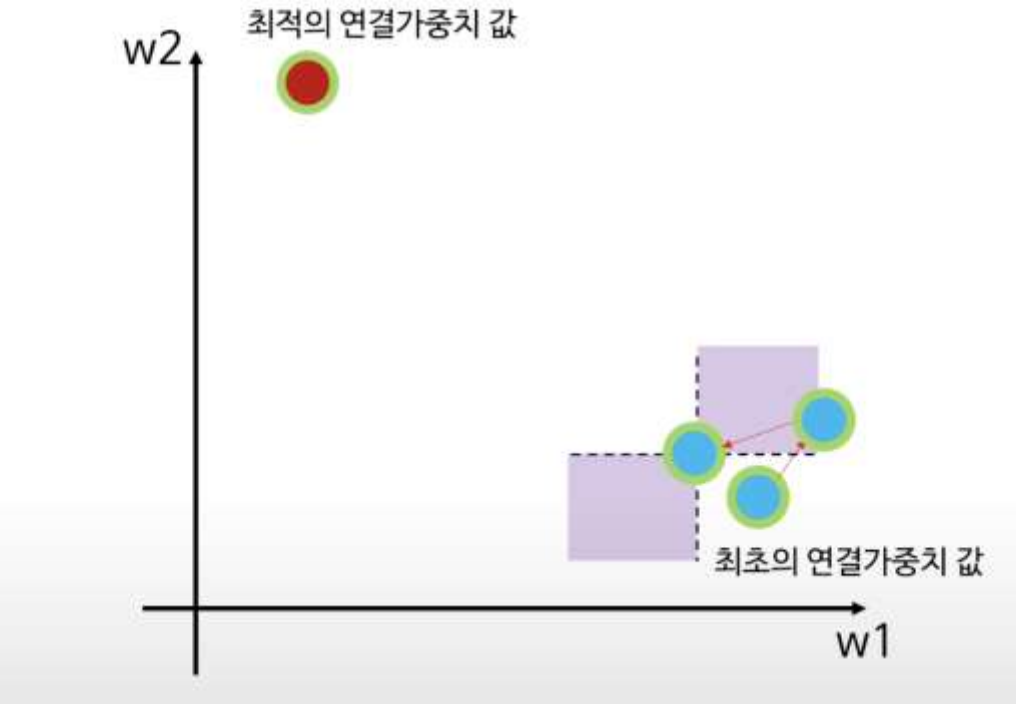

w2
w2’
w1’
w1

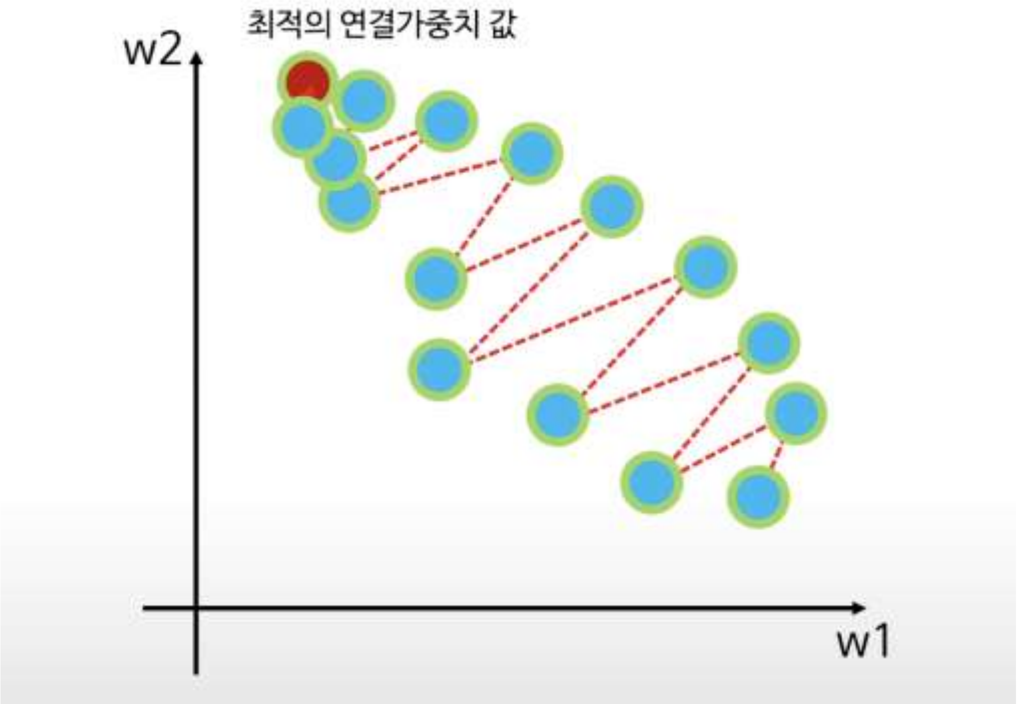

최적의 연결가중치 값
w2
최초의 연결가중치 값
w1

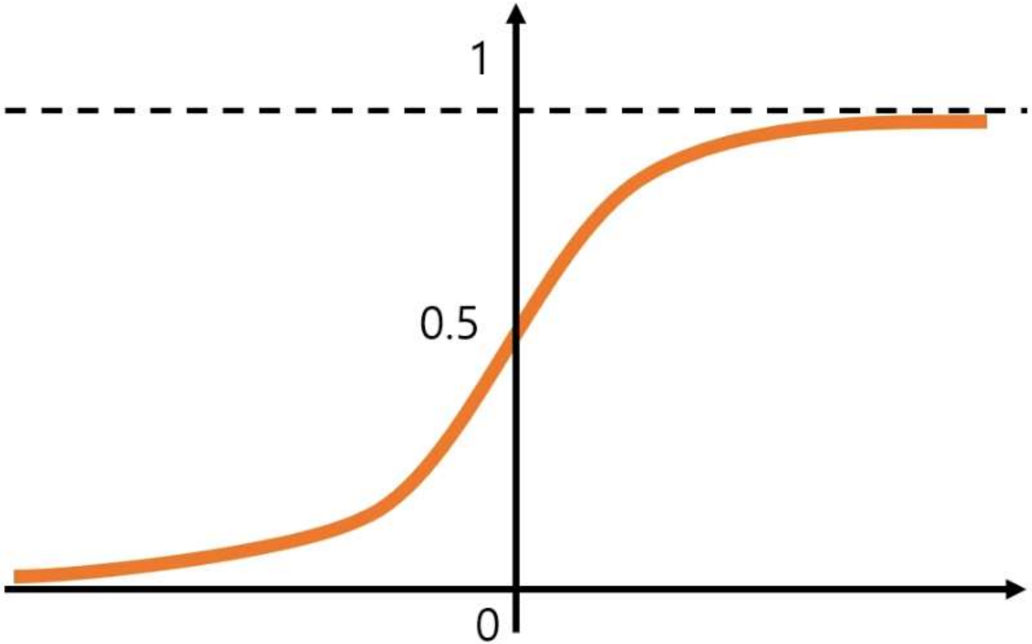


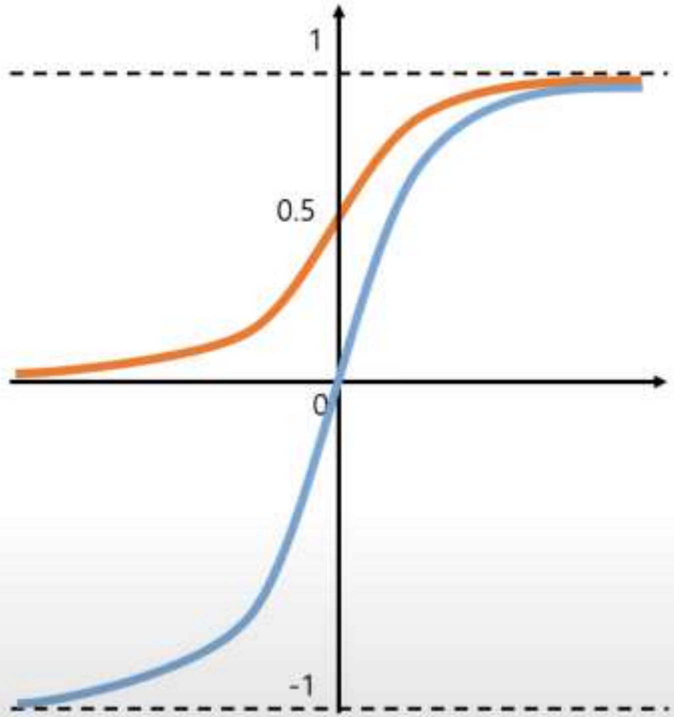
최적의 연결가중치 값
w2
최초의 연결가중치 값
w1


최적의 연결가중치 값
w2
최초의 연결가중치 값
w1

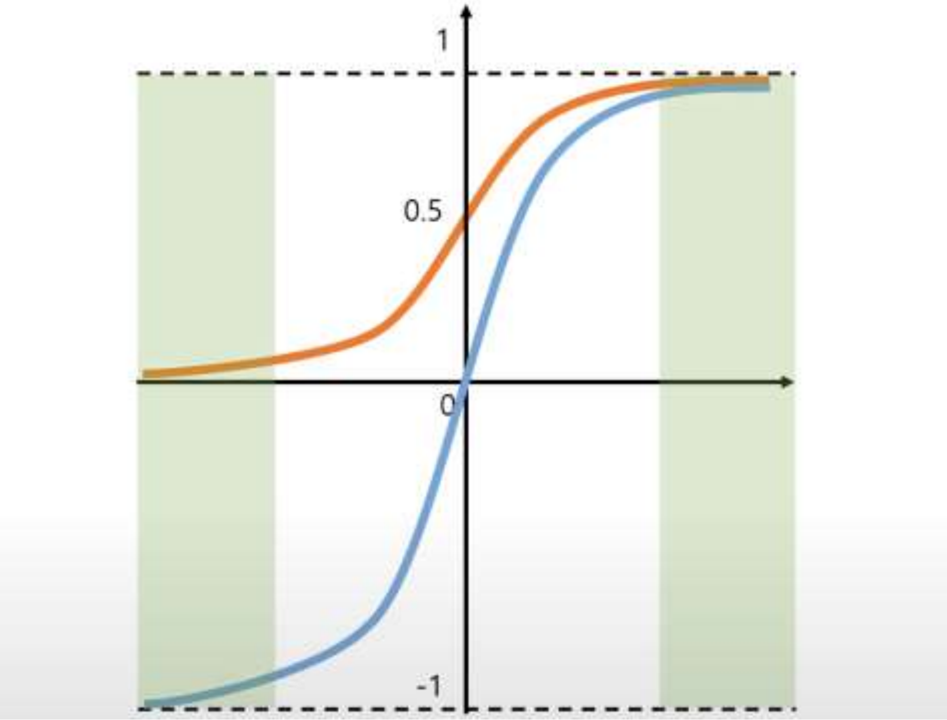

최적의 연결가중치 값
w2
최초의 연결가중치 값
w1

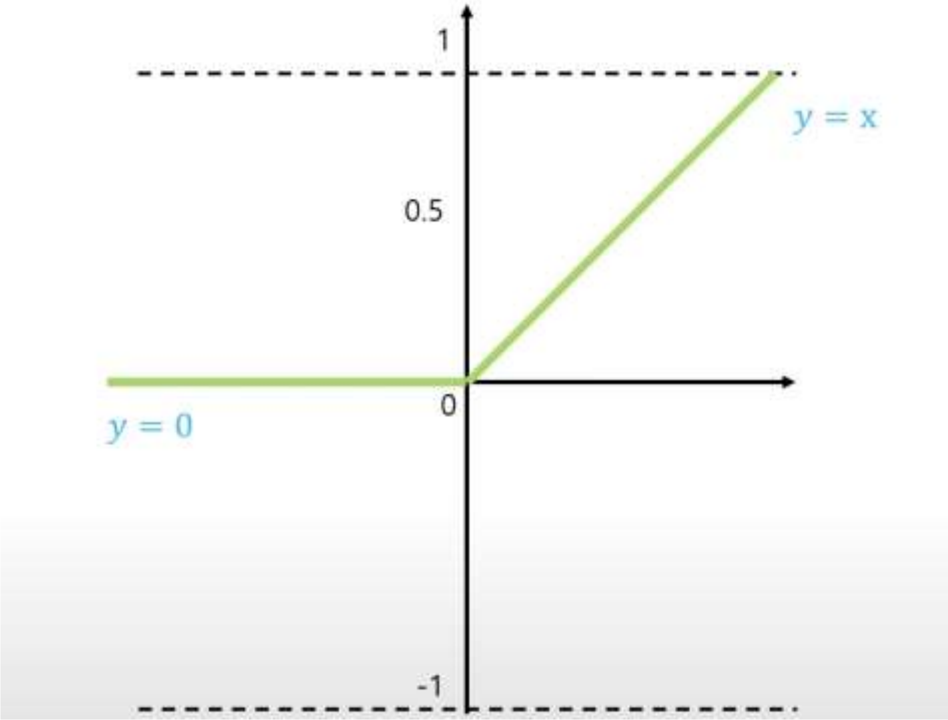

최적의 연결가중치 값
w2
w1

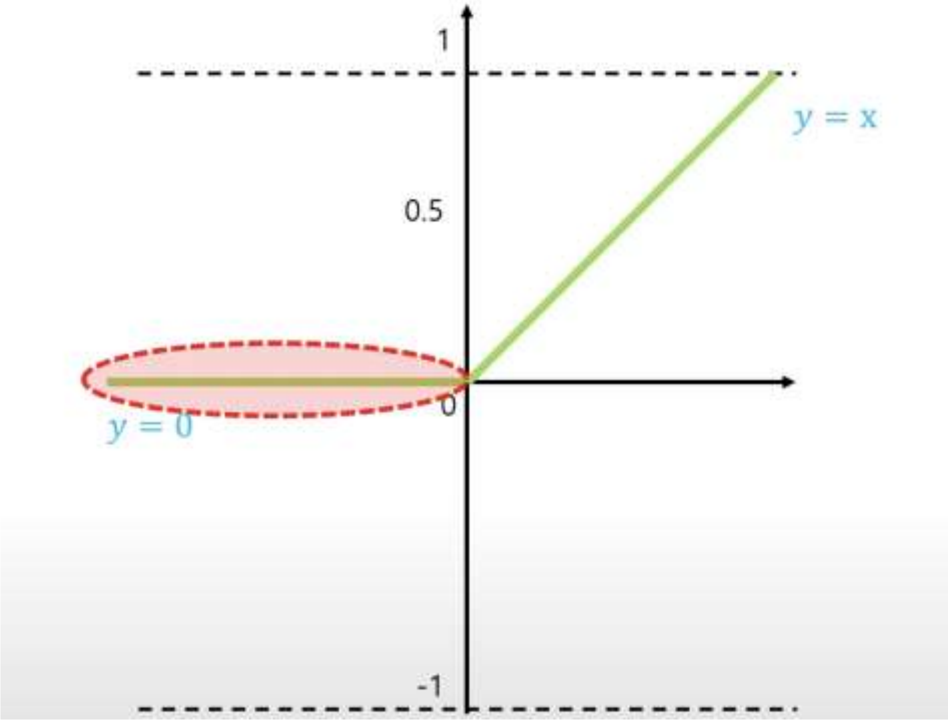

1
0.5
0

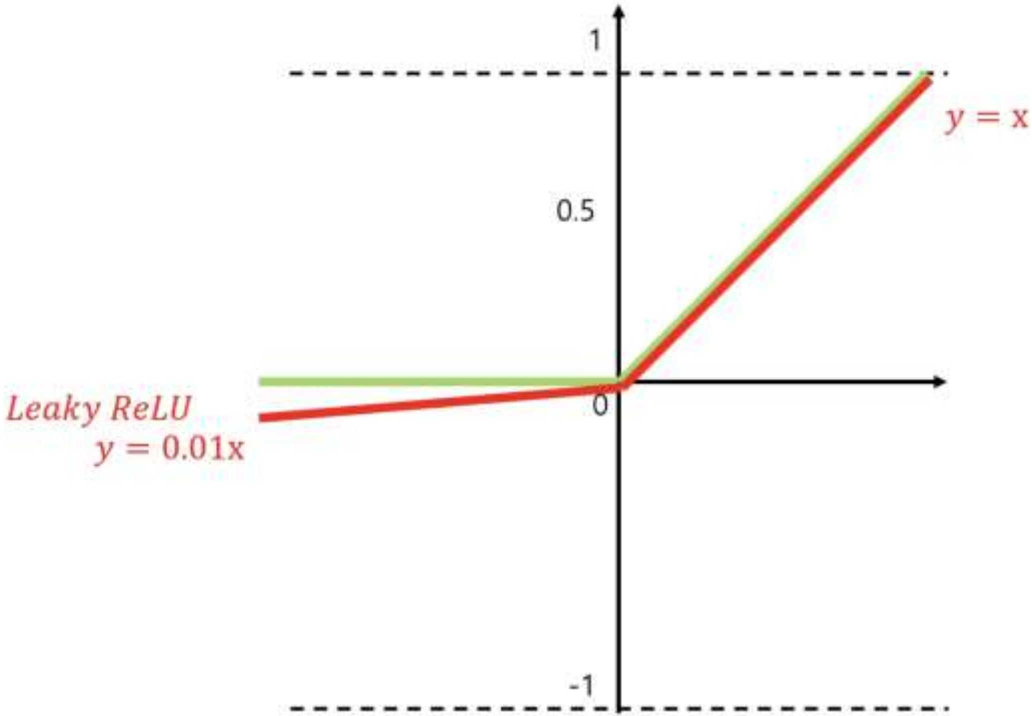


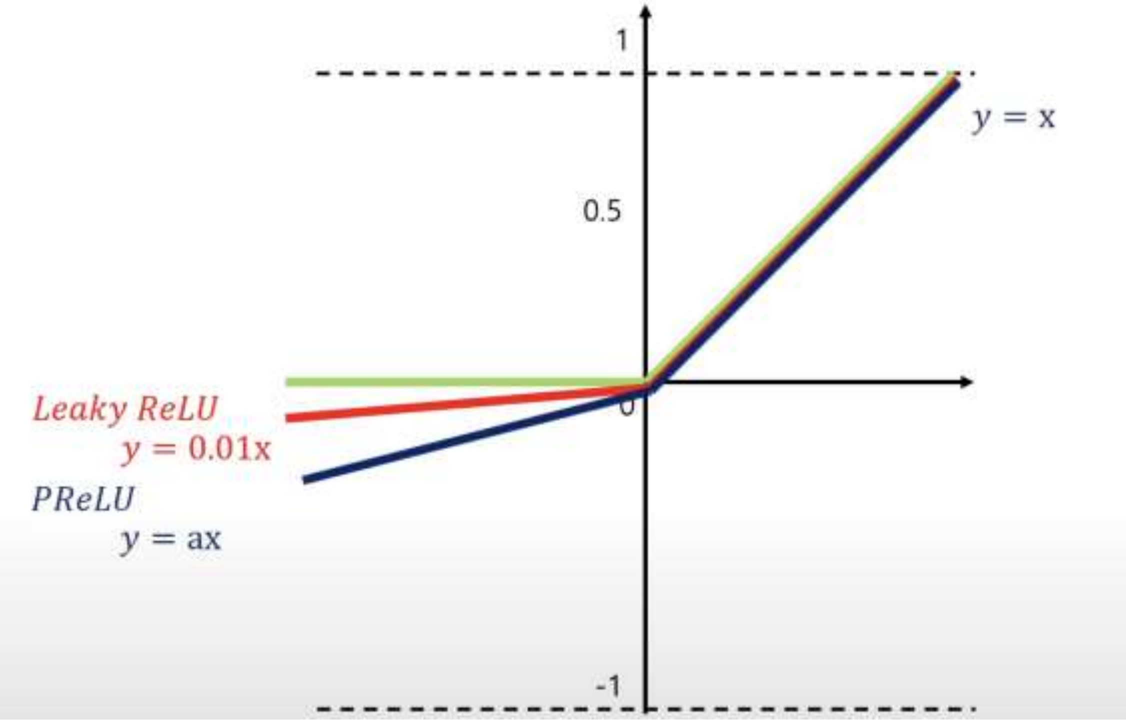
1
0.5
0
-1


1
0.5
0
-1

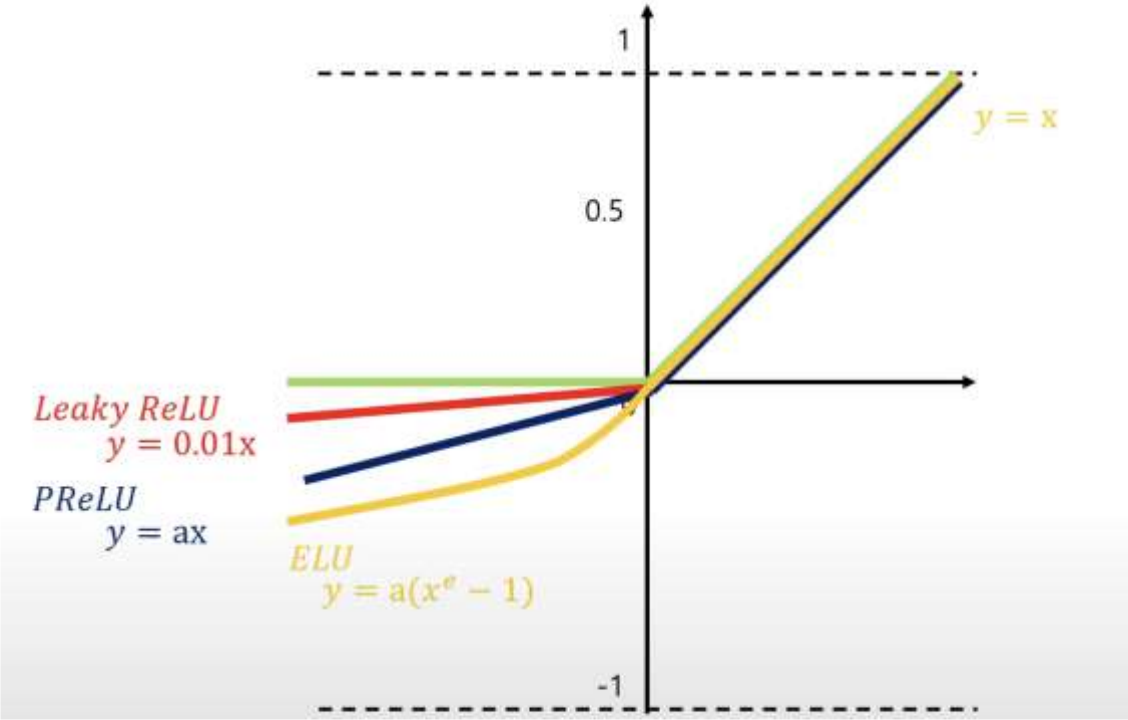

1
0.5
0
-1
y = x
y = 0


1
0.5
0
-1
y = x
y = 0

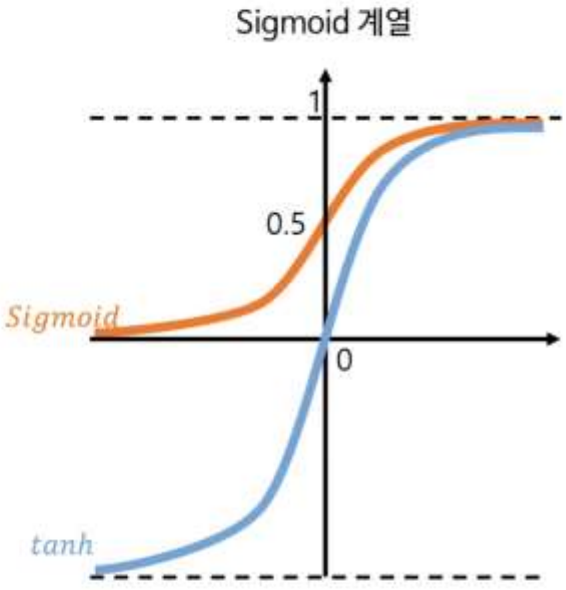

1
0.5
0
-1
y = x
Leaky ReLU
y = 0.01x

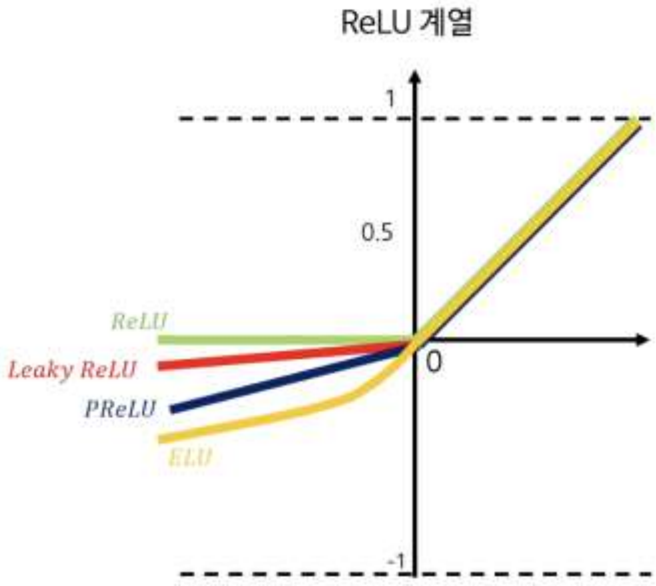


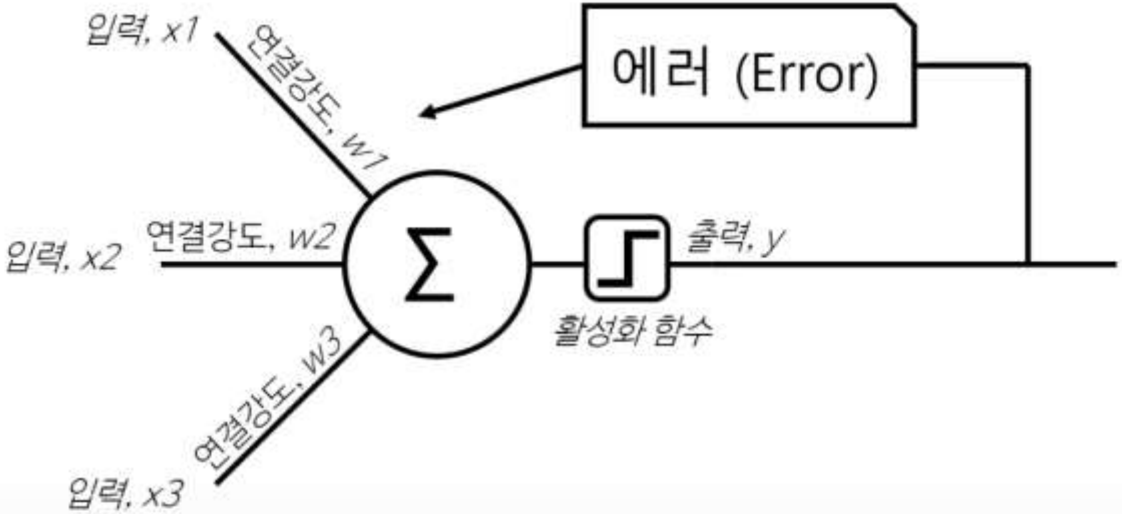
1
0.5
0
-1
y = x
Leaky ReLU
y = 0.01x
PReLU
y = ax

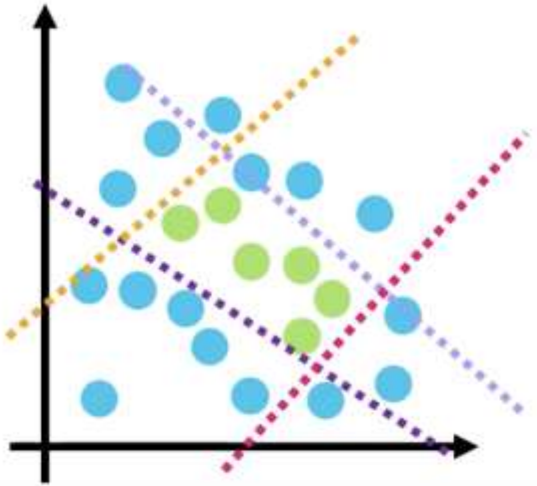


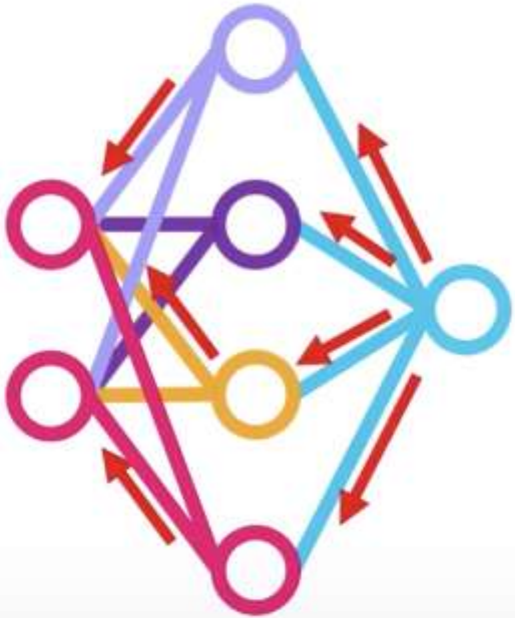
1
0.5
-1
y = x
Leaky ReLU
y = 0.01x
PReLU
y = ax
ELU
y = a(x^e − 1)

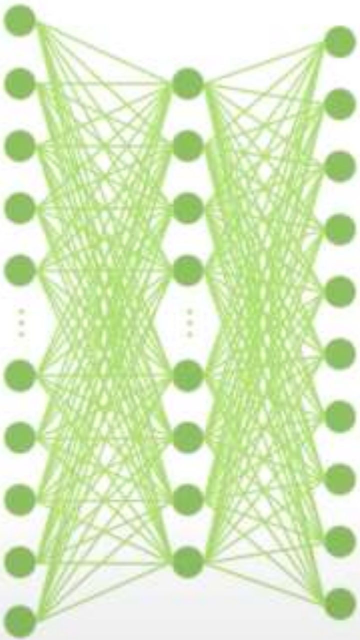

Sigmoid 계열
1
0.5
Sigmoid
0
tanh


ReLU 계열
1
0.5
ReLU
Leaky ReLU
PReLU
ELU
0
-1

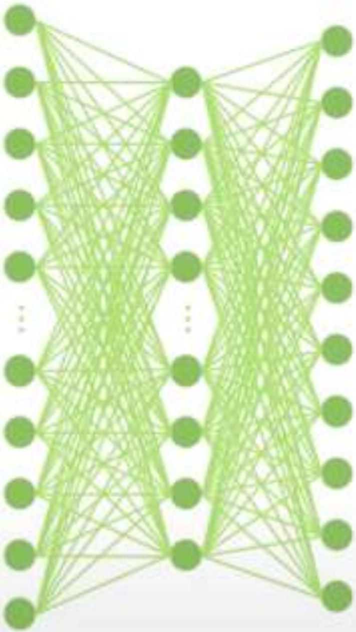

입력, x1
연결강도, w1
에러 (Error)
입력, x2
연결강도, w2
Σ
출력, y
활성화 함수
연결강도, w3
입력, x3

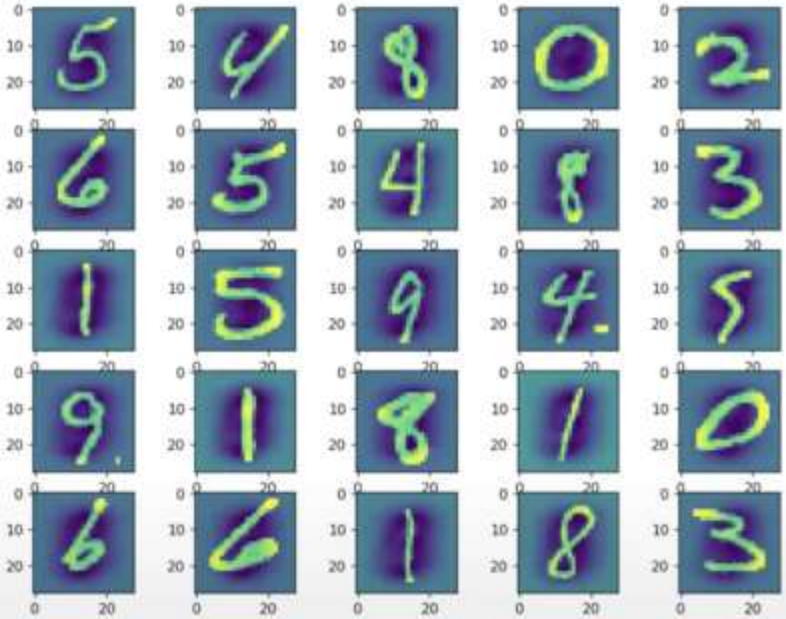


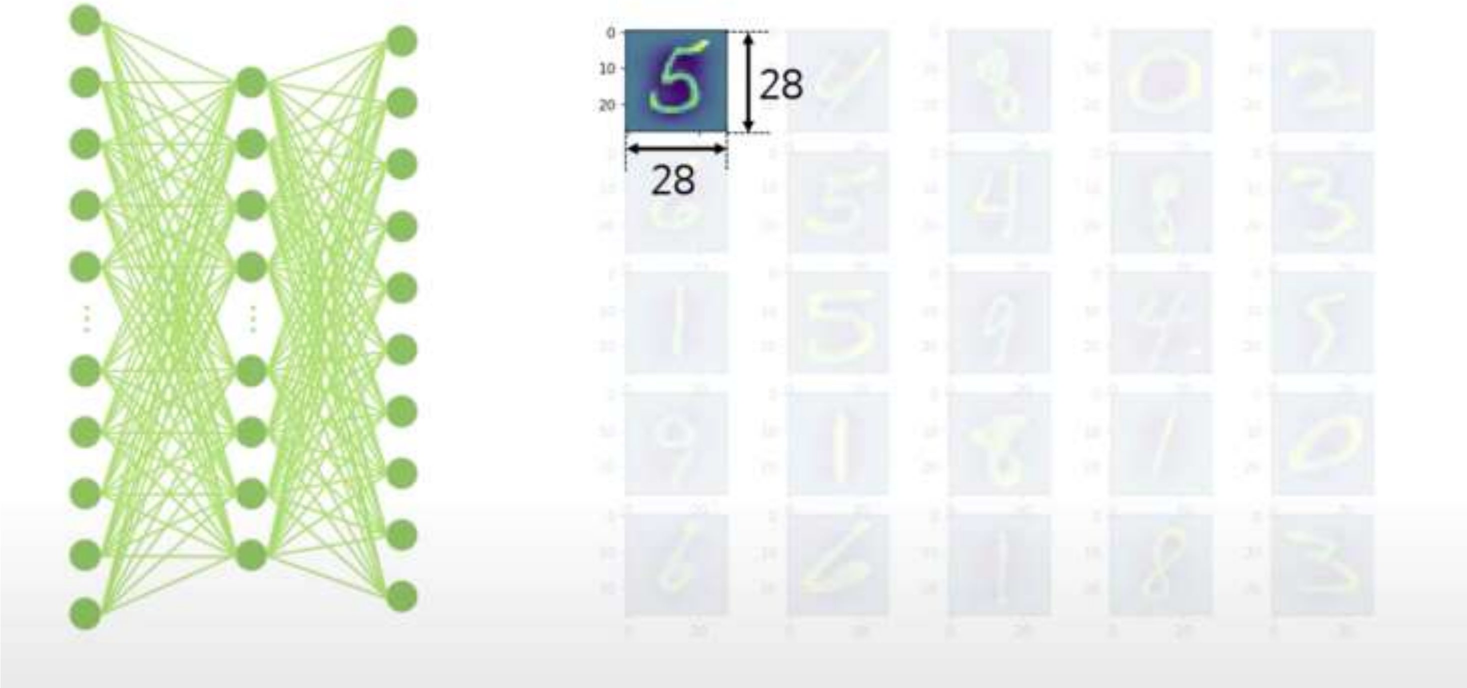


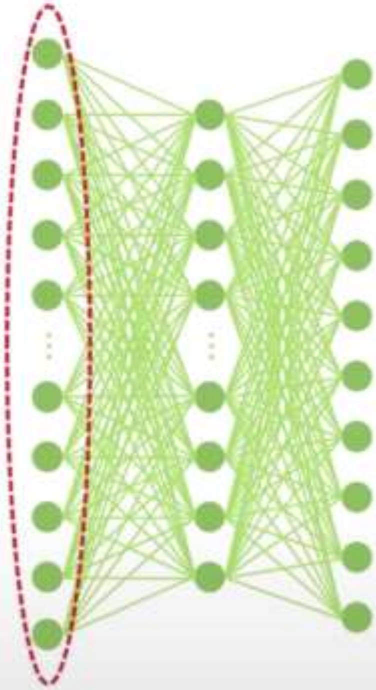


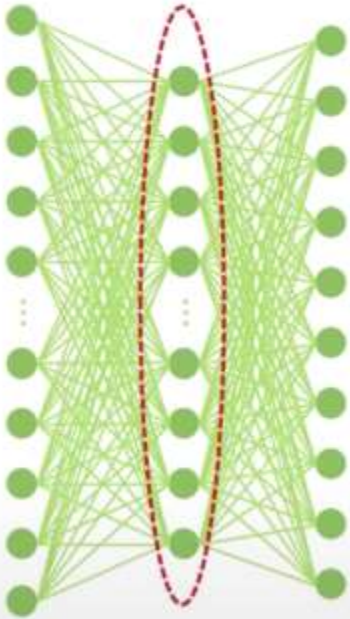
MNIST database. (2023, May 24). In *Wikipedia*.
https://en.wikipedia.org/wiki/MNIST_database

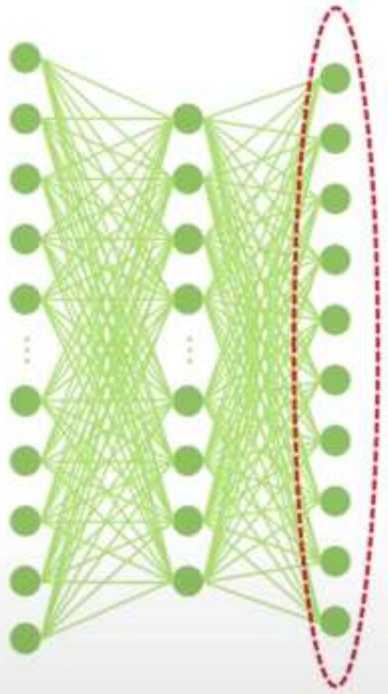


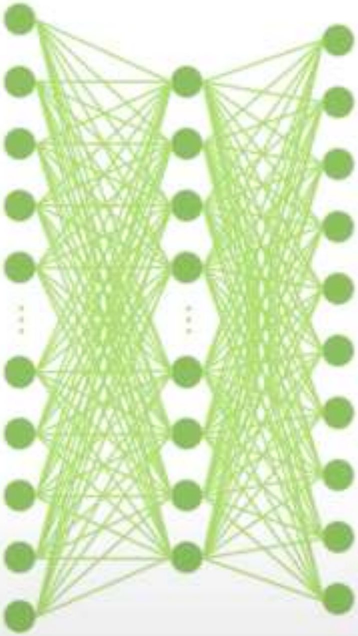



0
10
20
28
28

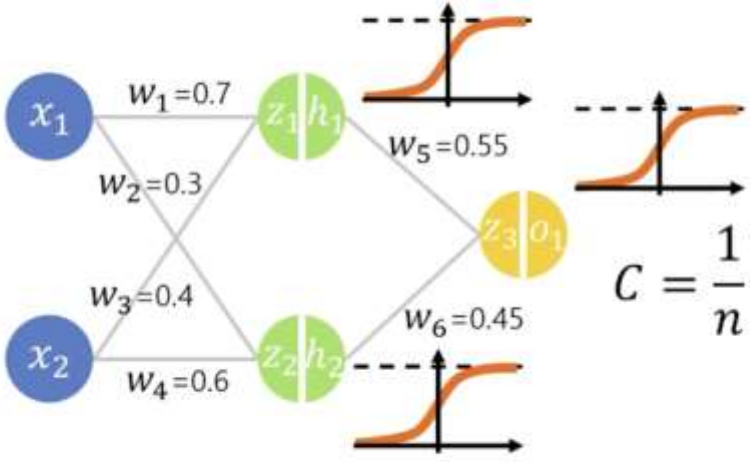






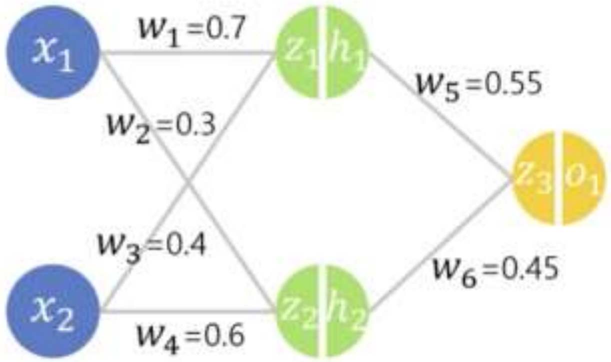


- 연결가중치 개수:
  784x100 + 100x10 = 79,400
- 편향 개수:
  100+10 = 110
- 한 개의 이미지 손실계산:
  79,400번의 연산량 (가중치만 고려)
- 한 개의 파라미터를 업데이트 하기 위해서는:
  4,764,000,000(=79,400x60,000)
- 모든 파라미터들에 대해 각각 연산하여야 하므로
  4,767,000,000 x 79,511(=79,400+110+1) = 378,790,404,000,000
- 378,790,404,000,000 / 850,000,000 = 445,635초 = 123시간



학습률=0.1

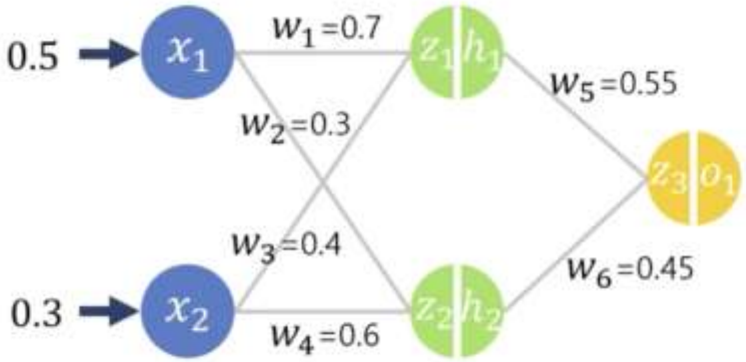
$$C = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2$$

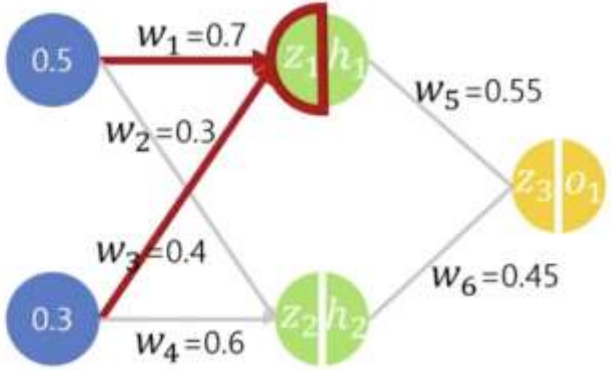


1단계: feedforward 순전파

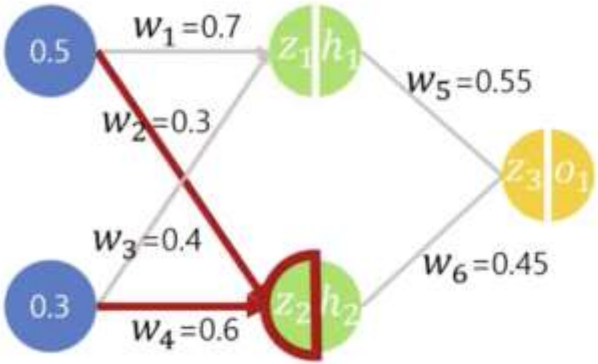
2단계: 손실계산

3단계: backpropagation 역전파


0.5
0.3
x_1
x_2
w_1 =0.7
w_2=0.3
w_3=0.4
w_4 =0.6
z_1 h_1
z_2 h_2
w_5 =0.55
w_6 =0.45
z_3 o_1

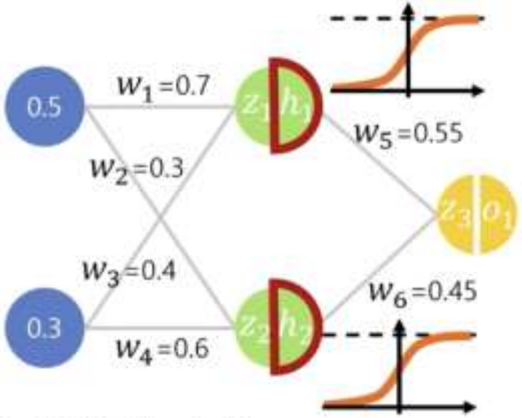


$z_1 = x_1 w_1 + x_2 w_3 = 0.5 \text{x} 0.7 + 0.3 \text{x} 0.4 = \mathbf{0.47}$



$z_1 = x_1 w_1 + x_2 w_3 = 0.5\text{x}0.7 + 0.3\text{x}0.4 = \mathbf{0.47}$

$z_2 = x_1 w_2 + x_2 w_4 = 0.5\text{x}0.3 + 0.3\text{x}0.6 = \mathbf{0.33}$

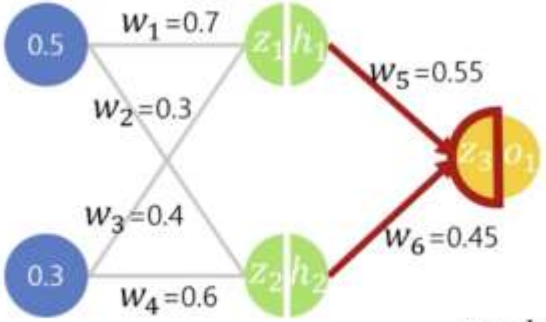


$z_1 = x_1 w_1 + x_2 w_3 = 0.5 \text{x} 0.7 + 0.3 \text{x} 0.4 = 0.47$

$h_1 = sigmoid(z_1) =$ **0.615**

$z_2 = x_1 w_2 + x_2 w_4 = 0.5 \text{x} 0.3 + 0.3 \text{x} 0.6 = 0.33$

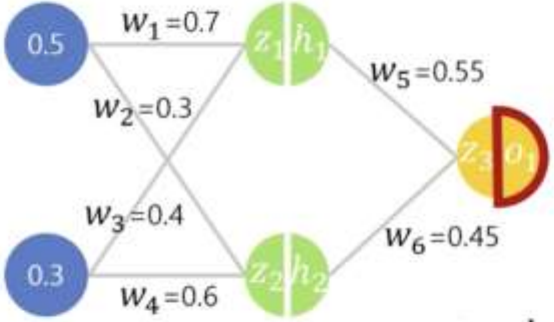
$h_2 = sigmoid(z_2) =$ **0.582**



$$z_3 = h_1 w_5 + h_2 w_6$$
$$= 0.615 \times 0.55 + 0.582 \times 0.45 = \mathbf{0.6}$$

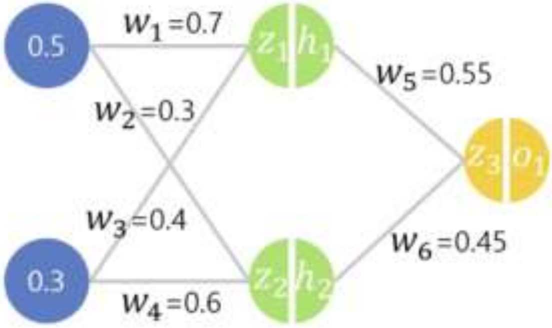
$$h_1 = sigmoid(z_1) = 0.615$$

$$h_2 = sigmoid(z_2) = 0.582$$



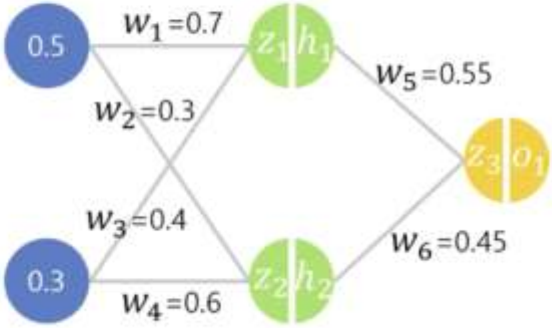
$o_1 = sigmoid(z_3)$=**0.645**

$z_3 = h_1 w_5 + h_2 w_6$
=0.615x0.55+0.582x0.45 =0.6



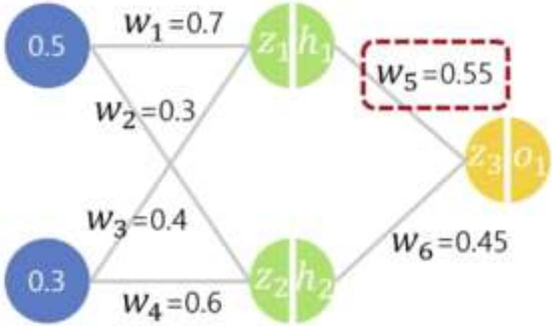
$o_1 = sigmoid(z_3) = \mathbf{0.645}$

$$C = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2$$



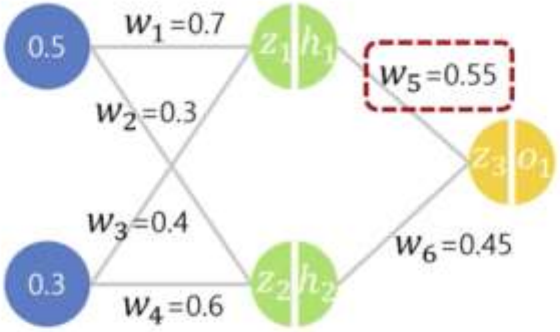
$o_1 = sigmoid(z_3) = 0.645$

$C = \mathbf{0.126}$



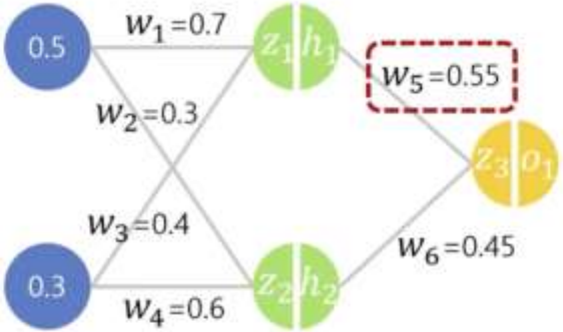
경사하강법을 이용한 가중치 학습법:

새 연결강도 = 현 연결강도 + $-\nabla C$ x 학습률



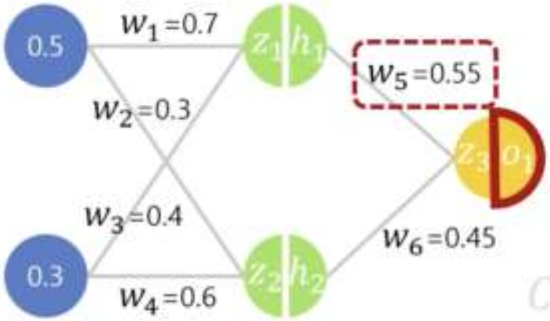
경사하강법을 이용한 가중치 학습법:

새 연결강도 = 현 연결강도 + ~~현 입력값 x 오차~~ $\frac{\partial C}{\partial w_5}$ x 학습률



$$\frac{\partial C}{\partial w_5} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_5}$$

Chain rule



$o_1 = sigmoid(z_3) = \mathbf{0.645}$

$$C = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2$$

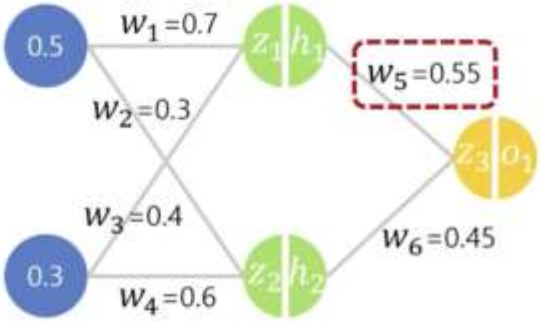
$$C = (y - o_1)^2$$

$$\frac{\partial C}{\partial w_5} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_5}$$

$$\frac{\partial C}{\partial o_1} = -2(y - o_1)^{2-1}$$

$$\frac{\partial C}{\partial o_1} = -2(1 - 0.645)$$

$$= -0.71$$



$o_1 = sigmoid(z_3) = \mathbf{0.645}$

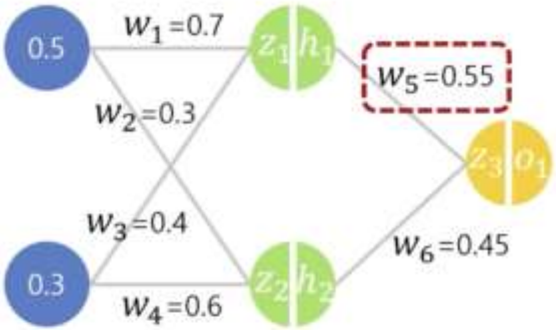
$$\frac{\partial C}{\partial w_5} = -0.71 \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_5}$$

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

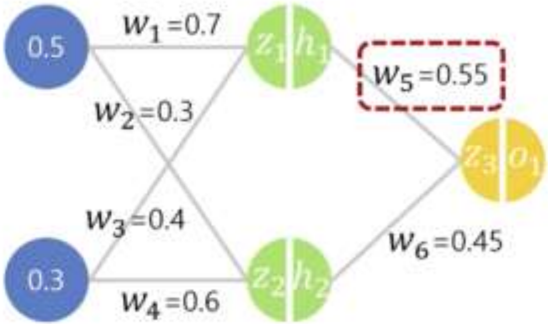
$$O(z) = \frac{1}{1+e^{-z}}$$

$o_1 = sigmoid(z_3) =$ **0.645**

$$\frac{\partial O}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = O(z)(1 - O(z))$$



$$\frac{\partial C}{\partial w_5} = -0.71 \cdot 0.229 \cdot \frac{\partial z_3}{\partial w_5}$$



$$z_3 = h_1 w_5 + h_2 w_6$$

$$\frac{\partial z_3}{\partial w_5} = h_1 \cancel{w_5} + \cancel{h_2 w_6}$$

$$= h_1 = 0.615$$

$$\frac{\partial C}{\partial w_5} = -0.71 \cdot 0.229 \cdot 0.615$$

경사하강법을 이용한 가중치 학습법:

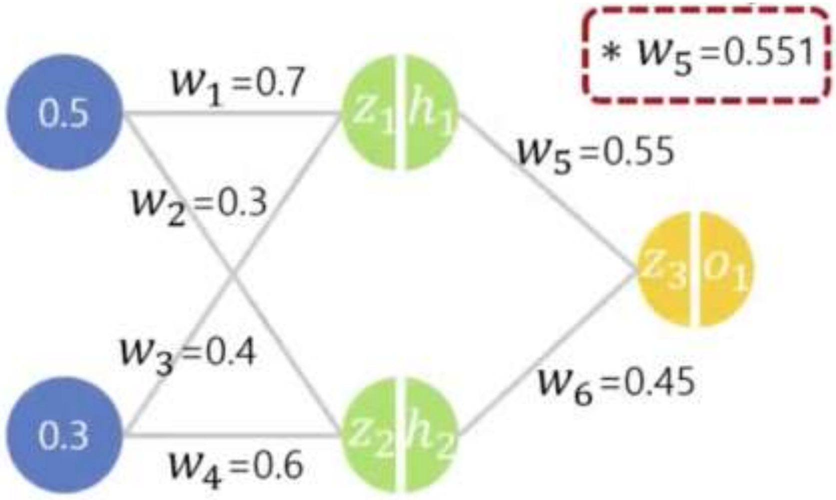
새 연결강도 = 현 연결강도 + $\frac{\partial C}{\partial w_5}$ x 학습률

$$\frac{\partial C}{\partial w_5} = -0.71 \cdot 0.229 \cdot 0.615 = -0.01$$

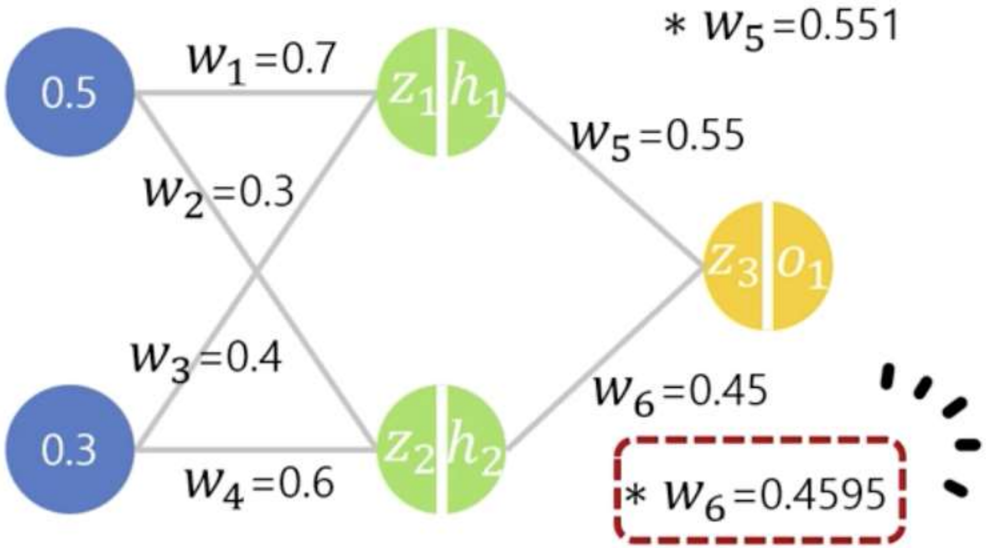
경사하강법을 이용한 가중치 학습법:

새 연결강도 = 현 연결강도 + 0.01 x 학습률

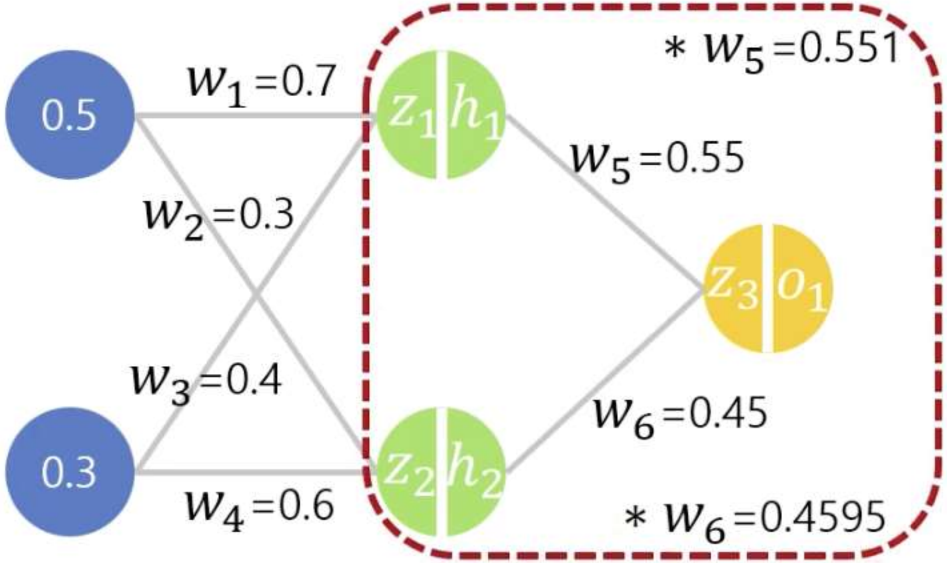
$$\frac{\partial C}{\partial w_5} = -0.71 \cdot 0.229 \cdot 0.615 = -0.01$$



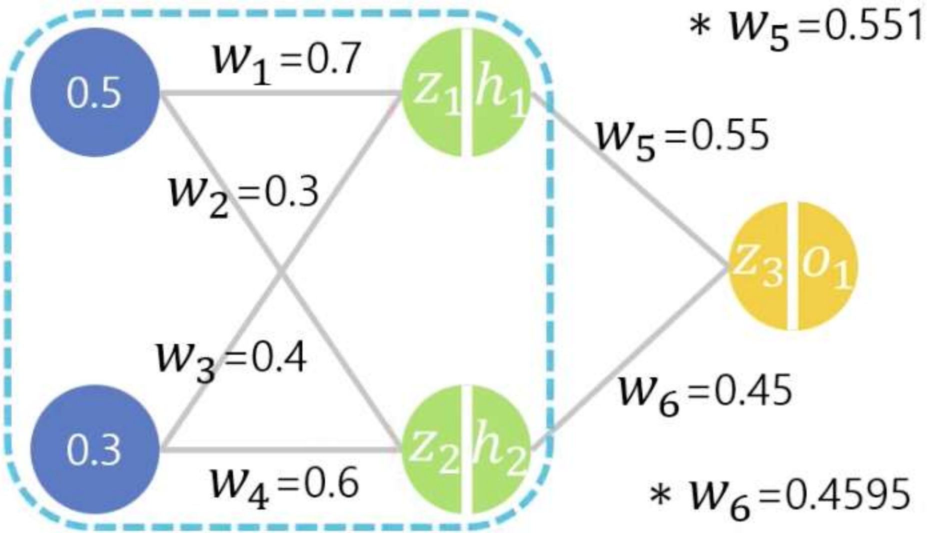
* w_5=0.551
0.5
w_1=0.7
z_1 h_1
w_5=0.55
w_2=0.3
z_3 o_1
w_3=0.4
w_6=0.45
0.3
z_2 h_2
w_4=0.6



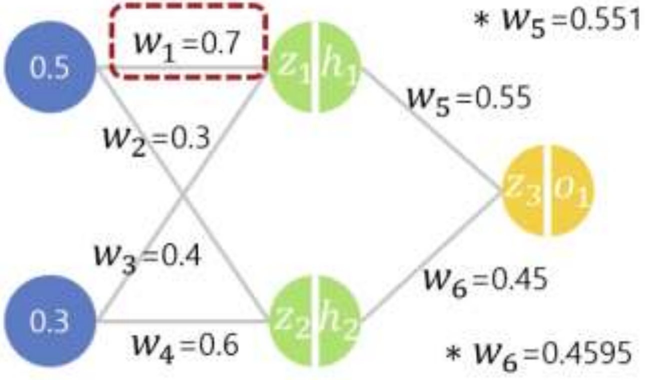
0.5
0.3
w_1 =0.7
w_2 =0.3
w_3 =0.4
w_4 =0.6
z_1 h_1
z_2 h_2
w_5 =0.55
w_6 =0.45
z_3 o_1
* w_5 =0.551
* w_6 =0.4595



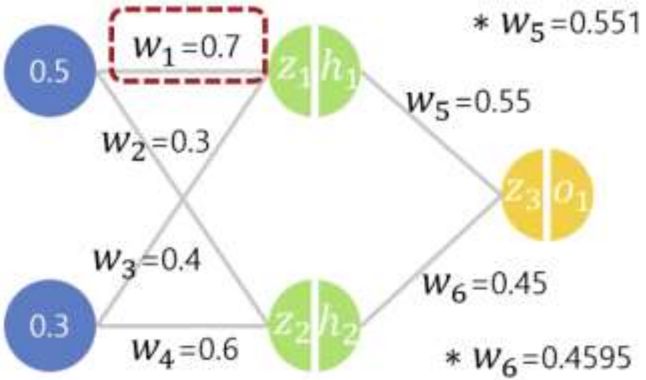
0.5
0.3
w_1 =0.7
w_2 =0.3
w_3 =0.4
w_4 =0.6
z_1 h_1
z_2 h_2
w_5 =0.55
w_6 =0.45
z_3 o_1
* w_5 =0.551
* w_6 =0.4595



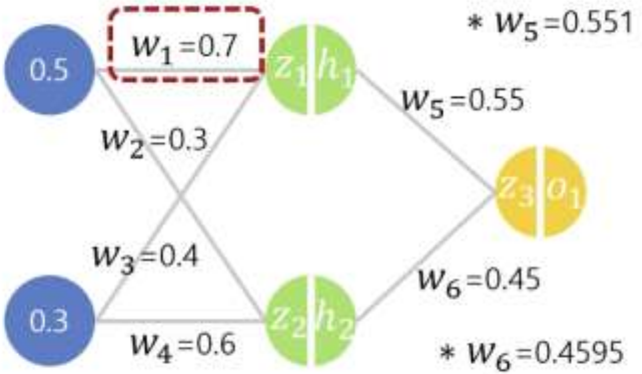
* w_5 =0.551
0.5
w_1 =0.7
z_1 h_1
w_5 =0.55
w_2 =0.3
z_3 o_1
w_3 =0.4
w_6 =0.45
0.3
z_2 h_2
w_4 =0.6
* w_6 =0.4595



$$\frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$

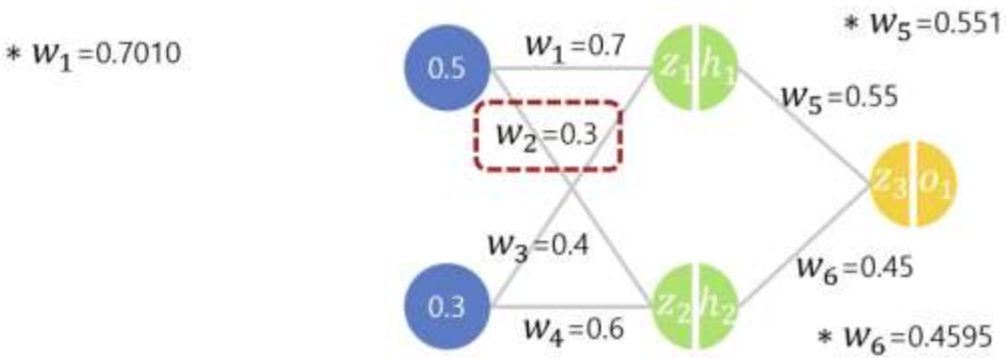


$$\frac{\partial C}{\partial w_1} = -0.0106$$



$$\frac{\partial C}{\partial w_1} = -0.0106$$ 새 연결강도 = 0.7 + 0.0106 x 0.1

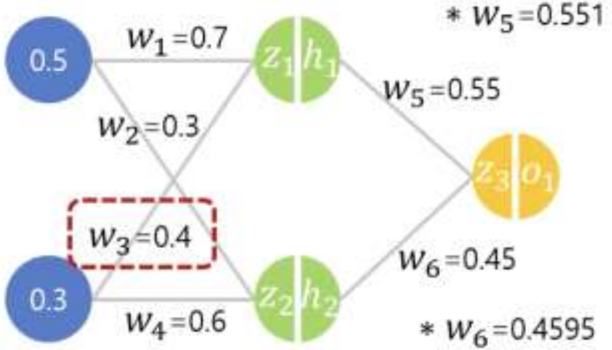
= 0.7010



$$\frac{\partial C}{\partial w_2} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = -0.009$$

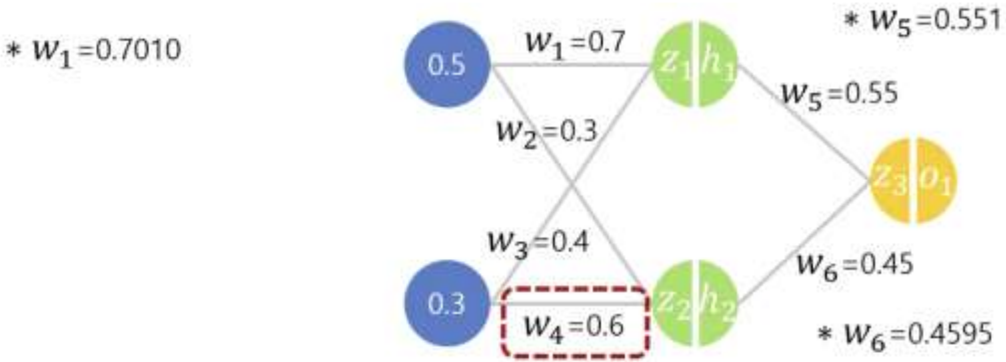
$-0.71 \cdot 0.229$ $\quad$ w_6 =0.45 $\quad$ $h_2(1-h_2)$ =0.243 $\quad$ x_1 =0.5

$* \ w_1 = 0.7010$



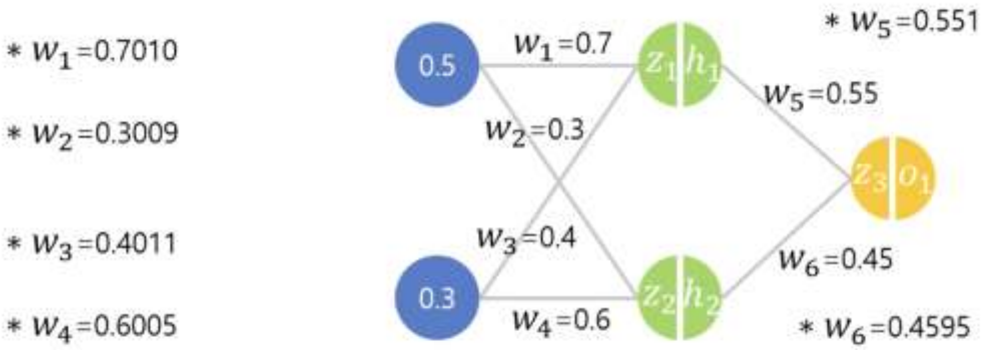
$$\frac{\partial C}{\partial w_3} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_3} = -0.011$$

$-0.71 \cdot 0.229$ $\quad w_5 = 0.55$ $\quad h_1(1 - h_1) = 0.237$ $\quad x_2 = 0.5$

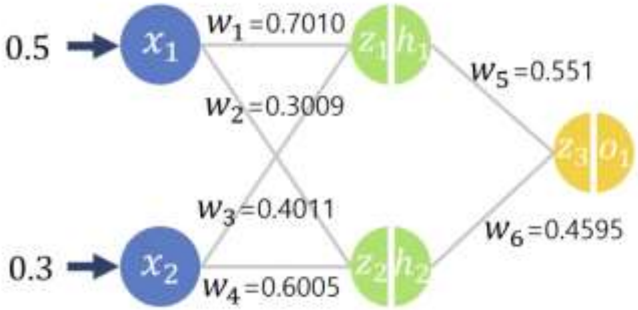


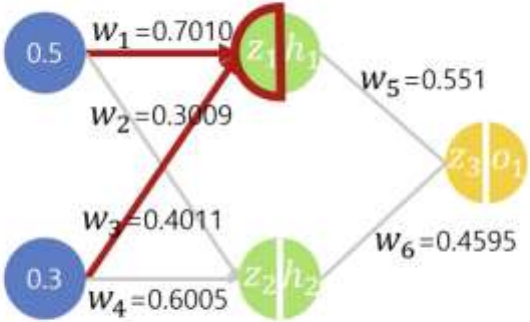
$$\frac{\partial C}{\partial w_4} = \frac{\partial C}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_3} \cdot \frac{\partial z_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_4} = -0.005$$

$-0.71 \cdot 0.229$ $\quad$ w_6 =0.45 $\quad$ $h_2(1-h_2)$ =0.243 $\quad$ x_1 =0.3

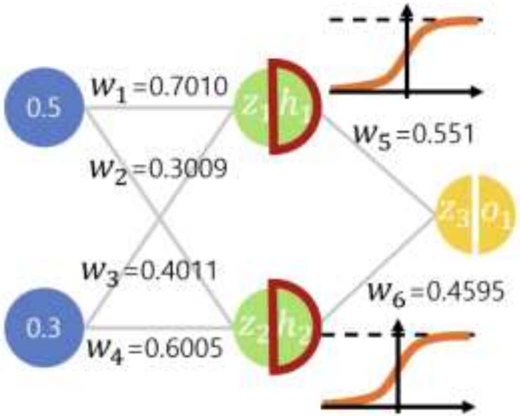


$$\frac{\partial c}{\partial w_2} = -0.009 \qquad \frac{\partial c}{\partial w_3} = -0.011 \qquad \frac{\partial c}{\partial w_4} = -0.005$$


0.5
0.3
x_1
x_2
w_1 =0.7010
w_2 =0.3009
w_3 =0.4011
w_4 =0.6005
z_1 h_1
z_2 h_2
w_5 =0.551
w_6 =0.4595
z_3 o_1



$z_1 = x_1 w_1 + x_2 w_3 = 0.5\text{x}0.7010 + 0.3\text{x}0.4011 =$ **0.4708**

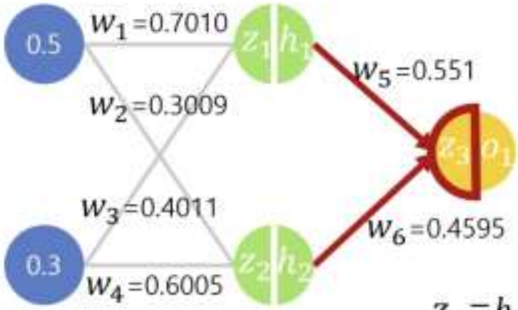


$z_1 = x_1 w_1 + x_2 w_3 = 0.5 \times 0.7010 + 0.3 \times 0.4011 = \mathbf{0.4708}$

$h_1 = sigmoid(z_1) = \mathbf{0.6156}$

$z_2 = x_1 w_2 + x_2 w_4 = 0.5 \times 0.3009 + 0.3 \times 0.6005 = \mathbf{0.3306}$

$h_2 = sigmoid(z_2) = \mathbf{0.5819}$



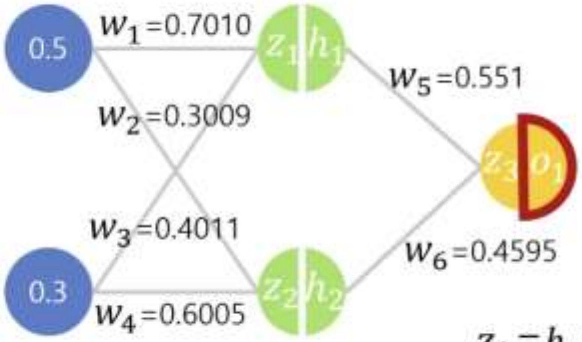
$z_3 = h_1 w_5 + h_2 w_6$

$= 0.6156 \times 0.551 + 0.5819 \times 0.4595$

$= \mathbf{0.6066}$

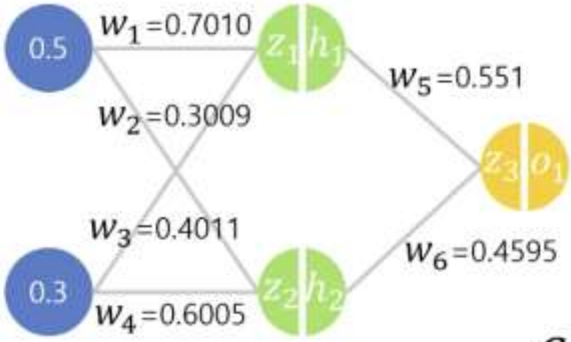
$h_1 = sigmoid(z_1) = \mathbf{0.6156}$

$h_2 = sigmoid(z_2) = \mathbf{0.5819}$



$o_1 = sigmoid(z_3) =$ **0.6472**

$z_3 = h_1 w_5 + h_2 w_6$
=0.6156x0.551+0.5819x0.4595
=**0.6066**

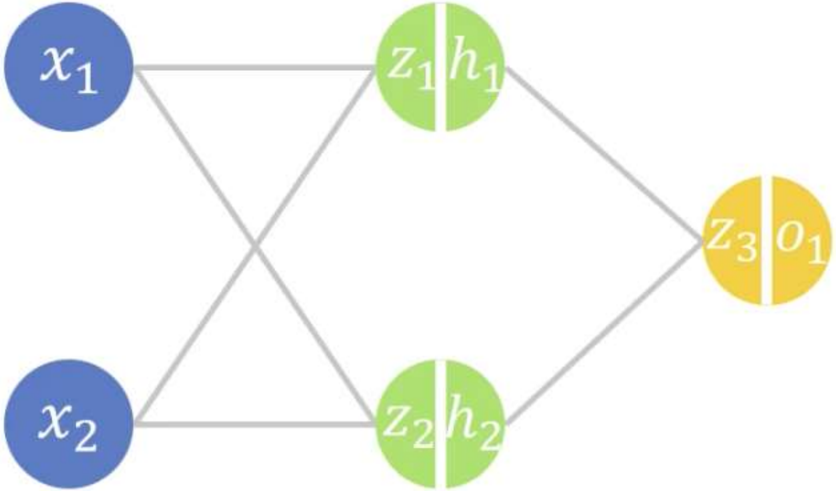


$o_1 = sigmoid(z_3) = 0.6443$

$$C = \frac{1}{1}\sum_{i}^{1}(1 - 0.6472)^2$$

$$C = \mathbf{0.1245}$$

$C = \mathbf{0.126}$ (이전오차)



x_1
x_2
z_1 h_1
z_2 h_2
z_3 o_1

Simple Neural Networks and Neural Language Models

Feedforward Neural Networks

Neural Language Models (LMs)

Language Modeling: Calculating the probability of the next word in a sequence given some history.

- We've seen N-gram based LMs
- But neural network LMs far outperform n-gram language models

State-of-the-art neural LMs are based on more powerful neural network technology like Transformers

But **simple feedforward LMs** can do almost as well!

Simple feedforward Neural Language Models

Task: predict next word w_t

given prior words w_{t-1}, w_{t-2}, w_{t-3}, …

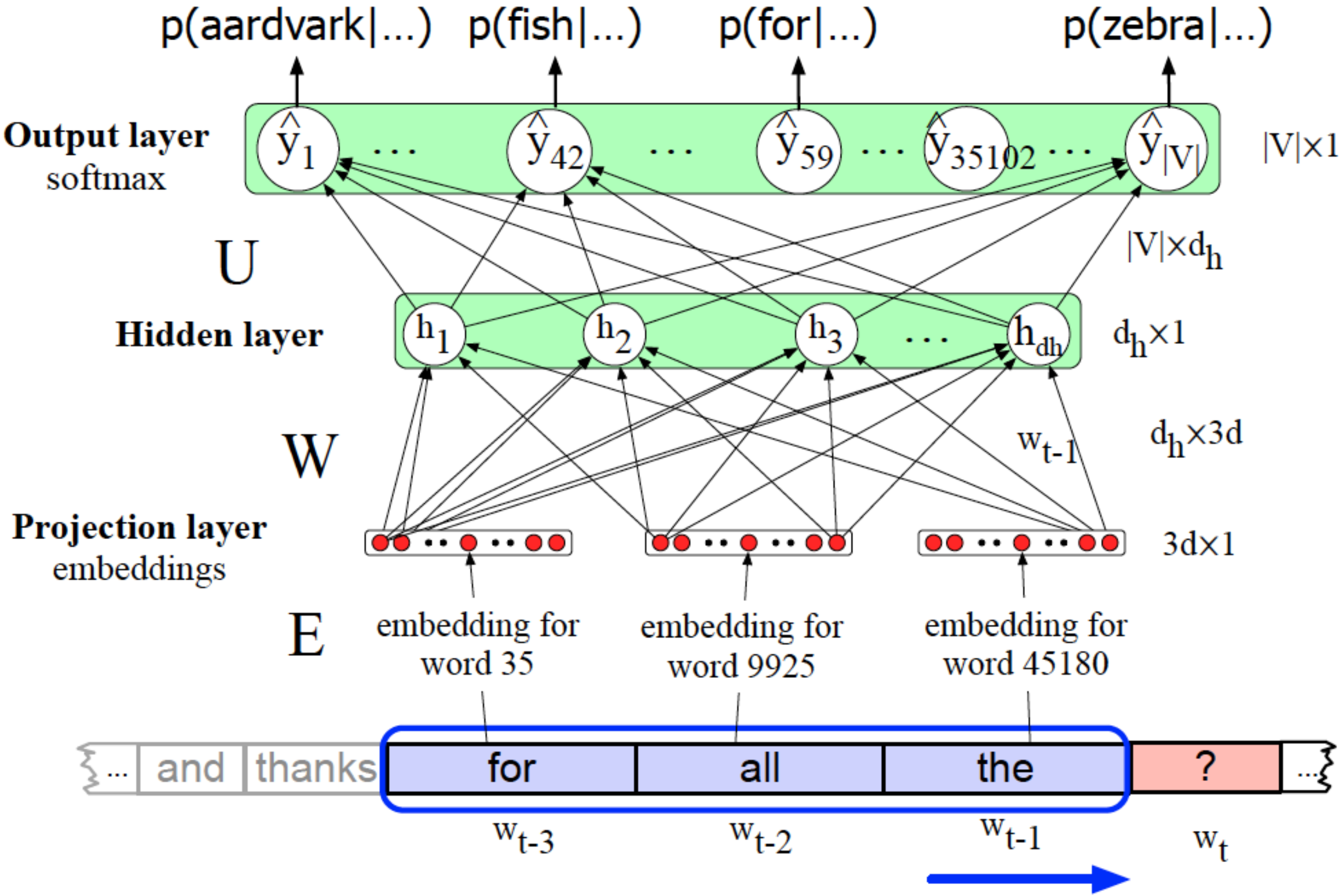
Problem: Now we're dealing with sequences of arbitrary length.

Solution: Sliding windows (of fixed length)

$$P(w_t|w_1^{t-1}) \approx P(w_t|w_{t-N+1}^{t-1})$$

$$w_t|w_{t-N+1},\ldots,w_{t-1}$$

Neural Language Model



Why Neural LMs work better than N-gram LMs

Training data:

We've seen: I have to make sure that the cat gets fed.

Never seen: dog gets fed

Test data:

I forgot to make sure that the dog gets ___

N-gram LM can't predict "fed"!

Neural LM can use similarity of "cat" and "dog" embeddings to generalize and predict “fed” after dog

감사합니다