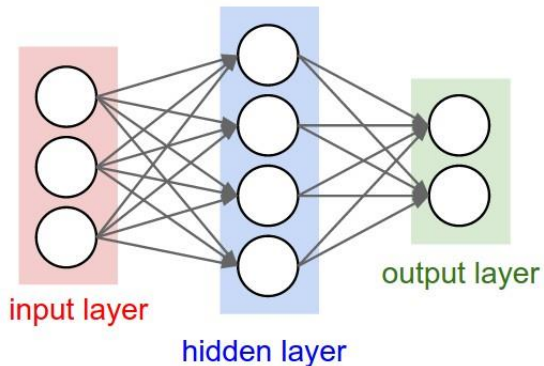


Ch3 신경망 모델 구현하기

- 활성화 함수

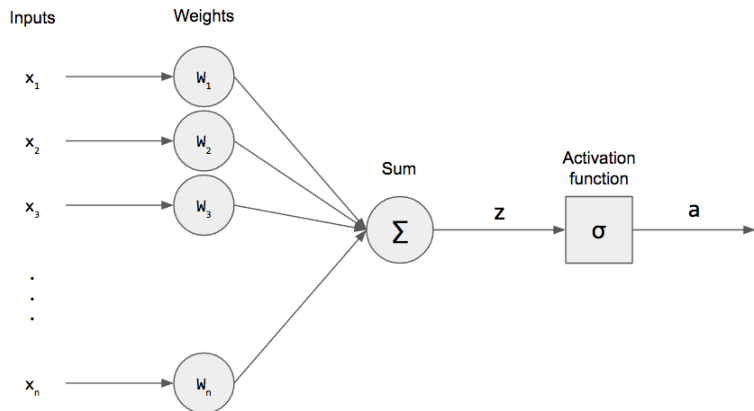
인공신경망과 활성화 함수



입력층 (input layer) 은 신호를 입력 받아 다음 노드에 전달한다.
입력층을 거치면 은닉층(hidden layer) 을 지나 마지막 뉴런인
출력층(output layer) 로 전달된다.

인공 신경망에선 입력 받은 신호에 특정한 수학 연산을 실행
⇒ 각 층에 존재하는 매개변수 가중치(weight)에 행렬곱 시킨 후 편향(bias) 를 더한다.

* 가중치 : 입력 신호가 출력에 주는 영향을 계산하는 매개변수
편향 : 각 노드가 데이터에 얼마나 민감한지 알려주는 매개변수



이 행렬곱의 결과가 활성화 함수 (Activation Function) 을 거치면
인공 뉴런의 결과값을 산출한다.

활성화 함수 -> 입력에 적절한 처리를 해 출력 신호로 변환하는 함수
입력 신호의 합이 활성화를 일으키는지 아닌지 정하는 역할

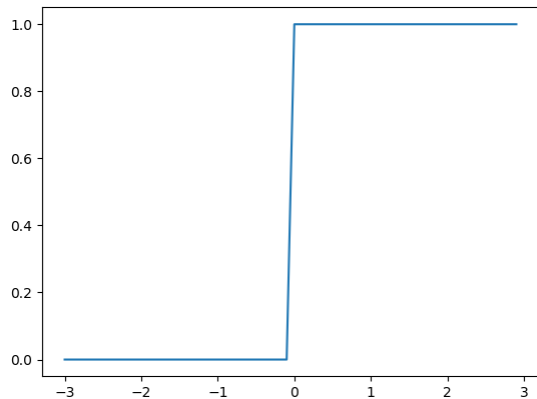
이 결과값은 은닉층의 인공뉴런으로 전달 -> 한번 더 가중치 곱 + 활성화 함수
몇 겹에 걸쳐 이 과정 반복 => 출력층에서 최종 결과값 생성

활성화 함수의 종류

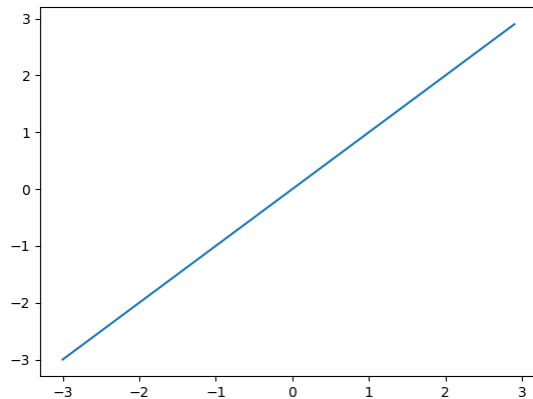
1. 이진 활성화 함수 (Binary step activation function) => 잘 안쓰임 (다중 출력 불가)

2. 선형 활성화 함수 (Linear activation function) => 잘 안쓰임

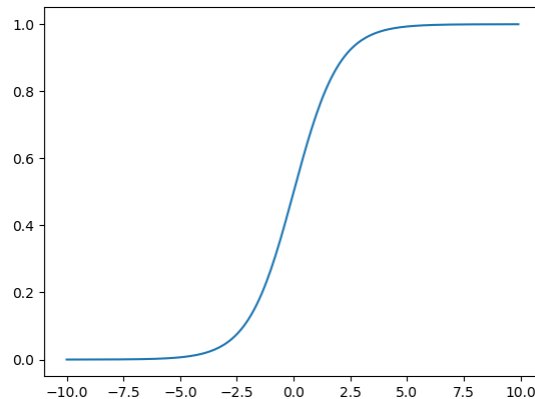
3. 비선형 활성화 함수 (Non - linear activation function)



1. 이진 활성화 함수 (Binary step activation function)

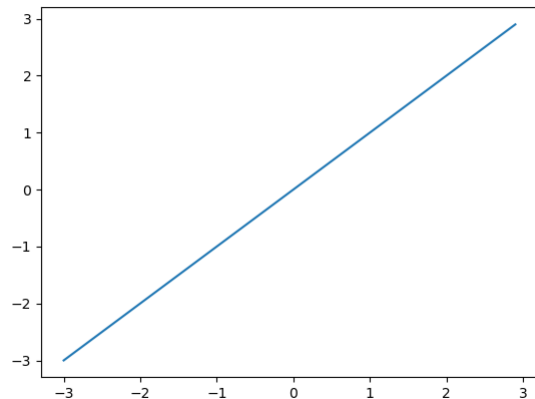


2. 선형 활성화 함수 (Linear activation function)



3. 비선형 활성화 함수 (Non - linear activation function)

선형 활성화 함수



$$h(x) = ax$$

치명적인 단점

1. 신경망 층을 깊게 쌓는 이유가 없어짐

ex) 활성화 함수를 $y = 3x$ 라는 선형 함수라 하자.

3개의 층으로 구성 했을 때, $y = 3(3(3(x)))$

이는 결국 $y = 27x$ 랑 같다.

3층이나 쌓아도 결국 1층만 사용한 것이라 같아짐

비선형 활성화 함수

장점

1. 역전파 알고리즘(Backpropagation) 사용 가능
2. 다중 출력 가능

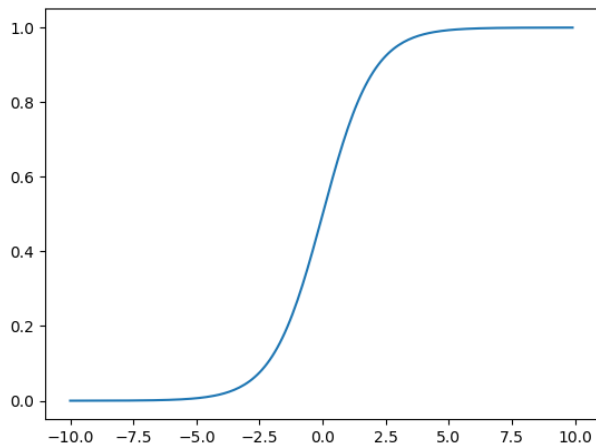
종류

1. Sigmoid
 2. Hyperbolic tangent
 3. ReLu(Rectified Linear Unit)
- ... 이외에도 여러가지 비선형 활성화 함수 존재

* 역전파 알고리즘

인공 신경망의 출력 결과값과 정답을 비교해 오차 계산
오차를 기반으로 신경망 전체를 학습시키기 위해 경사하강법을
이용해 출력층 가중치부터 입력층 가중치까지 뒤에서부터 차례
대로 조정하여 최적화 시키는 알고리즘

Sigmoid activation function



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

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

장점

1. 출력값의 범위가 0~1 사이로 제한

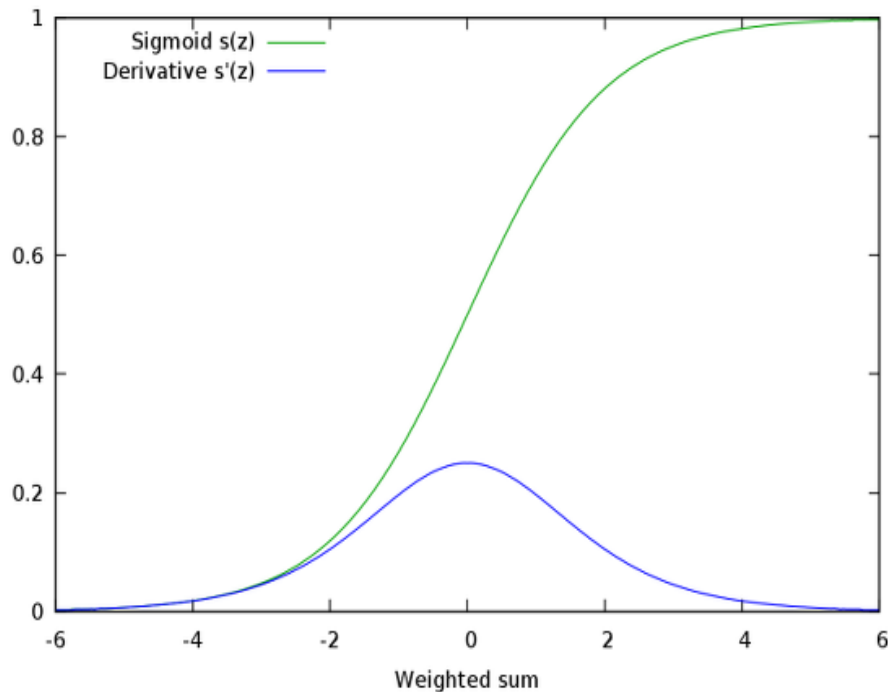
=> classification 문제, 확률 문제에서 유용

2. 단순한 미분식

단점

1. vanishing gradient problem

Vanishing gradient problem



Sigmoid 함수를 미분하면 최대 $\rightarrow 0.25$

양 끝으로 갈 수록 0에 가까워짐

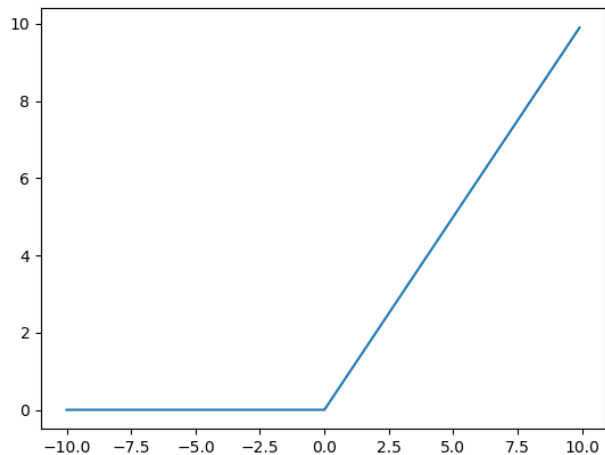
역전파 알고리즘 수행 시

Gradient 가 0에 수렴해 계산이 어려워짐

\Rightarrow Vanishing gradient problem

**\Rightarrow 이런 단점으로 보통 출력층에 sigmoid
를 쓰거나 아예 사용하지 않는 추세**

ReLU activation function



$$h(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

장점

1. 빠른 연산
2. 0 이상인 곳에선 수렴하는 구간이 없음
=> vanishing gradient problem X

단점

1. Dead ReLU

- => 값이 음수일 경우 0만 출력 하여 학습 불가
=> 대안 Leaky ReLU, PReLU, ELU ...



Thank you