

⊛ 2단 - 3단의 NW 정도면

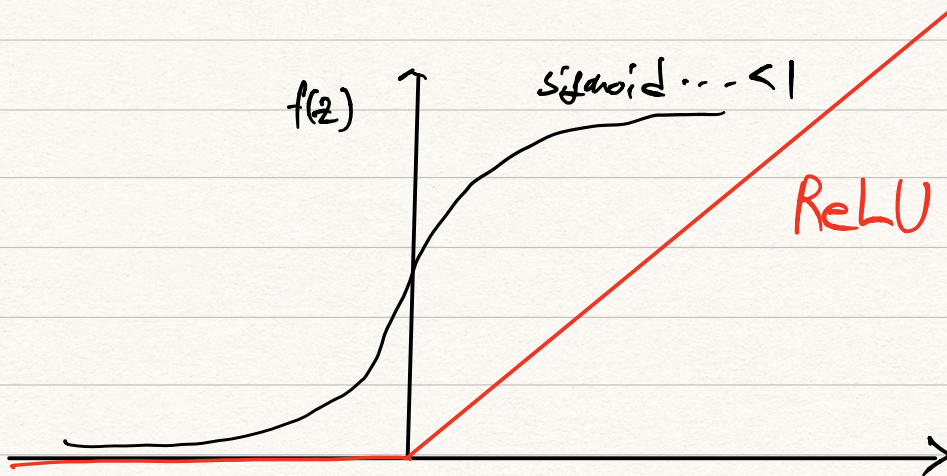
Back propagation으로 해결 가능함.

But 9단 10단 정도 가면 accuracy가 확↓

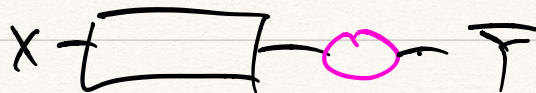
⇒ BP중에 Sigmoid를 통과하다 실패

chain rule 상에서... 0.0000001 이 관측어  
된다.

→ Vanishing gradient 문제이다.



⊛ ReLU = Rectified Linear Unit



ReLU 적용할 때도 '마지막 단'은 Sigmoid 사용해!