

Part. 04

Back Propagation

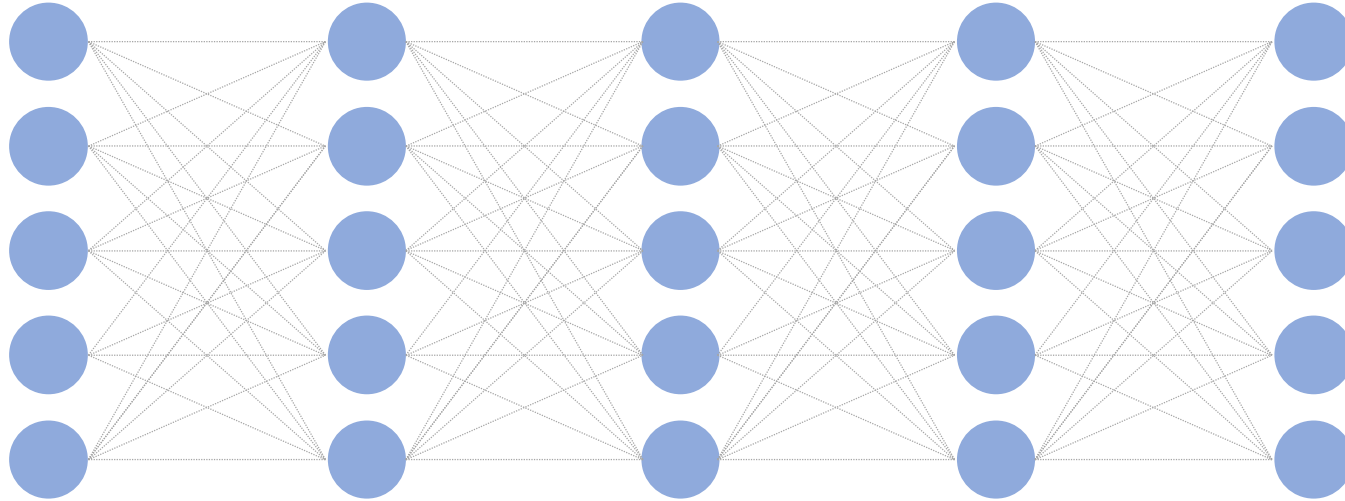
| 기울기 소실 문제

FASTCAMPUS
ONLINE

강사. 신제용

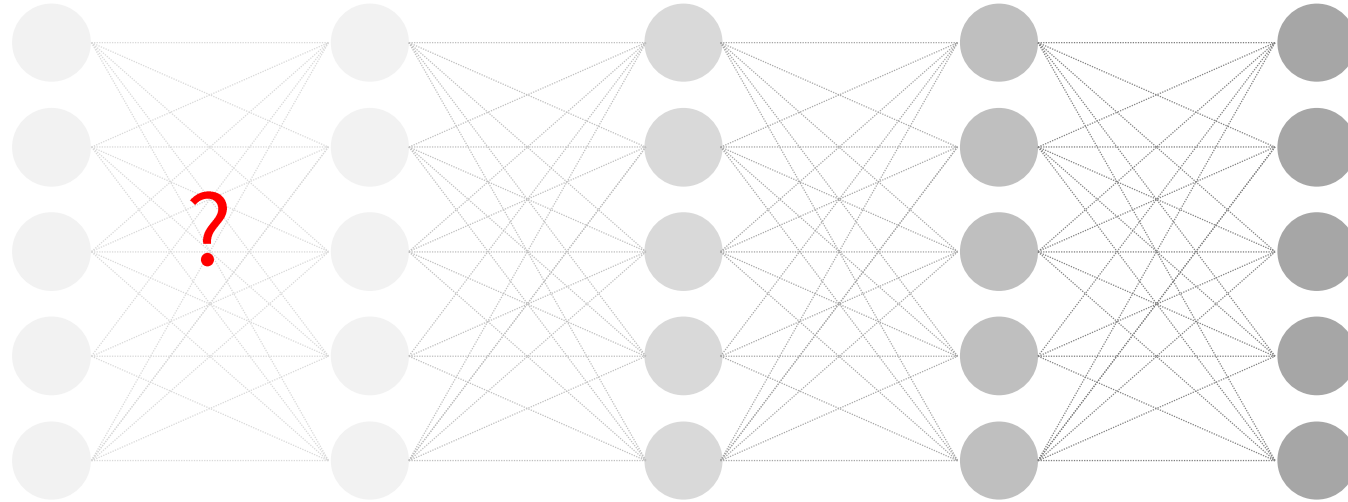
I 역전파면 다 된다?

Layer가 많아지면 표현력이 증가 → 역전파로 학습만 시키면 만사 형통?



역전파 알고리즘은 80년대에도 이미 사용되어 왔다. 그럼에도 딥러닝의 부흥은 왜 이리도 늦어졌는가?

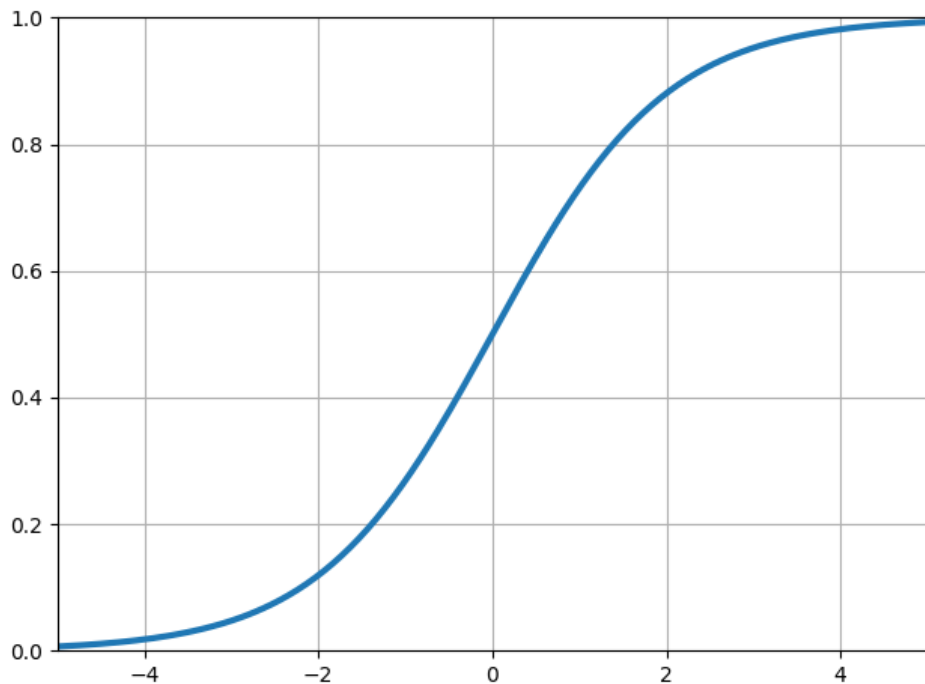
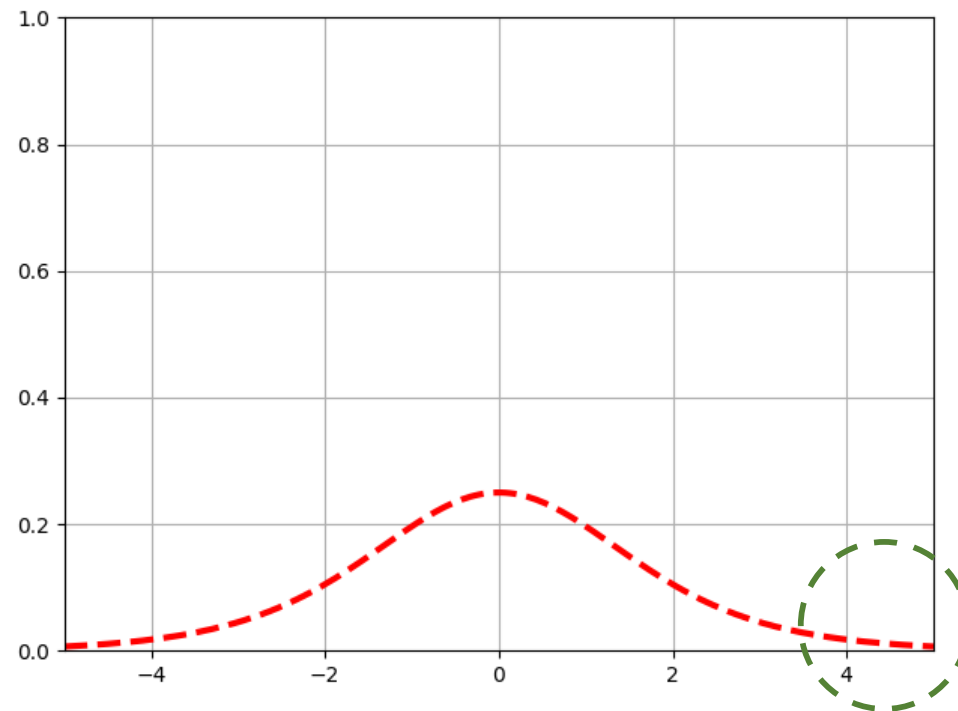
I 기울기 소실 문제



기울기 소실(Vanishing gradient) 문제. 역전파를 계층이 많아질 수록 학습이 잘 되지 않는다.
느낌상 그럴 수도 있을 것 같은데... 왜 그럴까?

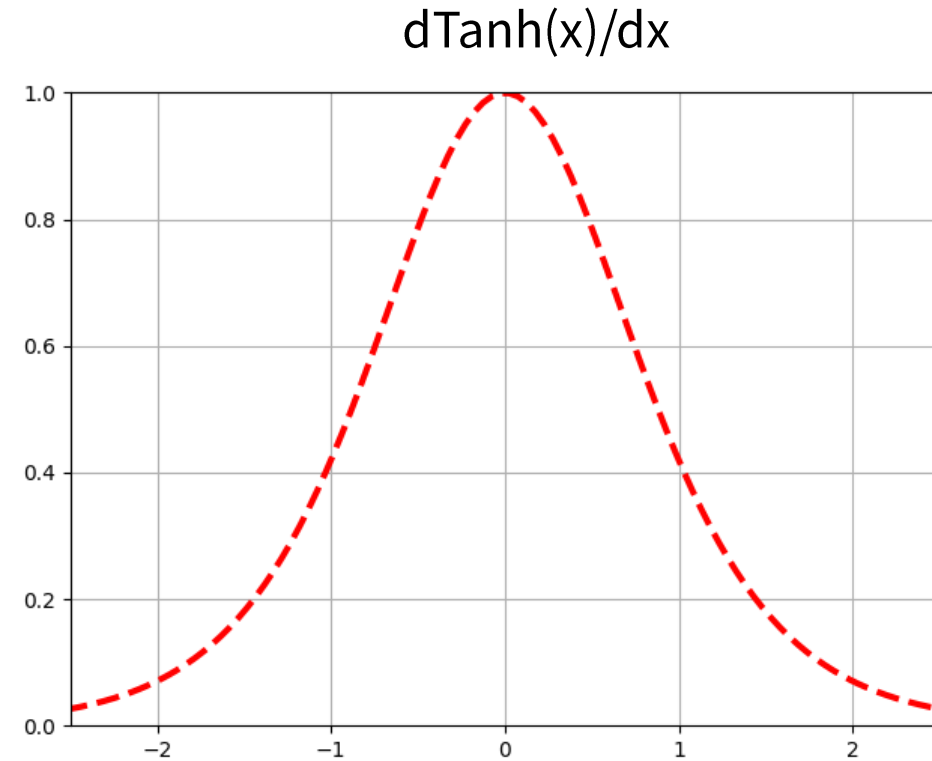
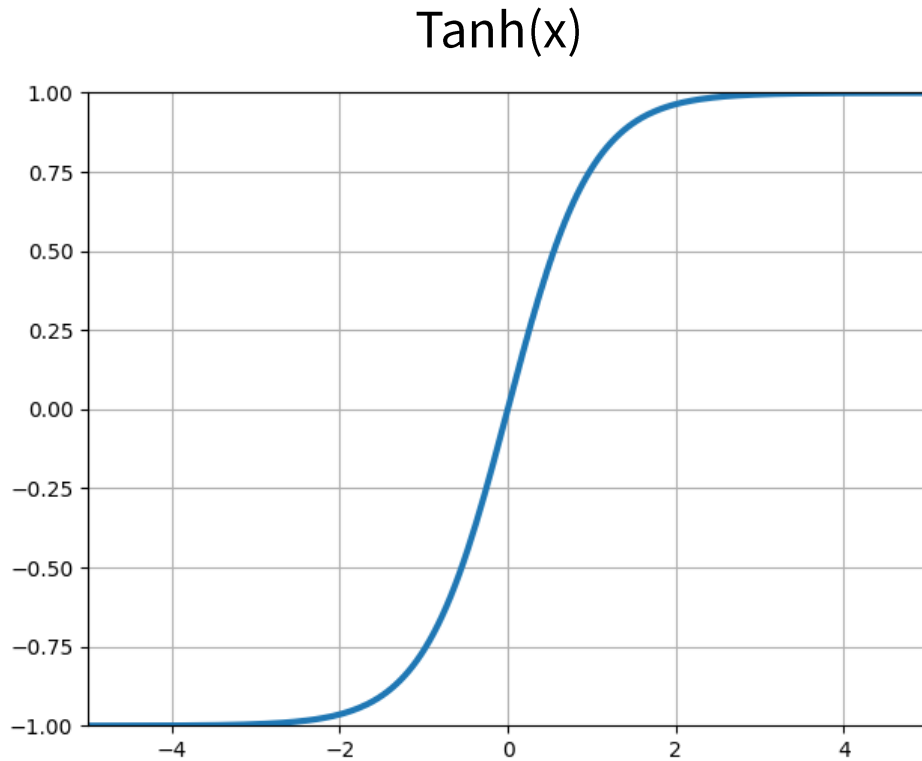
I 기울기 소실의 원인

Sigmoid(x)

 $d\text{Sigmoid}(x)/dx$ 

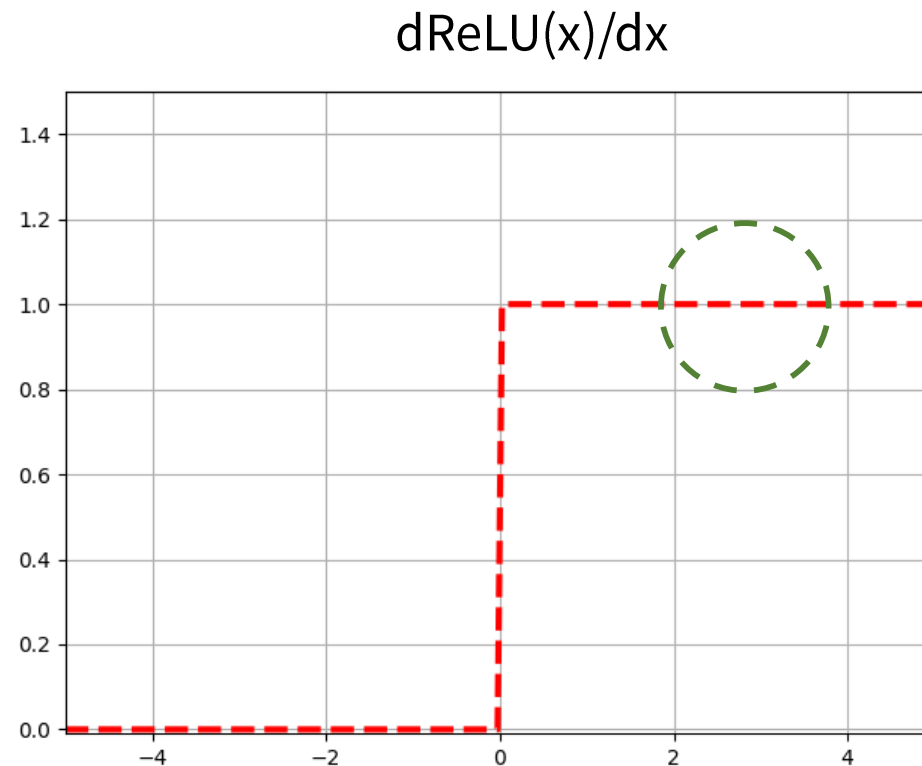
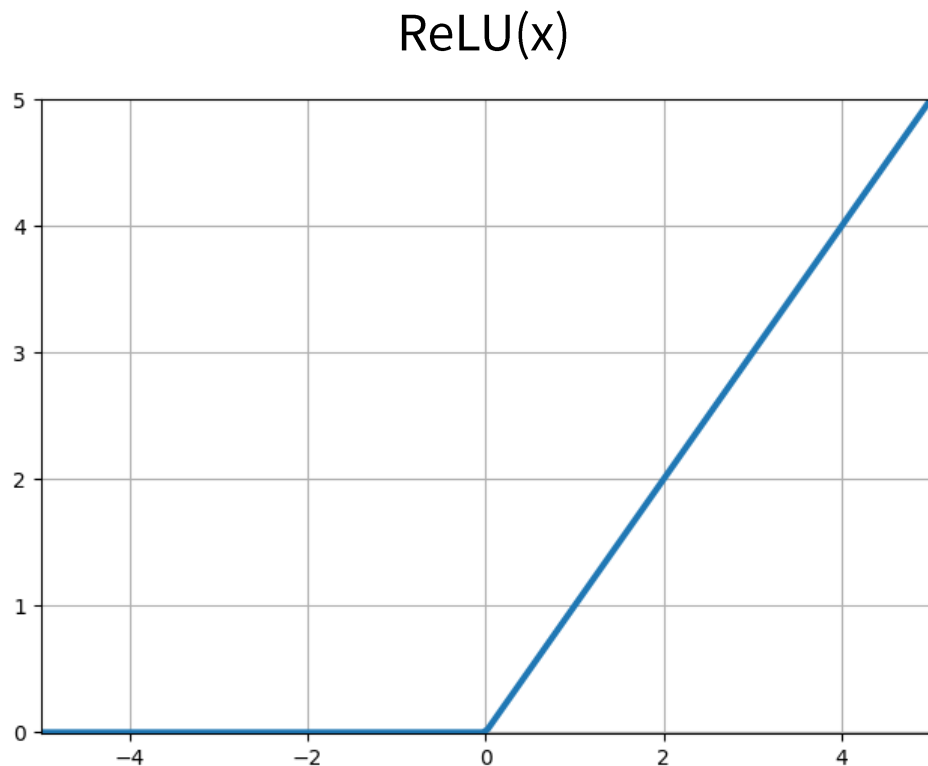
역전파 과정에서 Sigmoid 함수의 미분이 거듭 곱해지면서, 앞쪽 계층의 Gradient 값이 매우 작아진다.

I 활성화함수의 개선 – Hyperbolic Tangent



Sigmoid를 위아래로 늘려 (출력 범위가 2배) 문제를 해결하고자 하였다. 하지만 계층이 깊어지면 여전히 기울기 소실이 발생한다.

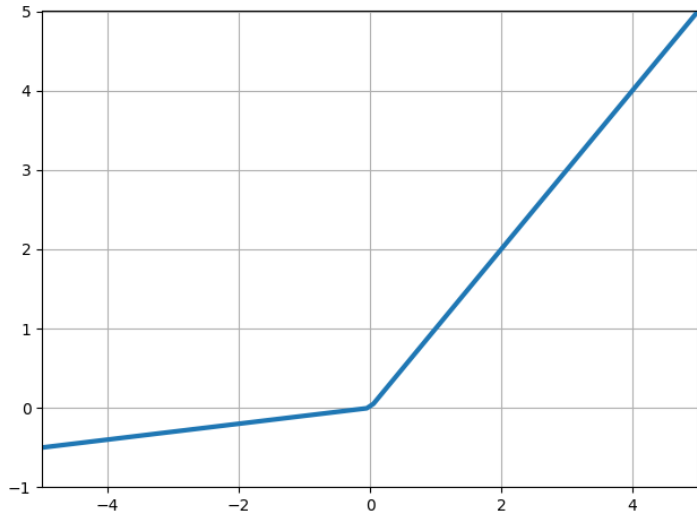
I 활성화함수의 개선 – Rectified Linear Unit (ReLU)



양수라면, 입력이 관계 없이 동일한 미분 값(1)이 나오게 하여 기울기 소실 문제를 해결. 간단하지만 뛰어난 결과를 보이면서 **딥러닝의 기본 활성화함수로 자리잡았다.**

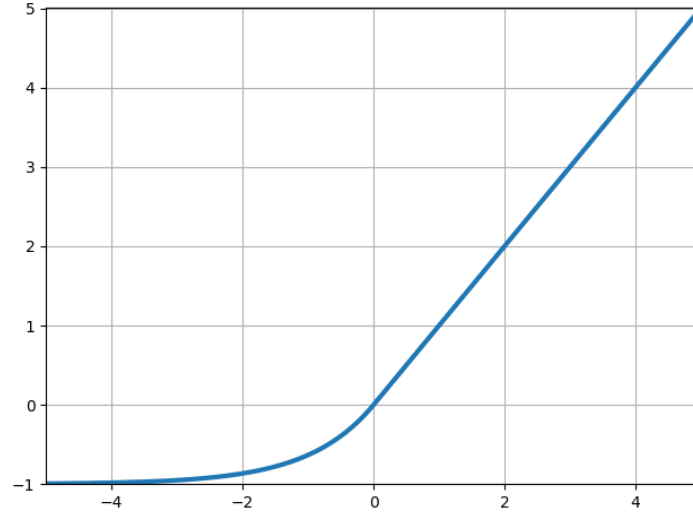
I ReLU의 변형

Leaky ReLU



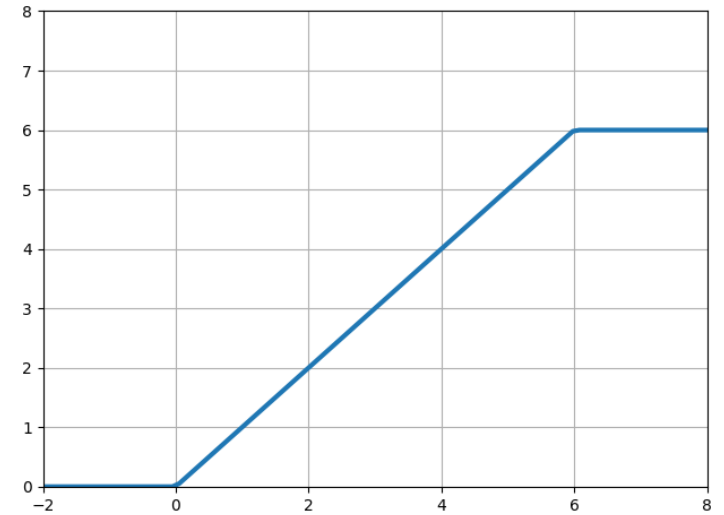
$$\max(0.1x, x)$$

ELU



$$\begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases}$$

ReLU6



$$\min(\max(0, x), 6)$$

다양한 변형된 함수들이 있으나, 특별한 이유가 없으면 보통 ReLU를 사용한다.