

2-7. 특이함수의 미분

시그모이드함수의 미분

$$\zeta_a(x) = \frac{1}{1 + \exp(-ax)}$$

$$\frac{d\zeta_a(x)}{dx} = \frac{ax \exp(-ax)}{\{1 + \exp(-ax)\}^2} = a\zeta_a(x) \{1 - \zeta_a(x)\}$$

시그모이드함수의 2계미분

$$\frac{d^2\zeta_a(x)}{dx^2} = \frac{d[a\zeta_a(x) \{1 - \zeta_a(x)\}]}{dx}$$

$$= \underbrace{a \frac{d\zeta_a(x)}{dx}}_{\text{1계 미분}} \{1 - \zeta_a(x)\} + a\zeta_a(x) \underbrace{\frac{d\{1 - \zeta_a(x)\}}{dx}}_{\text{1계 미분}}$$

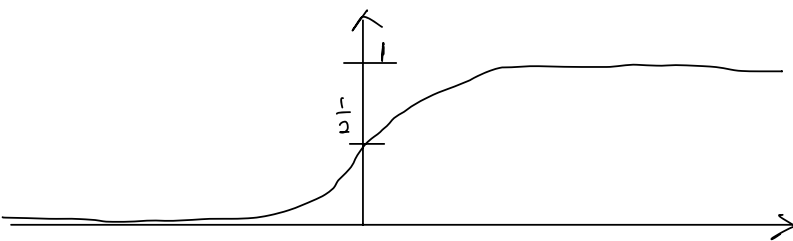
$$= a \frac{d\zeta_a(x)}{dx} \{1 - \zeta_a(x)\} - a\zeta_a(x) \frac{d\zeta_a(x)}{dx}$$

$$= a \frac{d\zeta_a(x)}{dx} \{1 - 2\zeta_a(x)\}$$

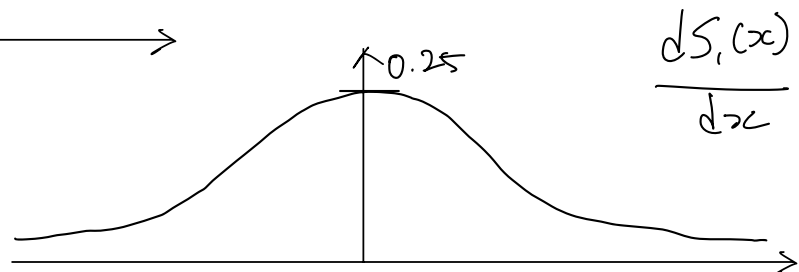
$$= a^2 \zeta_a(x) \{1 - \zeta_a(x)\} \{1 - 2\zeta_a(x)\}$$

표준 시그모이드함수

$\hookrightarrow a = 1$ 일 때



표준 시그모이드함수의 1계미분



ReLU 함수

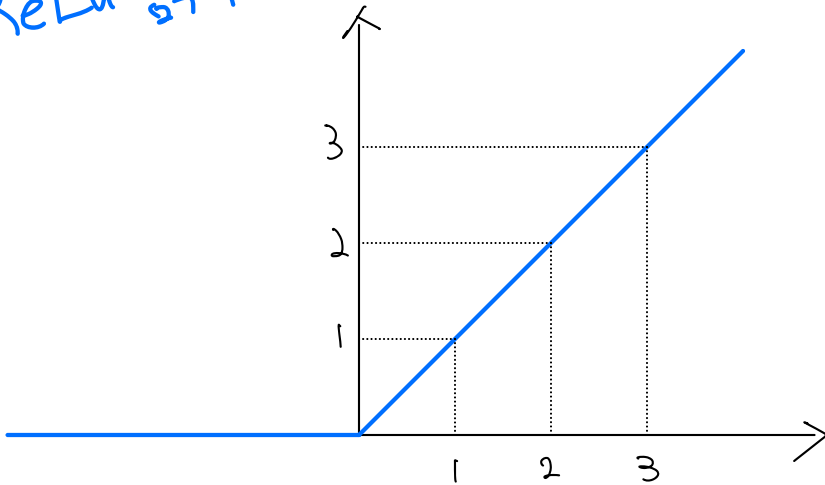
↳ 신경망의 표현력을 높일 수 있는 함수 중 하나

$$\text{ReLU} = \varphi(x)$$

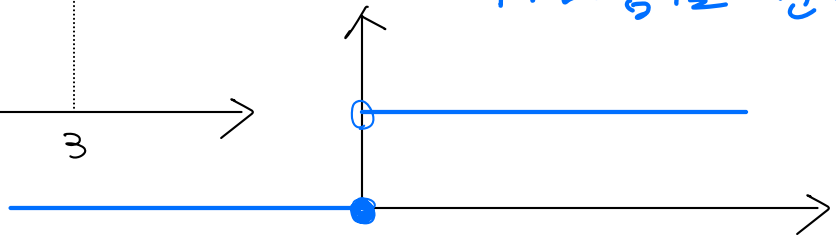
$$\text{공식 } \varphi(x) = \max(0, x) = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

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

ReLU 함수의 그래프



ReLU 함수를 미분한 그래프



인공지능에서?

- 신경망에서 가중치(w)의 조정량은 오차값을 기준으로 편미분한 것을 사용
활성화 함수도 미분해서 사용
- 따라서, 최근 시그모이드 함수를 미분하면 최대값이 0.25 가 되는데 신경망의 계층이 많아질수록, 오차역전파법에서 오차가 전파되기 어려워지는 상황이 발생할 수도 있음
↳ 기울기 소실 (Vanishing Gradient) 문제
- 기울기 소실 문제를 해결하기 위한 것이 ReLU 함수
ReLU 함수를 미분하면 0, 1 중 하나의 값을 갖게 되고,
이로써 기울기 소실 문제 해결에 도움을 줌
그래서 최근의 신경망에서는 활성화 함수로 ReLU를 많이 사용함