

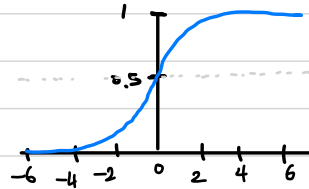
# Deep Learning

## 1. 알고있는 Activation function에 대해 알려주세요 (Sigmoid, ReLU, LeakyReLU, Tanh 등)

**Sigmoid function**은 Logistic 함수라고 불리기도 하며 미분이 되지 않는 계단 함수는 미분 가능하도록 곡선화를 적용한 함수이다.  
다층 퍼셉트론에서 비선형 값을 얻기 위해 사용하게 시작했다.

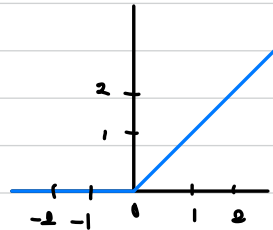
특징: 출력값은 0이로 보임

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



**ReLU function (Rectified Linear Unit)** 이란  $f(x) = \max(0, x)$  함수로, 정의해 보면  
"어떤 수가 0보다 크면 수가 출력, 어떤 수가 0보다 작으면 0으로 출력" 하는 함수를 의미한다.  
특징: 양의 값에서 Saturated 되지 않음. 계산 효율이 뛰어나.

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

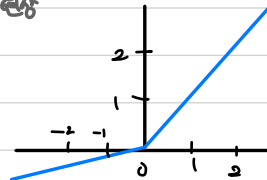


**Leaky ReLU**은 ReLU가 갖는 Dying ReLU을 해결하기 위해 나온 함수이다.

→ 음의 값을 가지면 전부 0으로 출력하기 때문에 점점 weight들이 업데이트 되지 않는 현상

특징: Dying ReLU 현상 x

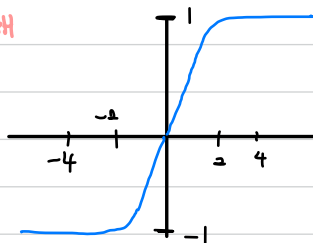
$$h(x) = \begin{cases} x & (x > 0) \\ 0.01x & (x < 0) \end{cases}$$



**Tanh function** 은 Sigmoid function의 함수값의 중간이 0이어서 한층 더 나은 단점을 극복한 함수이다.

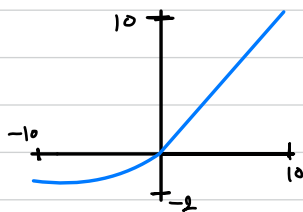
특징: 출력값은 -1 ~ 1로 압축. Dying gradient 문제

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



**ELU (Exponential Linear Units)** 은 ReLU와 Leaky ReLU의 중간점(?)이다. ELU는 Leaky ReLU처럼 zero-mean의 출력을 내지만 Saturation 관점에서는 ReLU의 특징도 갖는다.

$$f(x) = \begin{cases} x & (x > 0) \\ \alpha(\exp(x)-1) & (x < 0) \end{cases}$$



**Max out** 은 앞서 나온 activation function과는 다른 성격의 함수이다. Saturated Regime이 없으며 gradient가 0이 되는 지점이 없다는 장점이 있다. 하지만 parameter 수가 두배가 된다는 단점이 있다.

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

두 값중 최대값 선택

# 오버피팅인 경우 어떻게 해결할까요?

## 1. Dataset augmentation

Training data의 개수를 인위적으로 늘려 data의 양을 늘리는 기법.

→ 데이터의 특징은 고쳐야 한다.

## 2. parameter Norm penalties

cost function에 제곱을 더하거나 (L2) 절대값을 더하여 (L1) 웨이트의 크기에 제한을 준다.

• L2 weight decay (제곱값)

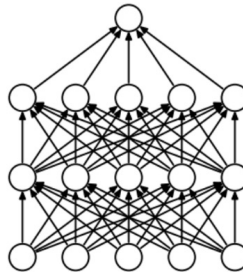
• L1 weight decay (절대값)

L2파라미터는 ridge 회귀분석 또는  
Tikhonov 정칙화로 알려져 있음.

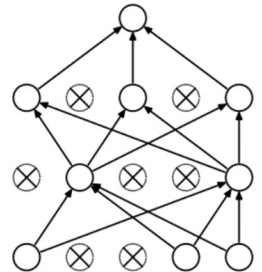
LASSO는 선형모델의 L1 패널티와  
최소제곱법 (LSM)을 합친 모델

## 3. Dropout

계층마다 일정 비율의 뉴런을 랜덤하게 온/오프  
drop 시키고 나머지 뉴런을 학습하는 방법이다.  
dropout을 통해 앙상블 학습처럼 여러  
모델을 학습시킨 것과 같은 효과를 주어  
오버피팅을 해결할 수 있다.



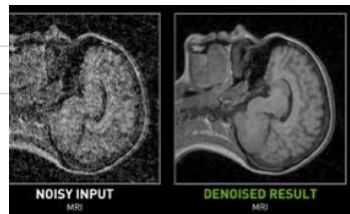
(a) Standard Neural Net



(b) After applying dropout.

## 4. Noise Rubustness

노이즈나 이상치 (outlier) 같은 엉뚱한 데이터가 들어와도 흔들리지 않는 모델을 만들기 위한 방법으로 인위적  
노이즈를 주는 방법.

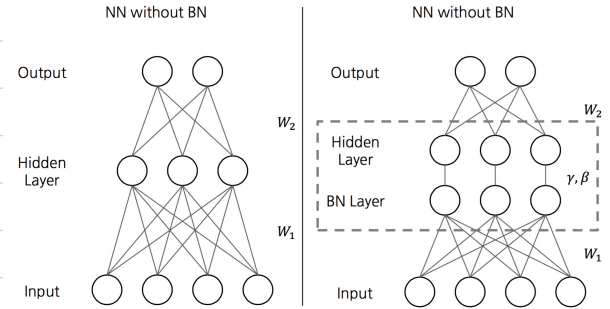


## 5. Batch Normalization

활성과 값이 적절하게 분포되도록 하는 값을 좋은 가중치의 초기 값으로 본다.

가중치의 초기값에 의존하지 않고 활성화 값을 강제로 적절히 분포되도록 하는 것은 배치 정규화라고 한다.

→ why? ....



## 6. Early Stopping

Epoch가 늘어날수록 Training Error는 줄어들지만, Testing Error는 증가하여 overfitting이 발생할 수 있다. 따라서 이전 Epoch에 비해 오차가 증가하면 overfitting이 발생하기 전에 학습을 멈추는 방법.

