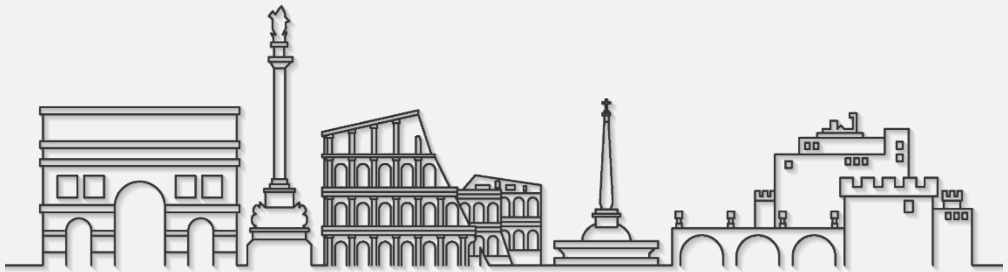


모두를 위한 딥러닝

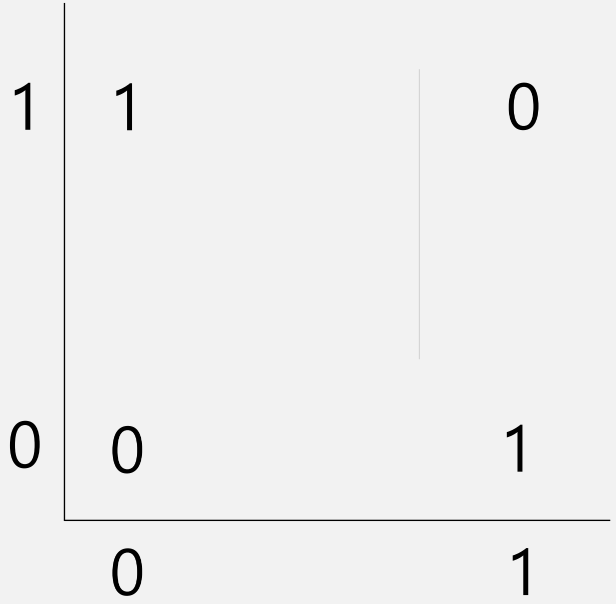
Ybigta science

2015147574
백진우

Contents

01. XOR 문제

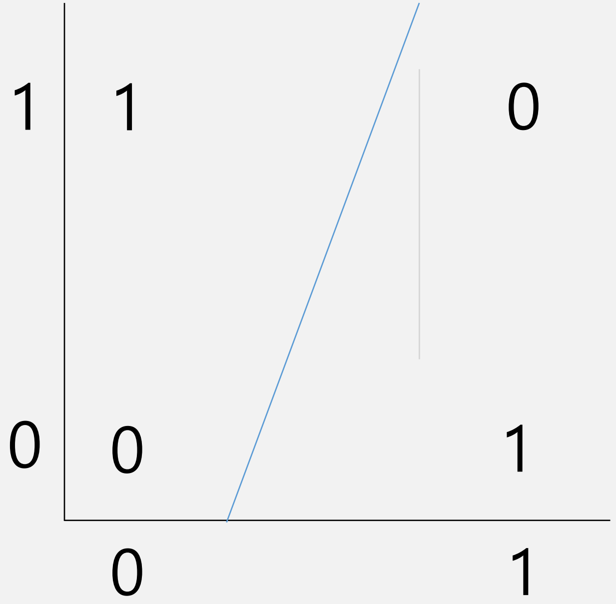
단일 네트워크의 XOR 문제



- XOR 연산은 두개의 값이 같은 경우에는 False, 다른 경우에는 True가 된다.
- 어떻게 해도 직선으로는 설명할 수 없음.

01. XOR 문제

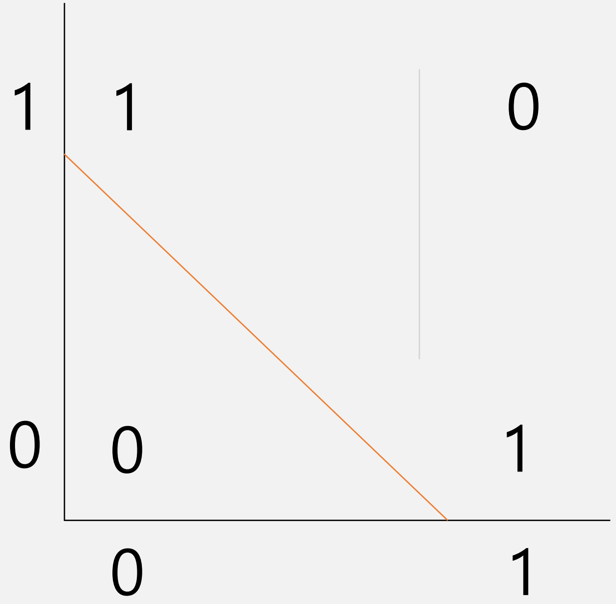
단일 네트워크의 XOR 문제



- XOR 연산은 두개의 값이 같은 경우에는 False, 다른 경우에는 True가 된다.
- 어떻게 해도 직선으로는 설명할 수 없음.

01. XOR 문제

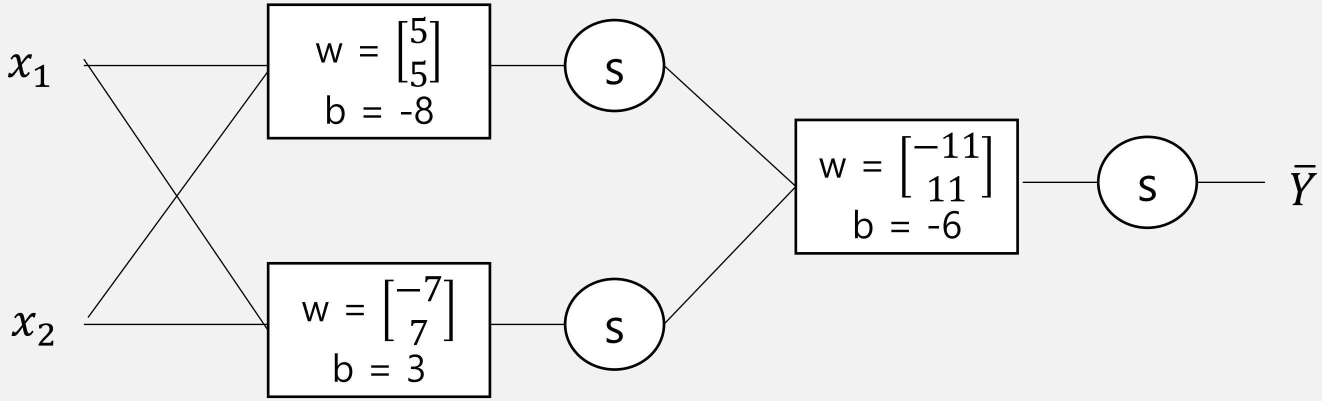
단일 네트워크의 XOR 문제



- XOR 연산은 두개의 값이 같은 경우에는 False, 다른 경우에는 True가 된다.
- 어떻게 해도 직선으로는 설명할 수 없음.

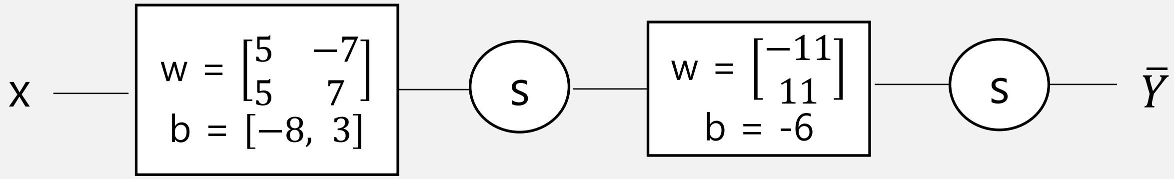
01. XOR 문제

XOR 문제 해결



여러 개의 뉴런 네트워크로 XOR 문제를 해결 할 수 있다.

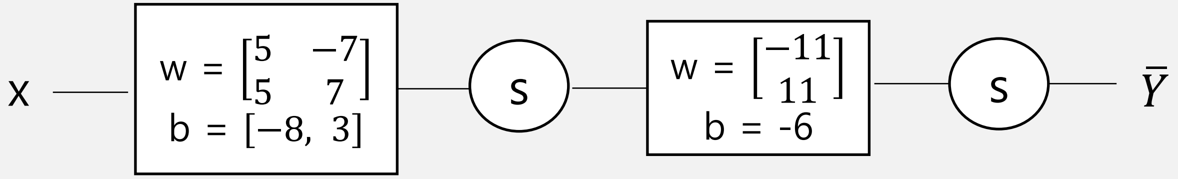
XOR 문제 해결



여러 개의 뉴런 네트워크로 XOR 문제를 해결 할 수 있다.

02. Back propagtion

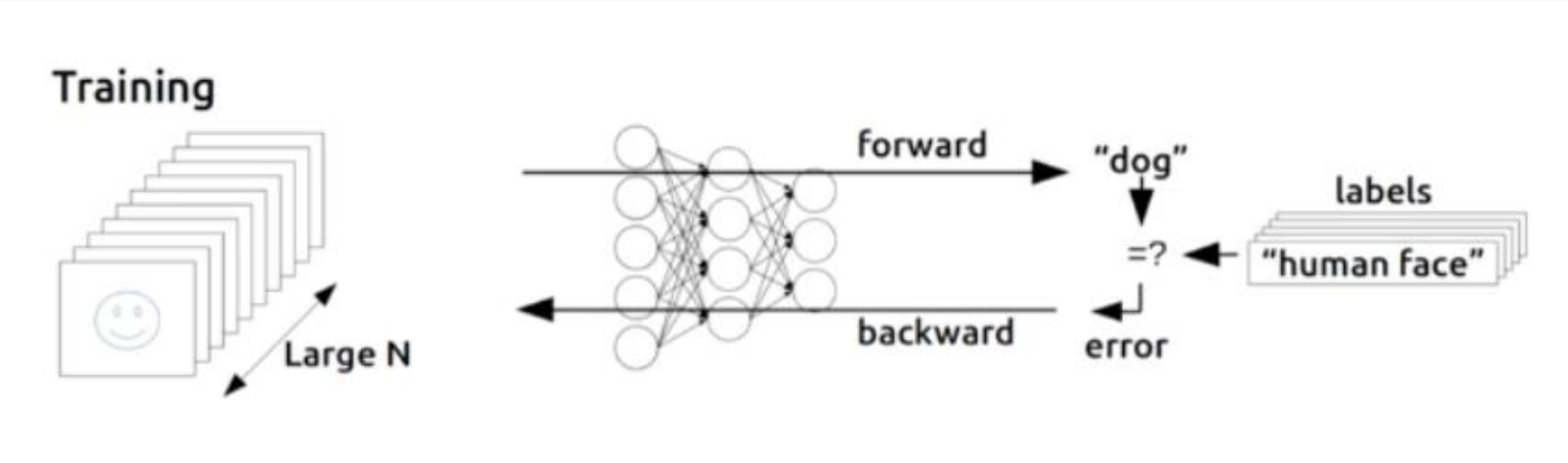
역전파



결론은 gradient descent 알고리즘!

02. Back propagtion

역전파



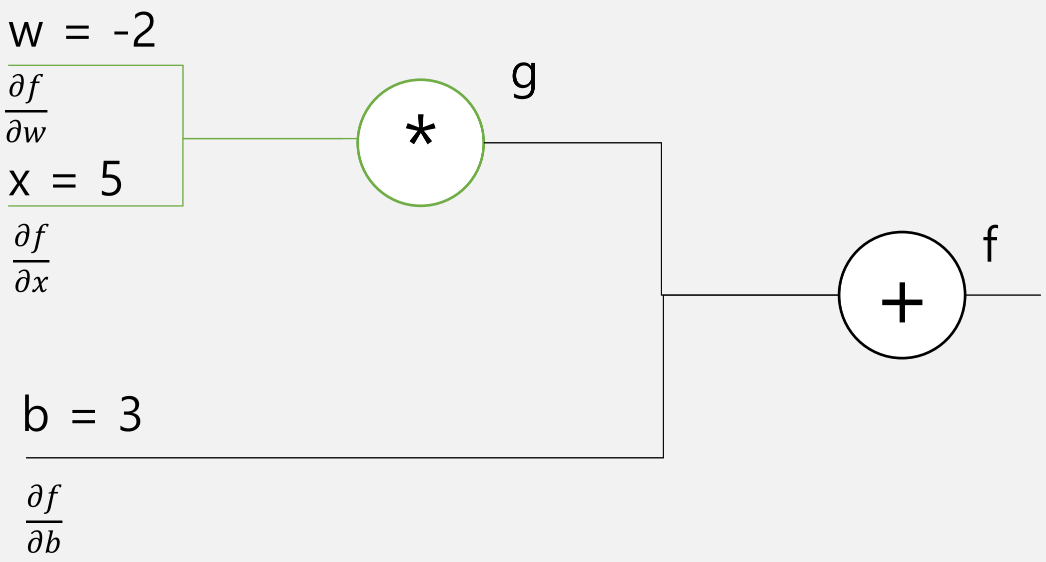
앞에서부터 뒤로 진행하면서 W와 b를 바꾸어 나갈 때 forward propagation이라고 하고, 뒤에서부터 앞으로 거꾸로 진행하면서 바꾸어 나갈 때 backward propagation이라고 한다.

02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$

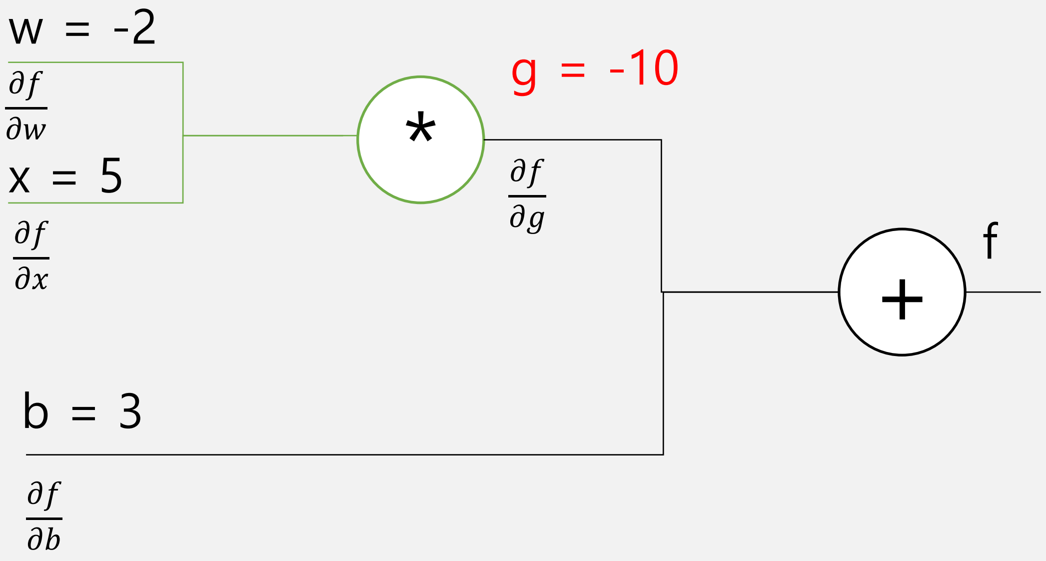


02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$

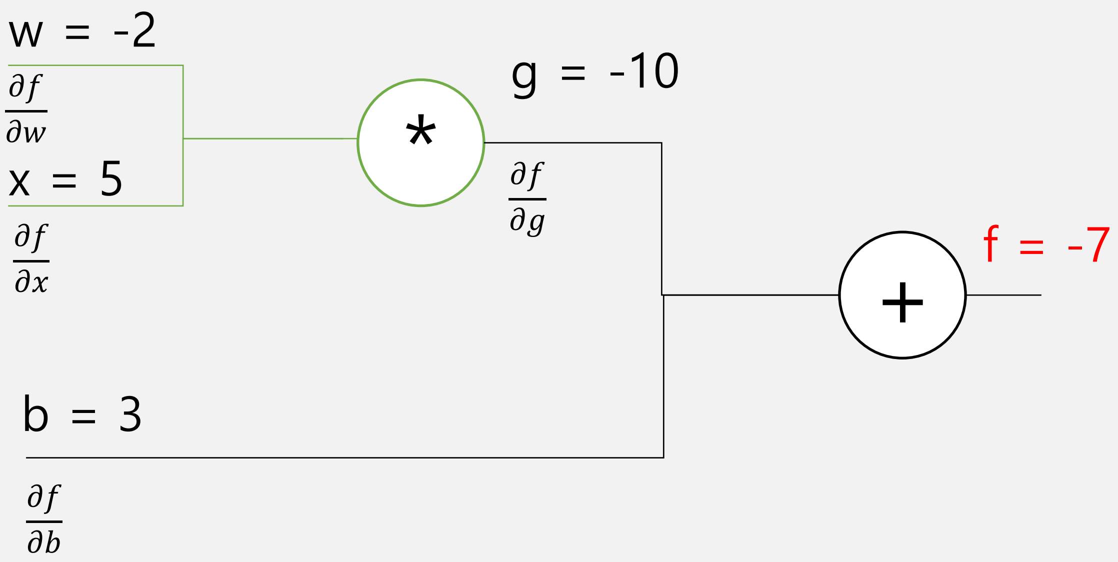


02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$

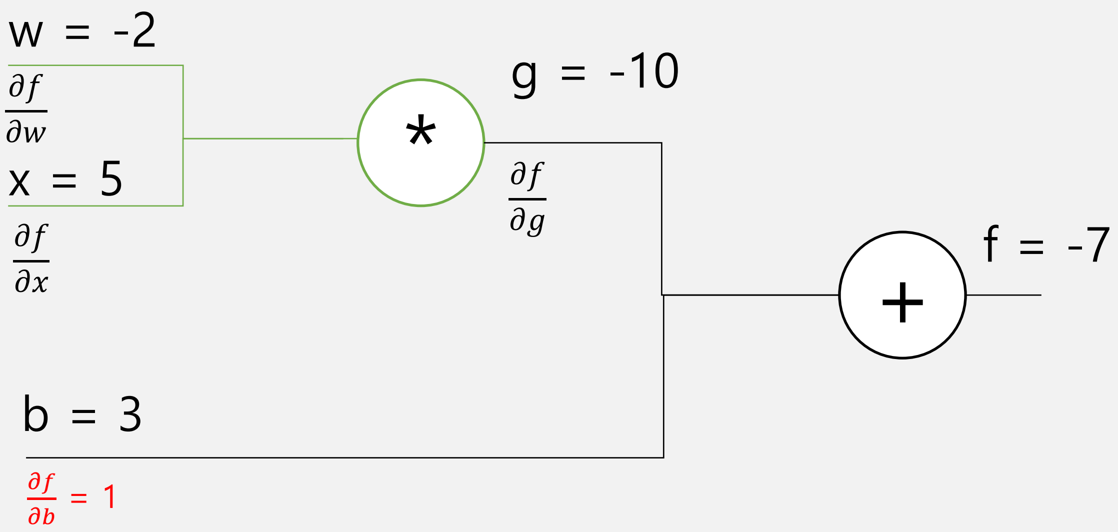


02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$

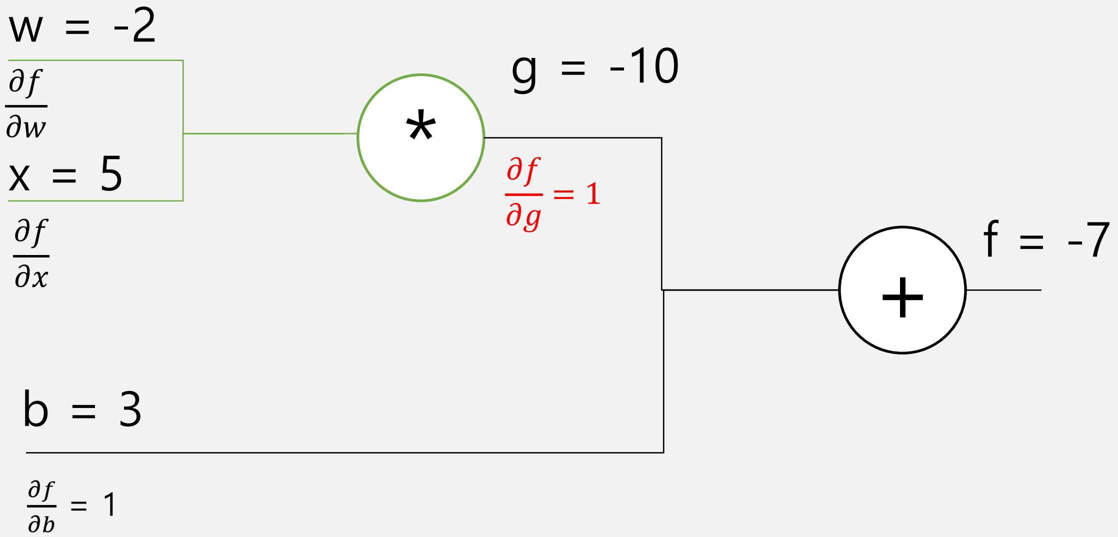


02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$

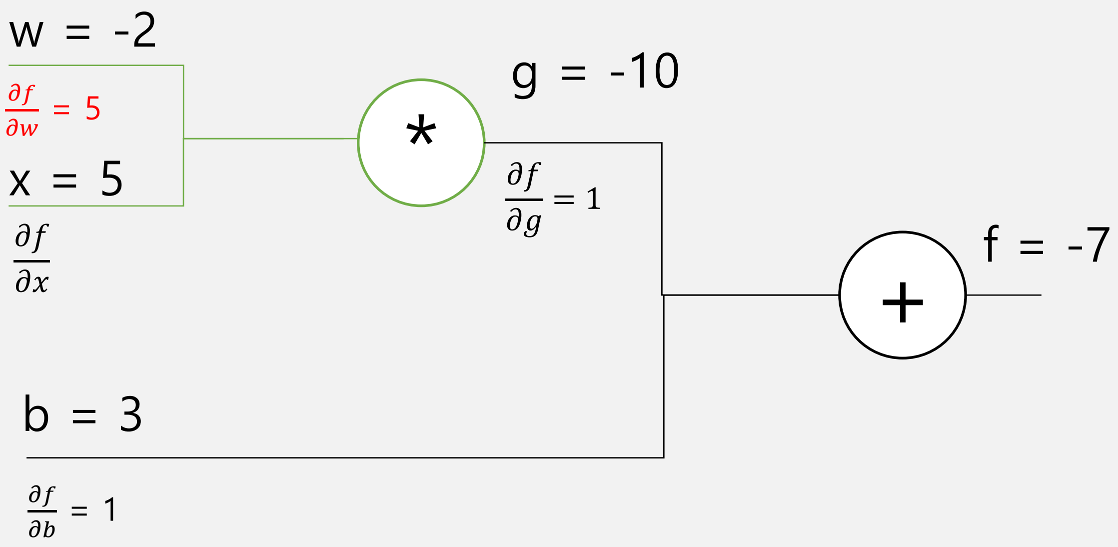


02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$



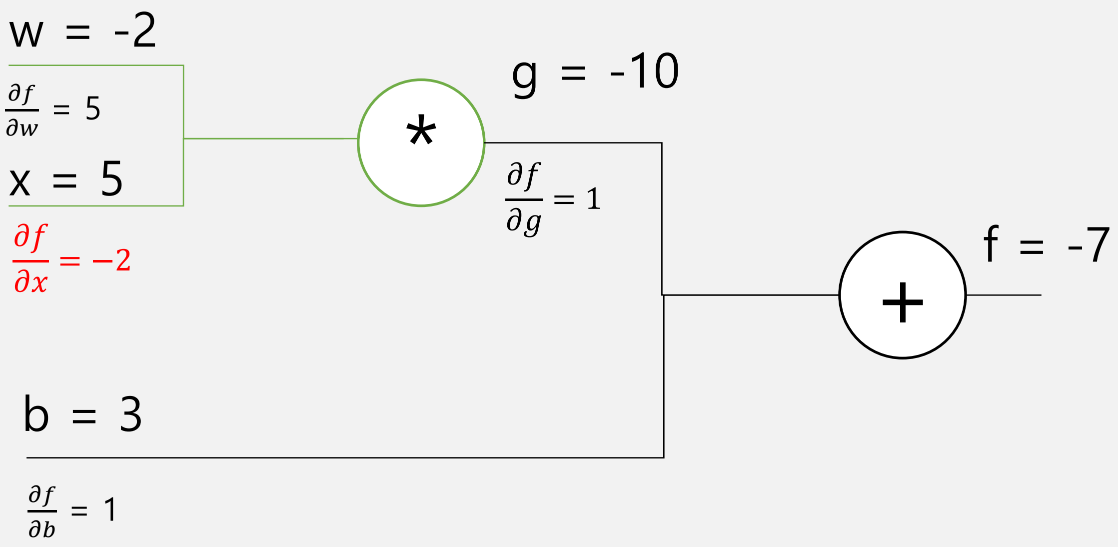
$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial w} = 1 * 5 = 5$$

02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$



$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial w} = 1 * 5 = 5$$

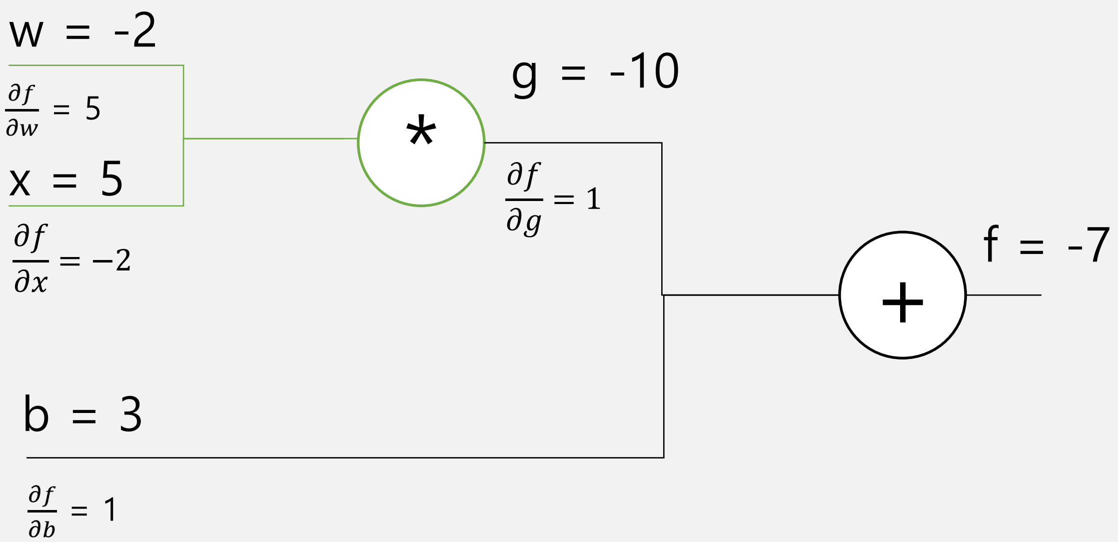
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial x} = 1 * -2 = -2$$

02. Back propagtion

Chain rule

$$f = wx + b, \quad g = wx, \quad f = g + b$$

$$\frac{\partial g}{\partial w} = x, \quad \frac{\partial g}{\partial x} = w, \quad \frac{\partial f}{\partial g} = 1, \quad \frac{\partial f}{\partial b} = 1$$



$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial w} = 1 * 5 = 5$$

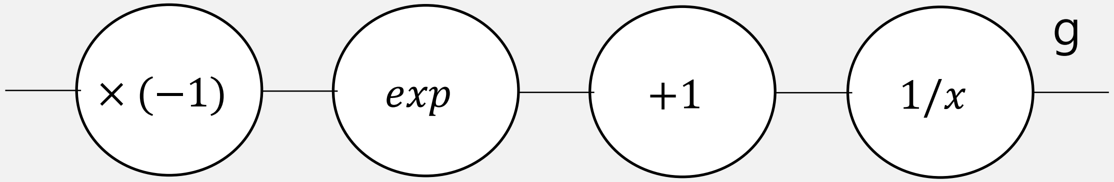
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial x} = 1 * -2 = -2$$

w, x, b 값이 1 변할 때 f값이 얼마나 변하는 지 알 수 있다.
즉, 경사하강법 식인 $w^+ = w - \gamma \frac{\partial f}{\partial w}$ 으로 학습 시킬 수 있다.

02. Back propagtion

Sigmoid

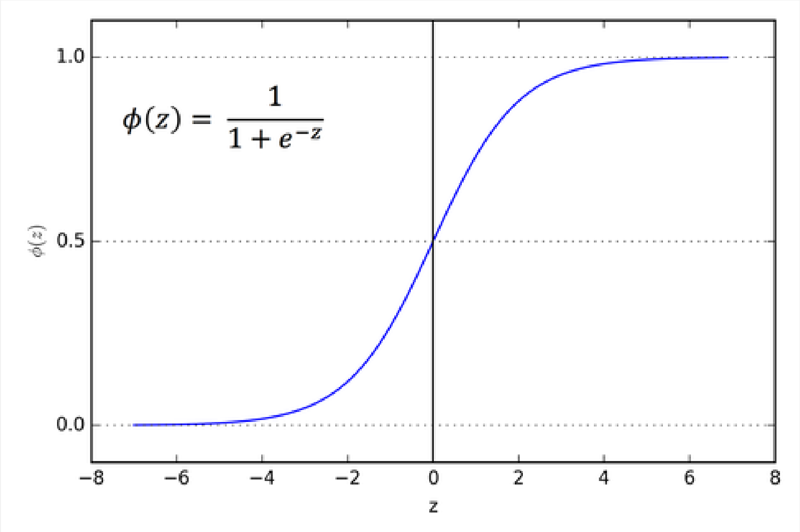
$$g(z) = \frac{1}{1+e^{-z}}$$



Sigmoid 또한 여러 개의 계산으로 나누어질 수 있고, 역으로 미분을 통해 어느정도의 영향을 줬는지 판단할 수 있다.

03. Sigmoid vs ReLU

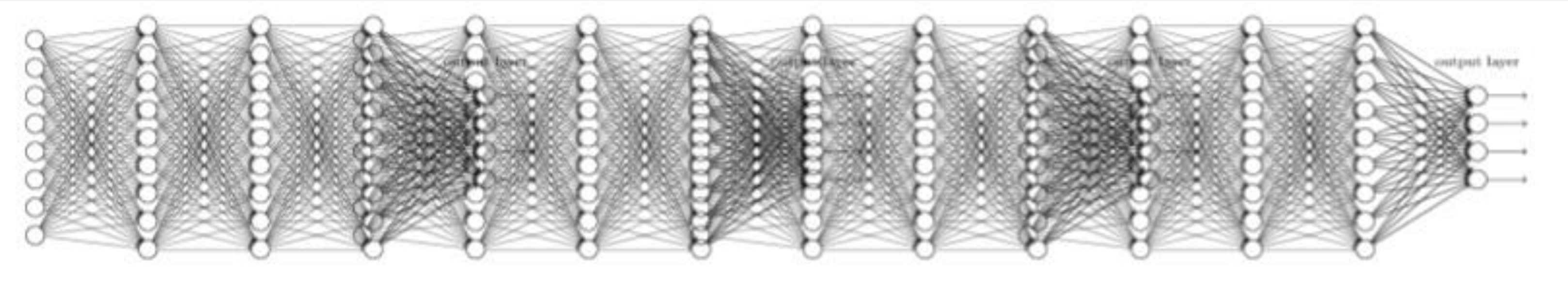
Sigmoid



Sigmoid 값은 -1 과 1 사이 값을 가진다.
-> 역전파 과정에서 문제가 생긴다

03. Sigmoid vs ReLU

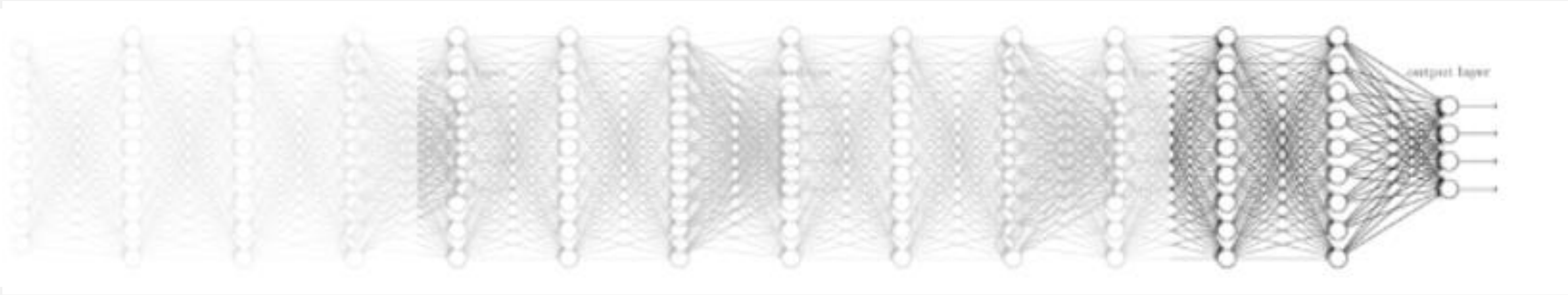
Back propagation



Layer가 많을 때는 미분 결과를 최초 layer 까지 전달 하는 것이 어렵다.

03. Sigmoid vs ReLU

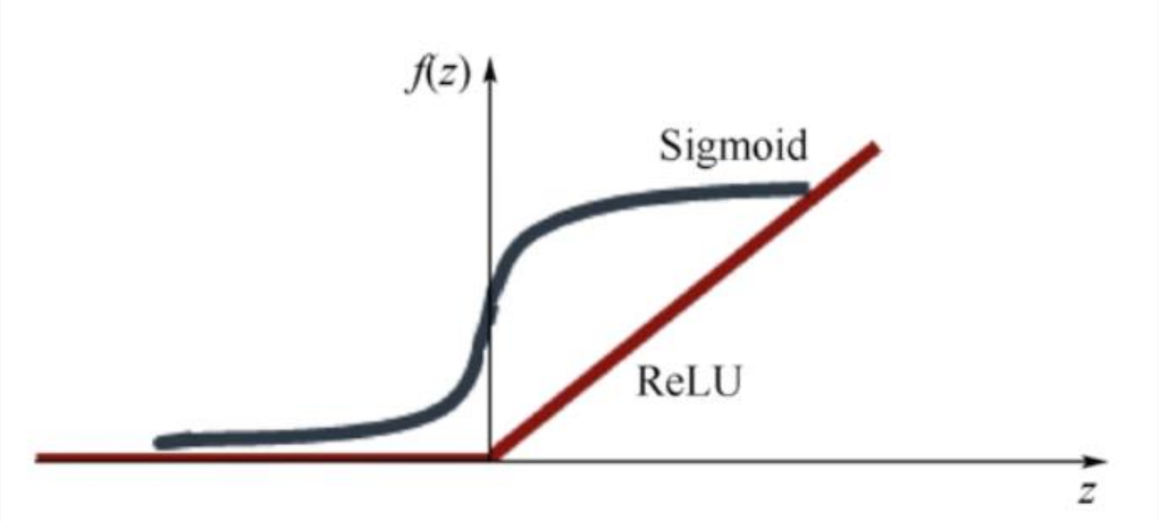
Vanishing gradient



layer을 지날 때마다 최초 값보다 현저하게 작아지기 때문에 값을 전달 해도 의미를 가질 수가 없다.

03. Sigmoid vs ReLU

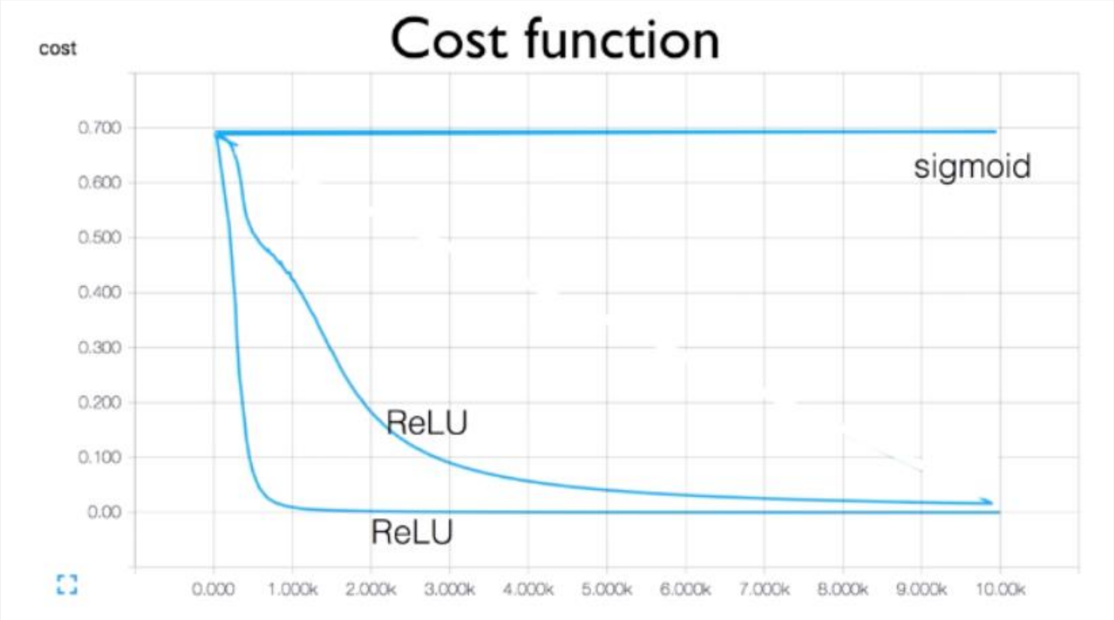
ReLU



기존의 sigmoid 문제를 해결하기 위해서 hinton 교수님이 ReLU 함수 제안!

03. Sigmoid vs ReLU

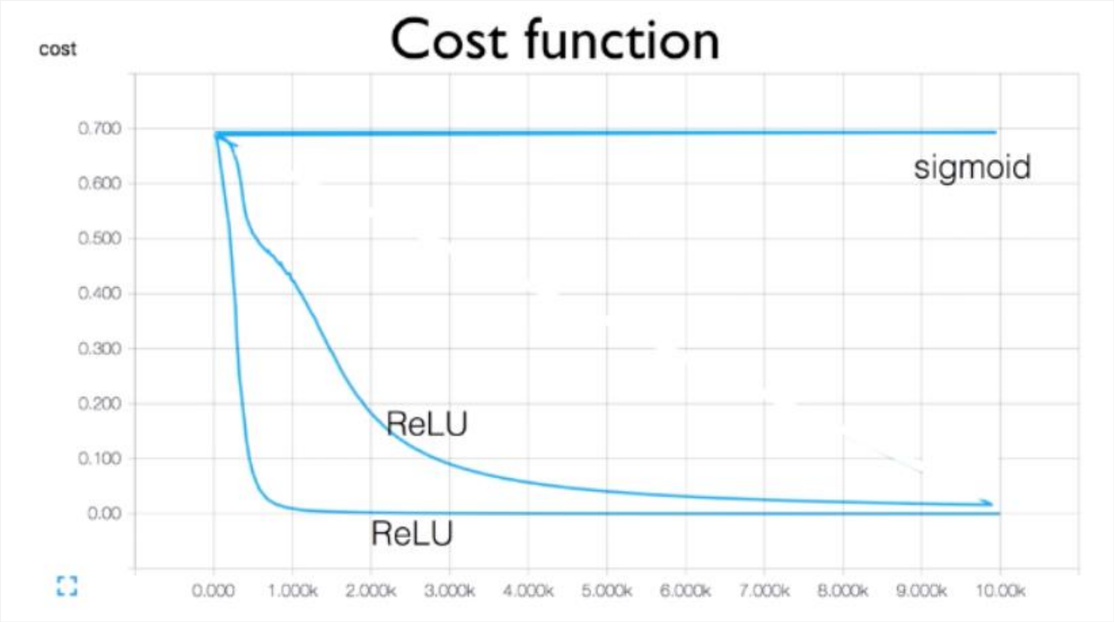
ReLU



ReLU가 sigmoid 보다 좋다!

03. Sigmoid vs ReLU

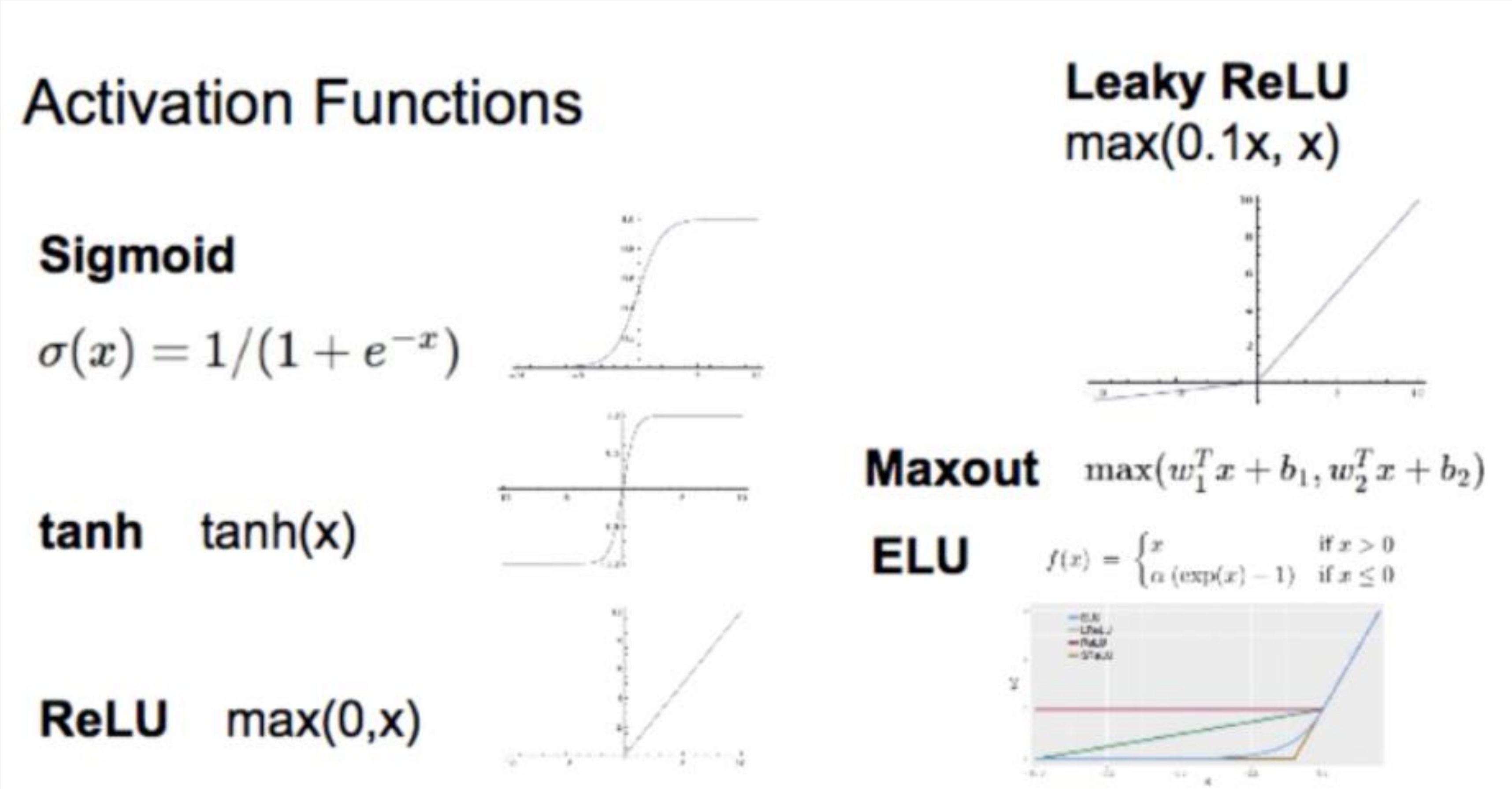
ReLU



ReLU가 sigmoid 보다 좋다!

03. Sigmoid vs ReLU

다양한 activation functions



03. Sigmoid vs ReLU

다양한 activation functions

maxout	ReLU	VLReLU	tanh	Sigmoid
93.94	**92.11**	92.97	89.28	n/c
93.78	91.74	92.40	89.48	n/c
–	91.93	**93.09**	–	n/c
91.75	90.63	92.27	**89.82**	n/c
n/c†	90.91	92.43	89.54	n/c

Activation 함수의 성능을 비교하고 있다.
sigmoid 함수가 n/c라고 나온 것은 결과를 낼 수 없기 때문이다.

감 사 합 니 다