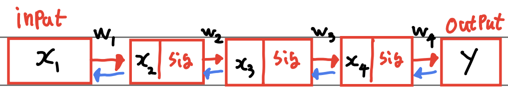
Sigmoid

- Back propagation 할때 변화량 계산때문에 시그모이드를 사용했다.

식 $\frac{1}{(1+e^{-x})}$ 미분식 $\frac{e^{-x}}{(1+e^{-x})}$

문제점 Vanishing Gradient



• 역전파를 진행 하면 그림 처럼 뒤로 움직이면서 weight 뿐만아니라 활성화 함수도 미분 해야한다.

이때 업데이트 공식을 보면서 설명하겠다

예) w_3 업데이트

$$W_3 = W_3 - \alpha \frac{\partial y}{\partial w_3}$$

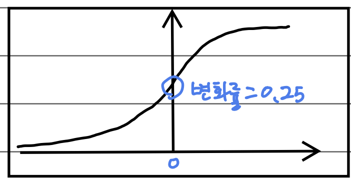
$$= W_3 - \alpha \left[\frac{\partial y}{\partial sig} \cdot \boxed{\frac{\partial sig}{\partial x_4}} \cdot \frac{\partial x_4}{w_3} \right]$$

이부분에서 시그마 함수를 x_4의 변화에 따른 변화량을 계산한다.

근데 sig의 최대 변화는 '0'인 부분에서 가장 크게 일어난다.

'0'인 부분의 변화량은 $\frac{1}{4}$ 이값이 deep한 모델로 가면 '0'으로 수렴하고

이렇게 되면 w_1 업데이트가 거의 이루어 지지 않는다.

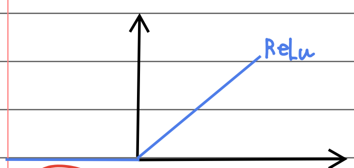


그래서 ReLu 함수등장

Relu

- 기울기가 1,0

식 $h(x) = \begin{cases} x & (x>0) \\ 0 & (x \le 0) \end{cases}$ 미분식 $h'(x) = \begin{cases} 1 & (x>0) \\ 0 & (x \le 0) \end{cases}$



- $y=x$ 라는 함수가 활성화함수가 안되는 이유 ➡ DNN 의 은닉층이 깊어 질수가 없다.

(아래 그림에 나오는것처럼 선형으로만 표현된다.)

└ 왜 굳이 비선형인가?

• $y=x$

$y = xw_1 \cdot W_2 \cdot W_3$

$= xw_1w_2w_3$ 처럼 그냥 hidden 값 하나로 표현이 가능하다.

• 선형으로 비선형을 표현할수있다.

그러면 엄청많은 수식이 필요해서

차라리 적은 수식으로 표현가능한 비선형을

처음부터 뽑을려고 한다.