Deep Learning Basic

Jaewon Kim, Dankook Univ.

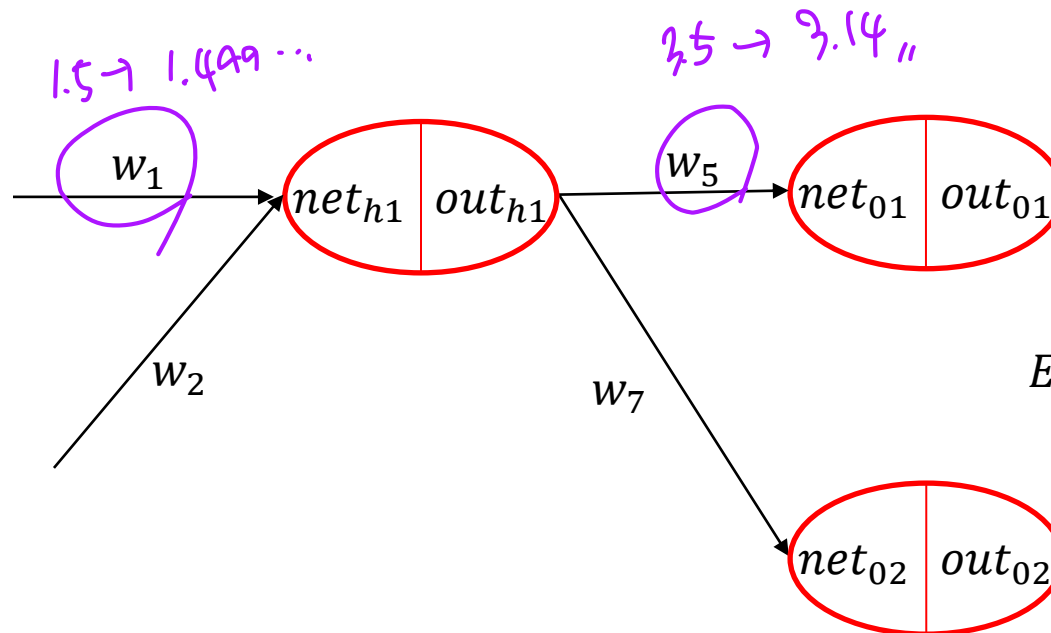**Chapter 3-2**

Contents

Part 1. Gradient Vanishing

Part 2. Activation Function

- If A.F is Linear ?
- Sigmoid
- Tanh
- ReLU
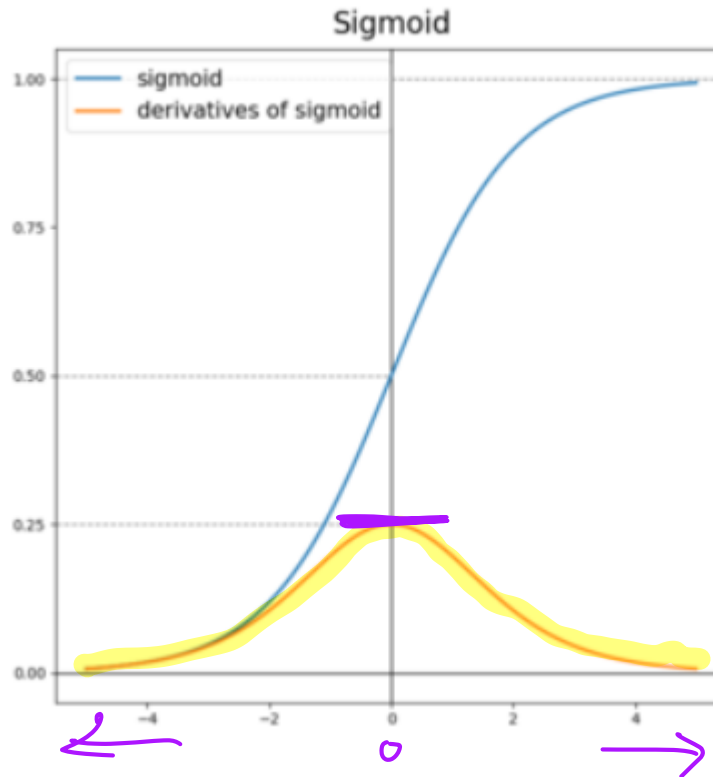- Leaky ReLU

Gradient Vanishing

- Backpropagation (w_1 update)

$1.5 \rightarrow 1.499 \cdots$

$3.5 \rightarrow 3.14 \, \text{''}$

w_1

net_{h1} | out_{h1}

w_5

net_{01} | out_{01}

w_2

w_7

net_{02} | out_{02}

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

w_1 parameter가 잘 update되지 않았다.
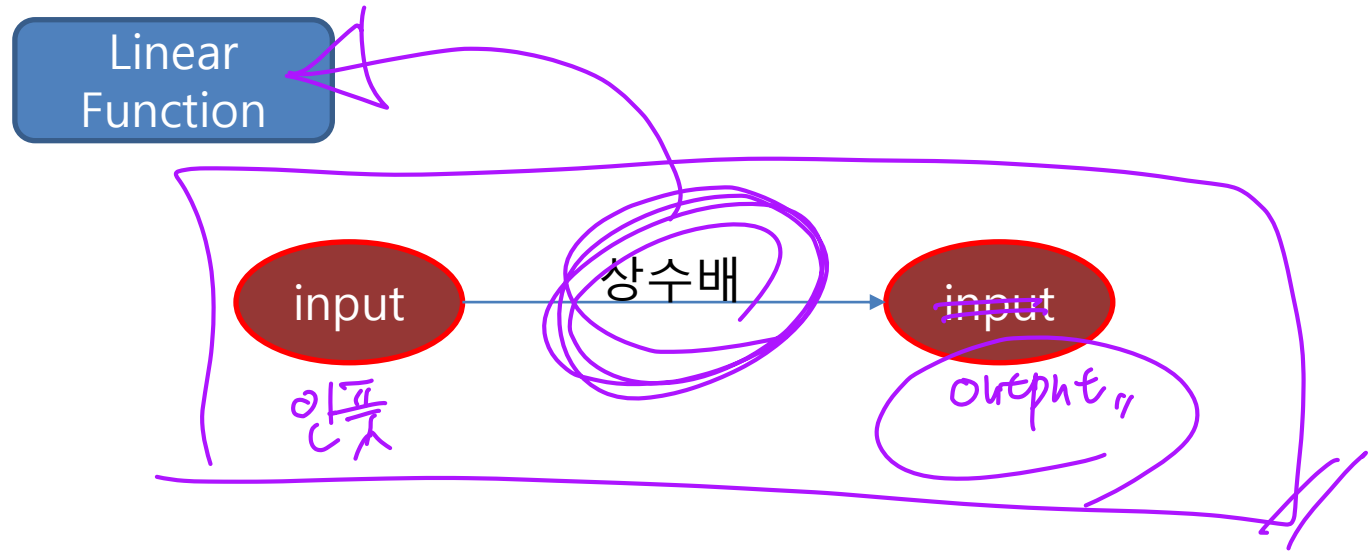
Gradient Vanishing

 - Sigmoid Function

$$Sig(x) = Sig(1-Sig)$$

Sigmoid

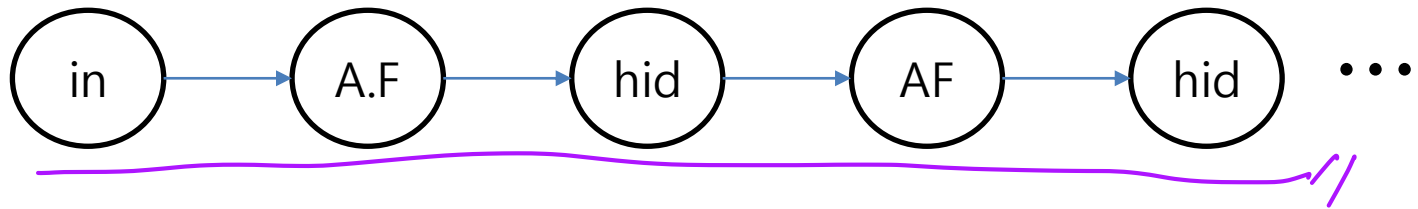Input이 0인 지점에서 기울기 =0

Activation Function

- Linear Function ?

If Activation Function is Linear ?

Activation Function

- Linear Function ?

If Activation Function is Linear ?

in → A.F → hid → AF → hid ...

Activation Function : $f(x) = wx$

↓

hidden layer 2개 증가.

$$y(x) = f(f(x))$$

↓

$$y(x) = f(f(f(x)))$$

$$f(wx) = w \cdot wx$$

$$f(w \cdot wx) = w \cdot w \cdot w \cdot x$$
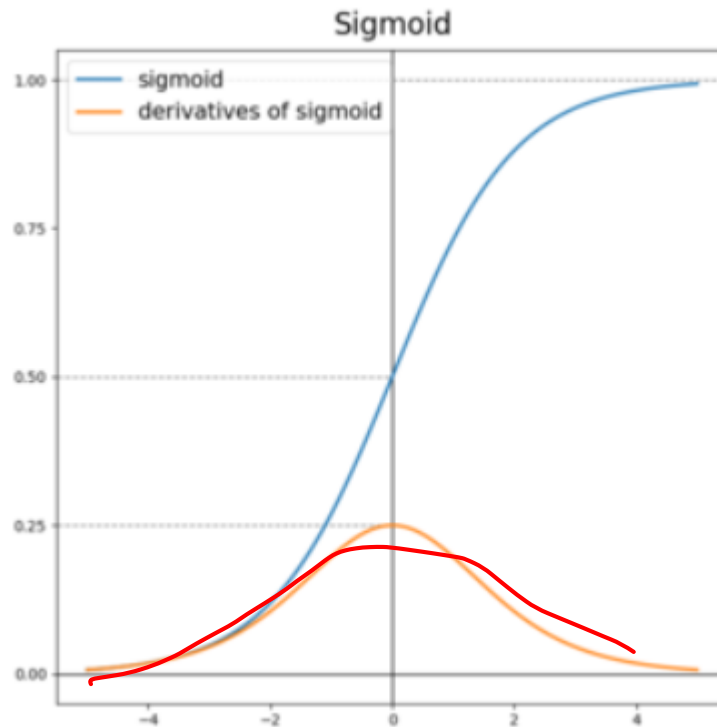
$$www \approx k$$

$$f(x) = wx$$

$$f(x) = k \cdot x$$

Activation Function

- Sigmoid Function

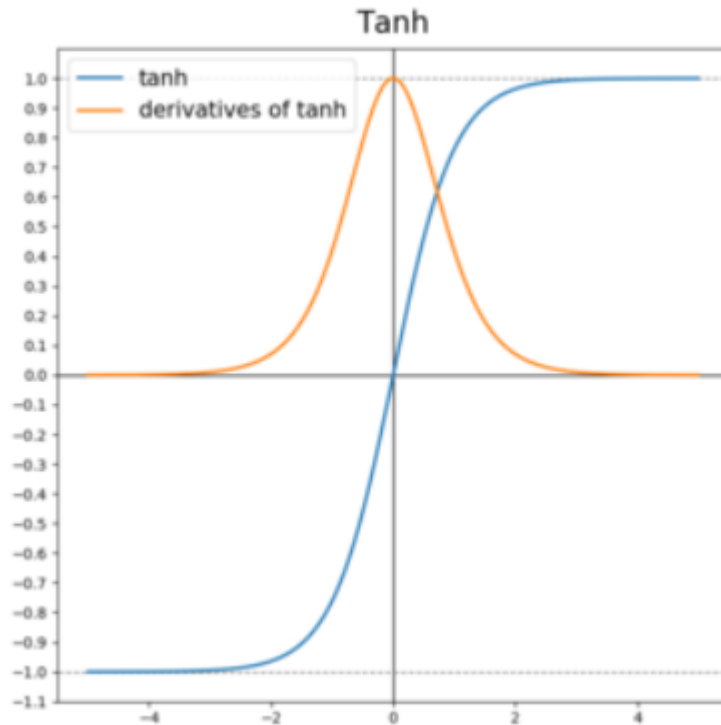Sigmoid

Pros   Cons

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

출력값 : 0 ~ 1

가장 보편적이

input= 0
too large , small

Gradient Vanishing.

Activation Function

- Hyperblic Tangent

Tanh

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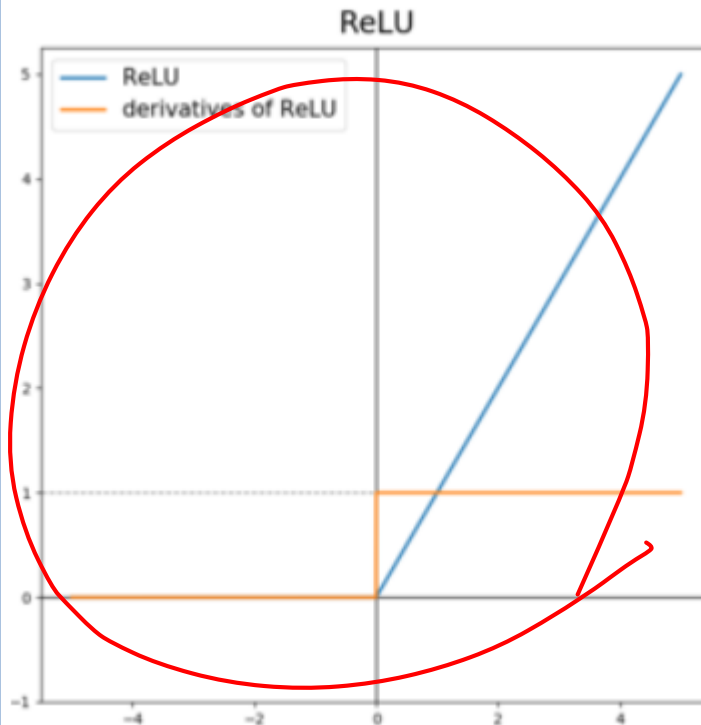
출력 −1 ~ 1

Pros

Cons

Sigmoid 보다는
다르게 있다.
↓
G . V 발생.

Activation Function
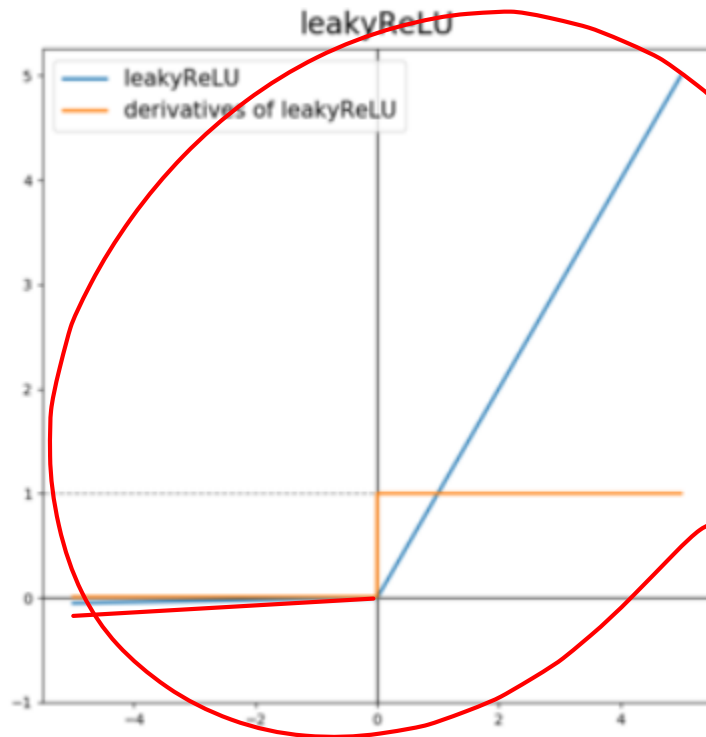
- ReLU

$$relu(x) = a = \max(0, z)$$

특징:

모든 가장 보편적.

계산효율 ↑ sig, tanh 6bu₄

Activation Function

- Leaky ReLU

$$leaky\ relu(x) = a = \max(0.01z, z)$$

특징 : 음수영역 이씀

↓

Saturation X

Pros   Cons

Thank you..!!!