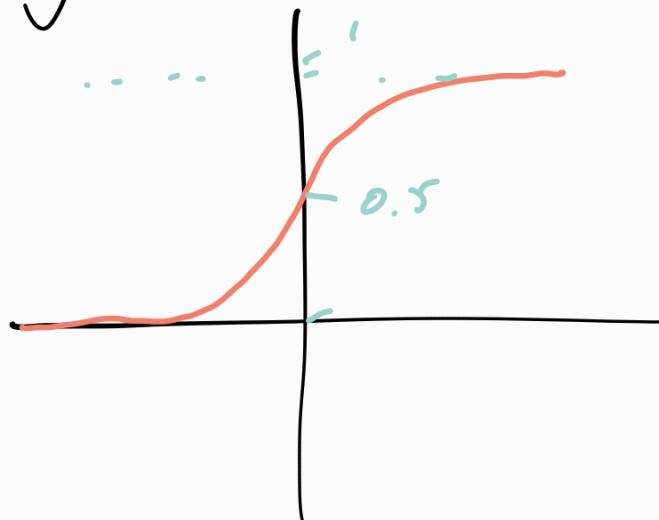


1. Sigmoid function

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



Step function izz 흐름에 대

: 모든 계층에서 비활성 기능

→ Backpropagation

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

$$= \frac{e^{-x}}{(e^x + 1)^2}$$

↑ 하위인 Sigmoid는 흐름에

Gradient Vanishing을 흐름에

0보다 작은 수를 계속 곱하기에, 흐름에

$\nabla S(x)$ 가 0으로 흐름에 → 흐름에 Sigmoid는

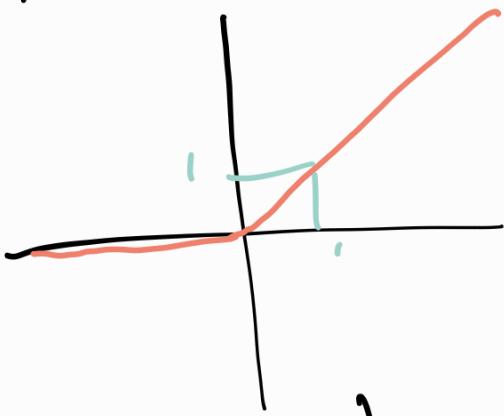
반대로 $x > 1$ 흐름에
Exploding gradient 흐름에
흐름에

형태 → 0.5 고정된 형태
0.5 고정된 형태
하는 문제
1.0 0.0

→ Gradient Clip / Norm²³
 Gradient L1 Explode \rightarrow 힘들어
 Gradient 합성

2. ReLU (Rectified Linear Unit)

$$\text{ReLU}(x) = \max(0, x)$$



ReLU Gradient Vanishing 문제 해결 가능함

$$\text{ReLU}'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

→ sigmoid $\rightarrow 0.00002 \times$

→ 무한대 / or 미분 불가능

But 'Dead ReLU' 문제가 등장함

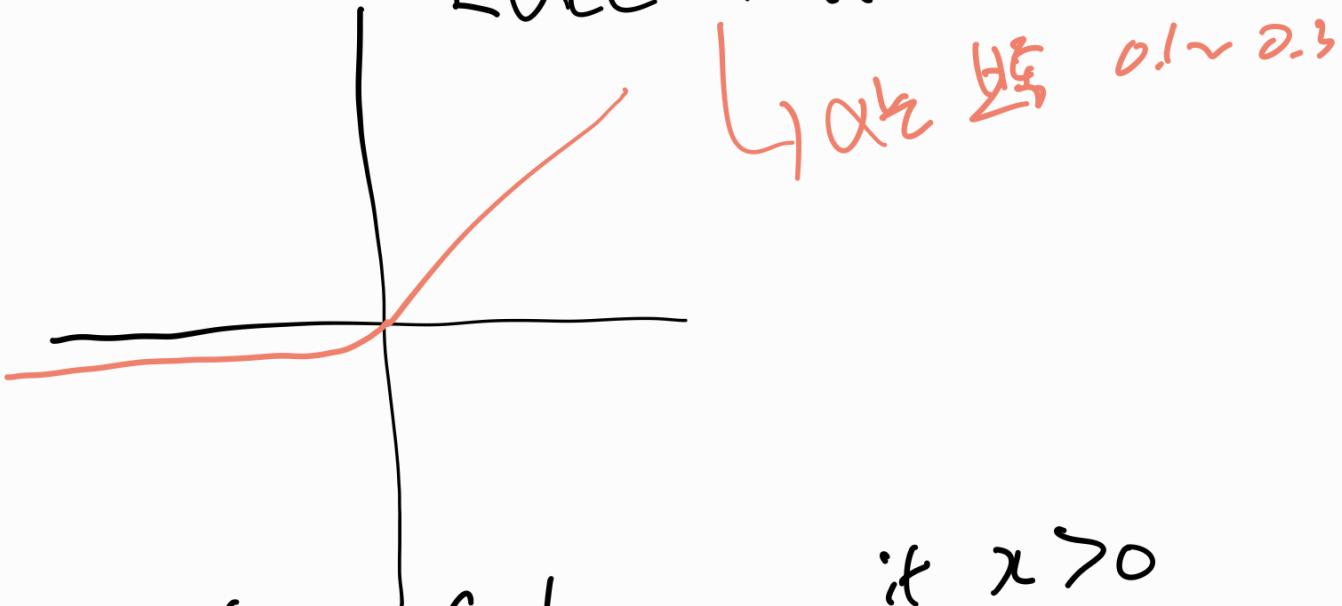
→ 대부분의 Unit들이 0 이하로 값이 초기화된
 Neural Net의 Sparse \rightarrow 차는 대수적

"간접적인 Sparse" \rightarrow 예상 \rightarrow 학습률 ↑
 But 대수 Sparse \rightarrow 차에 학습률 안정"

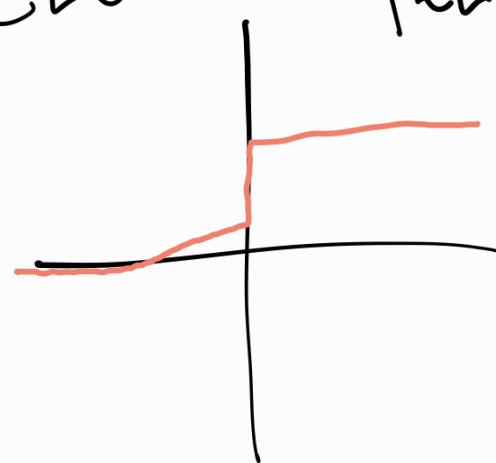
2번 Gradient Explode 9.2.7

3. ELU(Exponential Linear Unit)

$$ELU(x) = \begin{cases} a & \text{if } x > 0 \\ \alpha(e^{x-1}) & \text{if } x \leq 0 \end{cases}$$



$$ELU'(x) = \begin{cases} 1 & \text{if } x > 0 \\ ELU(x) + \alpha & \text{if } x \leq 0 \end{cases}$$



⇒ 음수의 Gradient는 0이 아님을 알 수 있음
Unit'd Output 0에 가까워질 때까지 계산
가능한 경우

Back/Gradient Exploding은 예전에 했던 모양

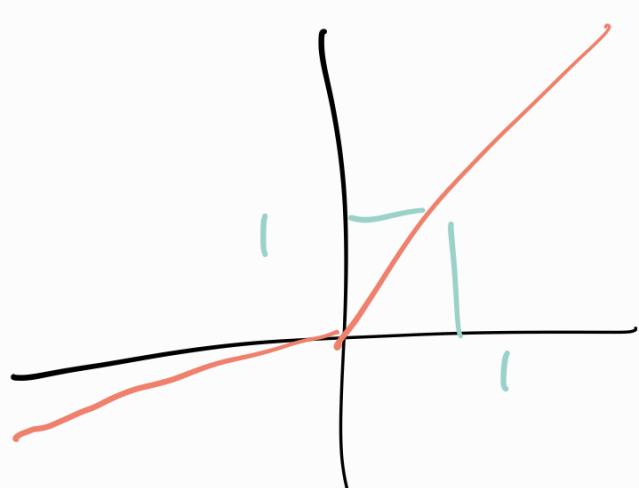
2. alpha를 0.1로 hyperparameter 설정

? Computation time?

5. Computation

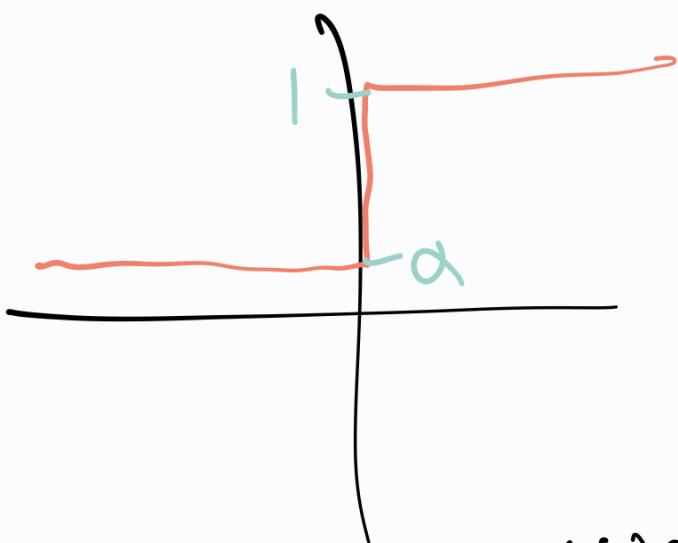
4. Leaky ReLU

$$\text{LReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{if } x \leq 0 \end{cases}$$



α : Normally small
above: ≈ 0.01

$$\text{LReLU}'(x) = \begin{cases} 1 & \text{if } x > 0 \\ \alpha & \text{if } x \leq 0 \end{cases}$$



ELU의 특성은 LReLU와 비교해

가장 큰 특징으로는 LReLU를 암시적

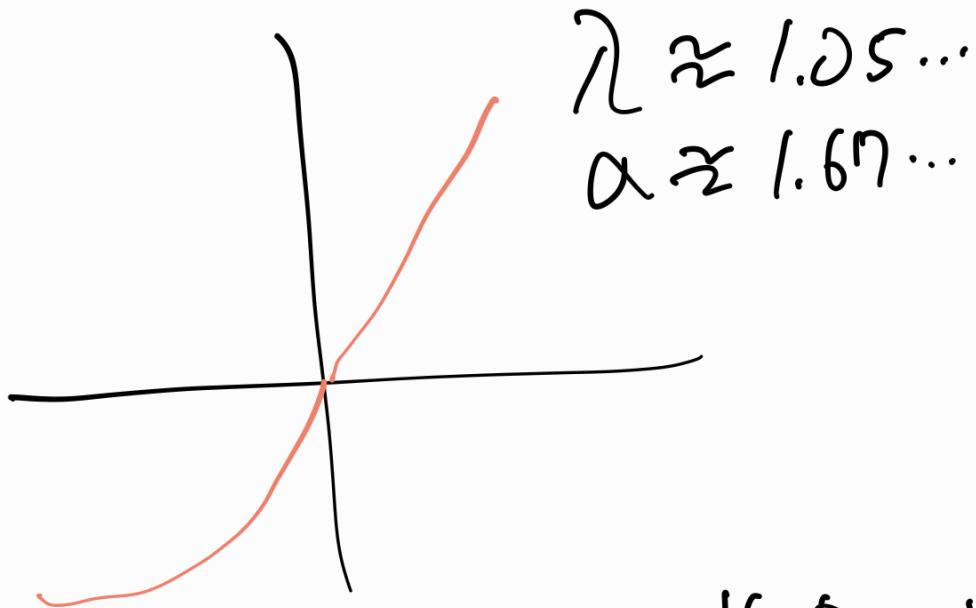
But 다른 중요한 Linear 특성

이 특성을 이용하는 차례

6. SELU (Scaled Exponential Linear Unit)

※ [Weight Initializer: Lecun-normal 사용
Dropout: Alpha Dropout 적용]

$$\text{SELU}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$



SELU는 Self Organizing Neural Network에 사용되는 활성화 함수입니다.

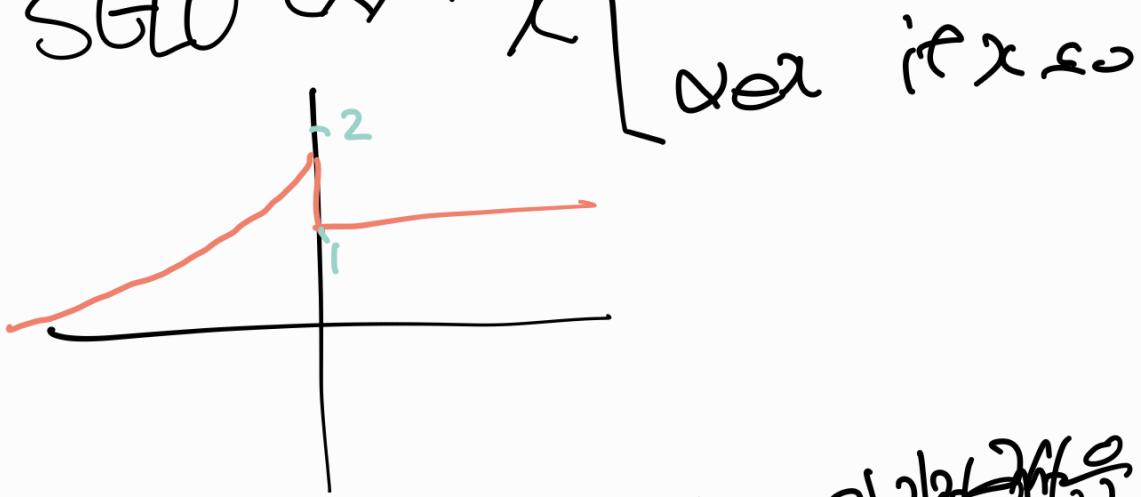
SELU는 정규 분포의 Gaussian distribution을 갖습니다.

위므로 \Rightarrow Internal Normalization

Batch Normalization 및 External Normalization

$$\rightarrow \mu \in [-0.1, 0.1] \quad \sigma \in [0.8, 1.5]$$

$$\text{CELU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases}$$



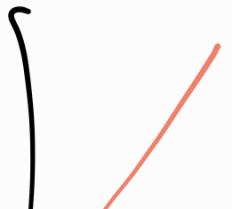
+ 그림 Dropout 설명 ~~지나가는 노드를 제거하는 것~~
 이는 \Rightarrow Dropout 전에 각 단위의 확률을 랜덤으로 0이나 1로 설정하는 것
 Dropout \neq Alpha dropout

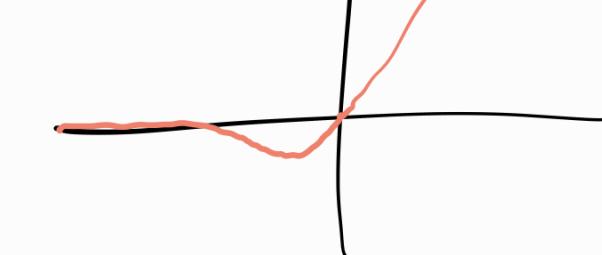
방법: External Normalization (Batch Normalization 대신 평균 제거)
 Gradient Vanishing/Exploding (평가 손실)

(예): LSTM의 예제
 ex) RNN, CNN

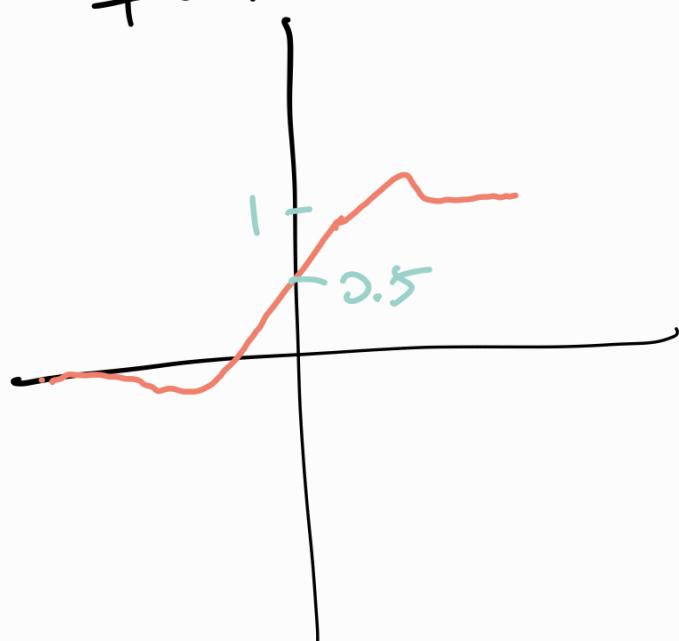
6. GELU (Gaussian Error Linear Unit)

$$GEIU(x) = 0.5x \left(1 + \tanh \left(\sqrt{2/\pi} (x + 0.044715x^3) \right) \right)$$





$$G_{GW}(x) = 0.5 \tanh(0.035677x) + 0.711485x \\ + (0.052516x^3 + 0.318442x) \text{sech}^2(0.035677x) + 0.711485x \\ + 0.5$$



\Rightarrow NLP of ~~is not unique~~.
 Gradient Vanishes +
 One ~~real~~ local minimum of $f(x)$ if $\nabla f(x) = 0$