실전 기계학습 Lab #P02 보고서

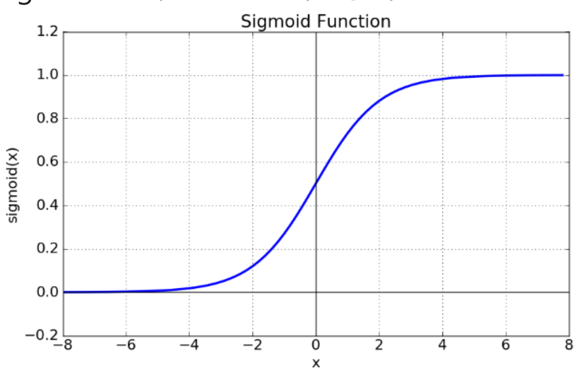
2018102111 소프트웨어융합학과 서보민

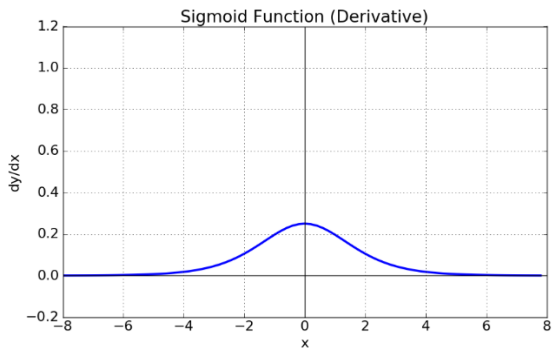
6. [Exercise 6-1] Lecture 06, pp 20 의 코드를 사용하여 아래 문제를 해결하시오

1) 아래의 Activation function 함수들에 대하여, 함수 형태, 함수의 미분 형태, 특성을 조사하시오

Sigmoid

- sigmoid 함수는 s 자와 유사한 완만한 시그모이드 커브 형태를 보이는 함수입니다.
- sigmoid(x) = 1 / 1+ e^(-x)
- sigmoid 함수는 모든 실수 입력 값을 0 과 1 사이의 미분가능한 수로 변환하는 특징을 가집니다.
- sigmoid 함수는 음수 값을 0 에 가깝게 표현하기 때문에 계층이 많아질수록 영향이 적어지는 vanishing gradient problem 을 가지고 있으며, 이 문제로 인해 실무에서 사용되지 않고 있습니다
- Sigmoid 함수 및 도함수 형태

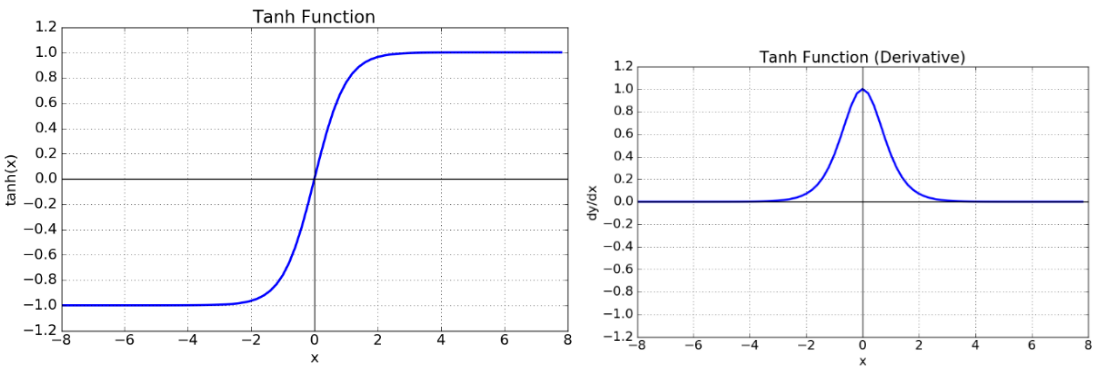




-

Tanh

- 쌍곡선 함수 중 하나로, sigmoid 함수를 변형해서 얻을 수 있습니다.
- tanh(x) = e^x - e^(-x) / e^x + e^(-x)
- tanh 함수는 함수의 중심점을 0 으로 옮겨 sigmoid 가 갖는 최적화 과정에서 느려지는 문제점을 해결하였습니다.
- 미분함수에서 일정 값이상에서 미분 값이 소실되는 문제점은 여전히 가지고 있습니다
- tanh 함수 및 도함수 형태



HardTanh

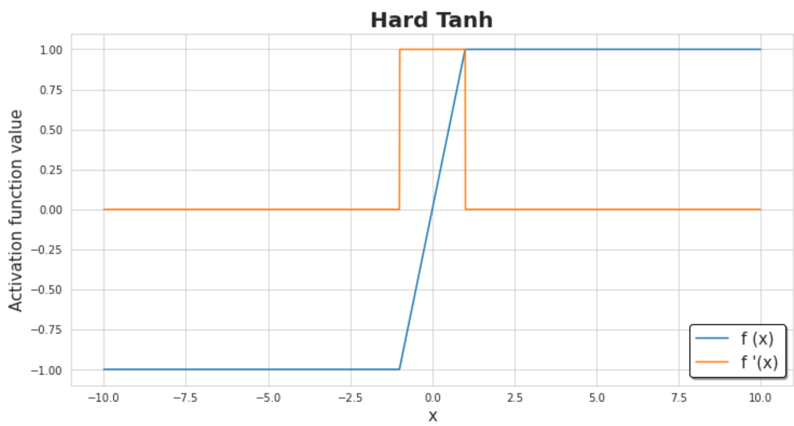
- Tanh 함수보다 각진 형태의 함수입니다.
- 입력값이 1 보다 크면 1 을 출력하고 입력값이 -1 보다 작으면 -1 을 출력하면 그 외에는 입력값을 출력합니다.

$$f(x) = 1 \ \ if \ \ x > 1$$

$$f(x) = -1 \ \ if \ \ x < -1$$

$$f(x) = x \ \ otherwise$$

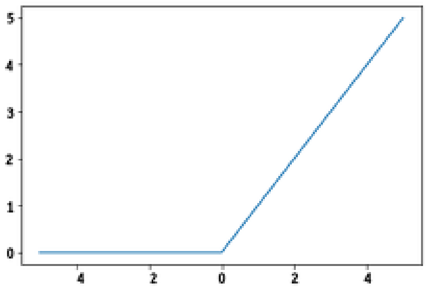
-



-

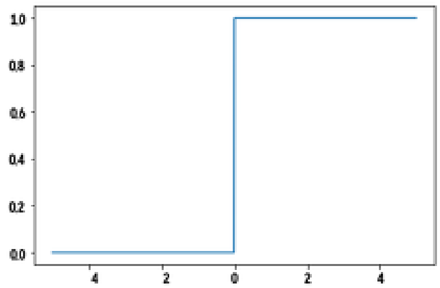
ReLU

- vanishing gradient 문제를 해결하기 위해 고안되었습니다.
- relu 의 도함수는 0 또는 1 의 값을 가지기 때문에 컴퓨팅 자원 면에서 효율적이며 최대값이 1 보다 큰 값도 가능하여 학습이 빠르다는 장점을 가지고 있습니다.
- 0 보다 작은 경우에는 0 을 반환하고 0 을 반환한 노드는 다시 값을 가지기 어렵다는 문제를 가지고 있습니다.
- ReLU 의 함수 및 도함수 형태



$$ReLU(x) = \max(0,\ x)$$

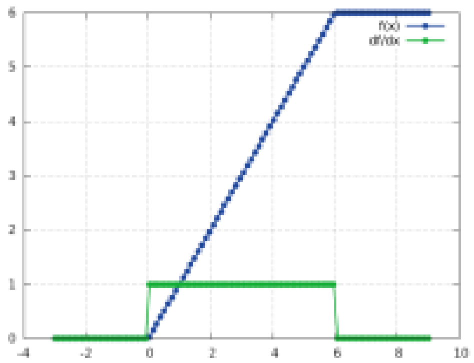
Rectified Linear Unit



$$R'(y) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

ReLU6

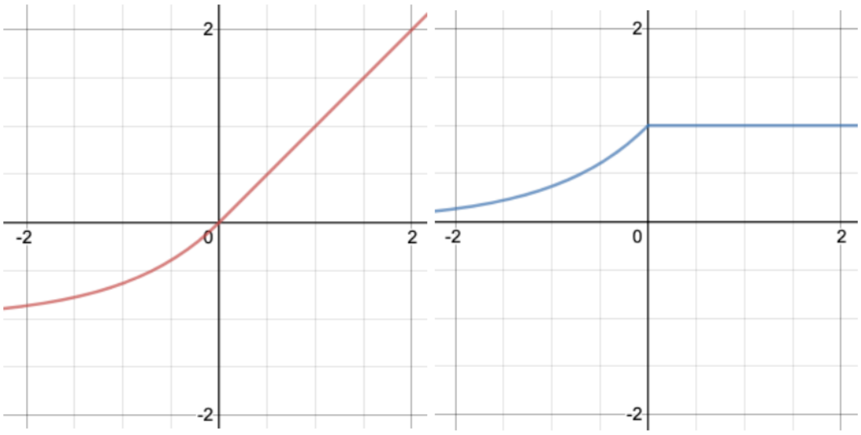
- ReLU 의 변형으로 6 의 상한값을 설정합니다.
- 정확도 면에서 장점을 가지고 있으며, 상한값이 6 이기 때문에 3bit 만 사용하여 표현이 가능합니다.
- ReLU6 의 함수 및 도함수 형태



ELU & SELU

- ReLU와 비슷한 형태를 가지면 지수를 이용하여 음수일 경우 0이 아닌 완만한 곡선으로 표현합니다.
- 출력값이 0에 가까우며 exp함수의 비용이 별도로 발생합니다.
- 음수의 표현 a(e^x – 1)로 표현하며 a = 1일 때 ELU, 그 외의 값일 때 SELU라고 지칭합니다.
- 함수와 도함수의 형태

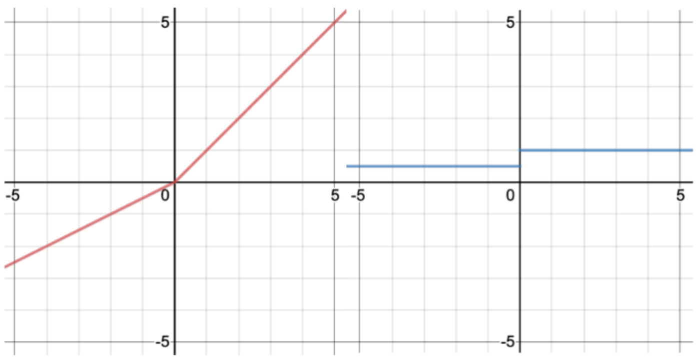
$$f(\alpha,x)=\begin{cases} x & (x>0) \\ \alpha(e^{x}-1) & (x\le 0)\end{cases} \qquad f'(\alpha,x)=\begin{cases} 1 & (x>0) \\ f(\alpha,x)+\alpha & (x\le 0)\end{cases}$$



LeakyReLU & PReLU

- LeakyReLU는 ReLU가 가지는 Dying ReLU현상을 해결하기 위해 고안되었습니다.
- LeakyReLU는 ReLU와 비슷한 형태이지만 음수의 표현에서 0이 아닌 ax형태로 표현함으로써 미분값이 0이 아닌 값인 작은 값으로 표현됩니다.
- LeakyReLU는 알파값으로 고정된 상수를 사용하며 PReLU는 알파값을 학습이 가능한 파라미터로 설정한 방법입니다.

$$f(\alpha,x)=\begin{cases} \alpha x & (x<0) \\ x & (x\ge 0)\end{cases} \qquad f'(x)=\begin{cases} \alpha & (x<0) \\ 1 & (x\ge 0)\end{cases}$$



-

threshold

- 이진함수의 형태를 가지고 있습니다.
- 입력값이 임계값보다 높으면 1을 출력하면 아니면 0을 출력합니다,
-