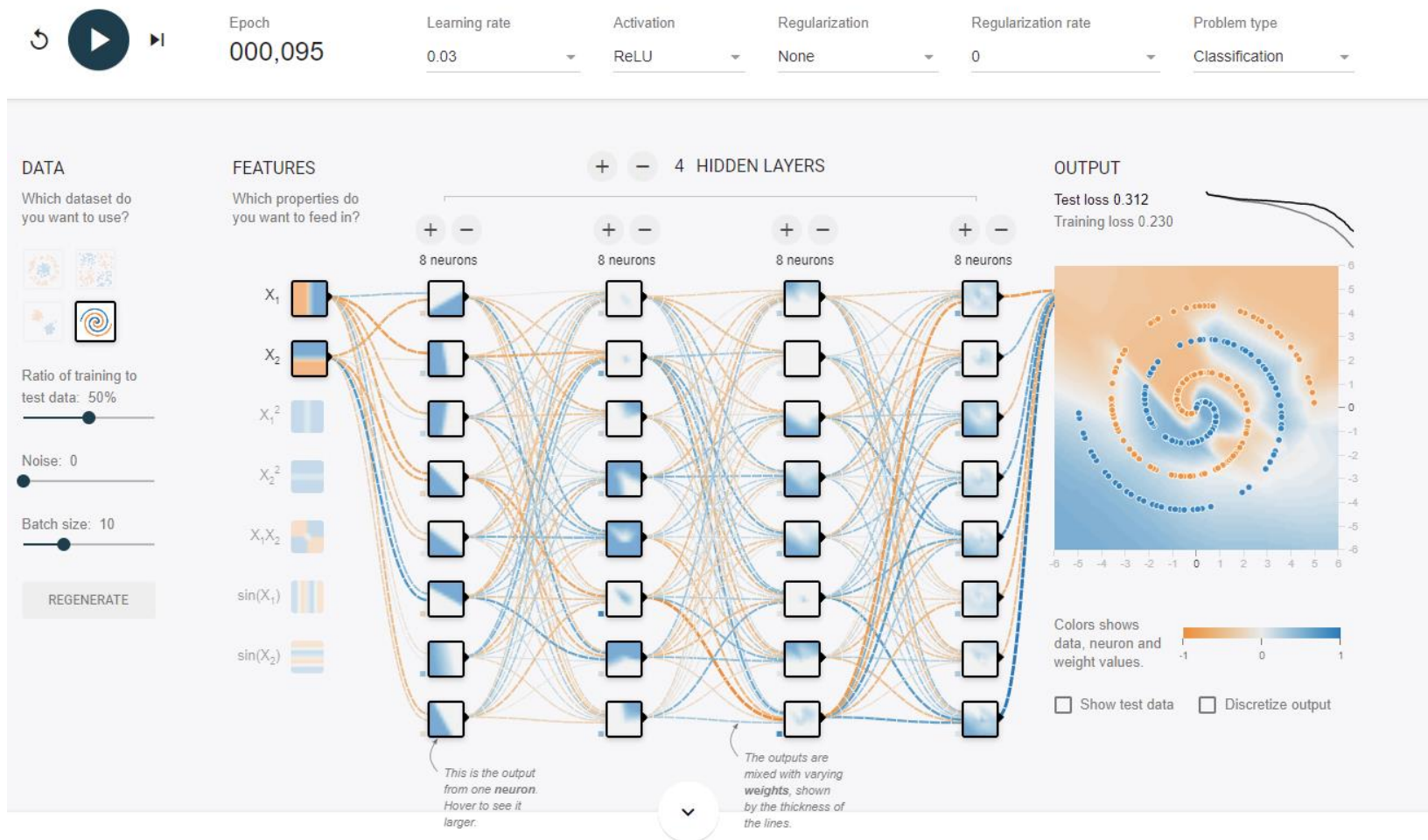

TAVE 8기 전반기 13팀 ‘딥러닝 스터디 '10-4 연습문제’

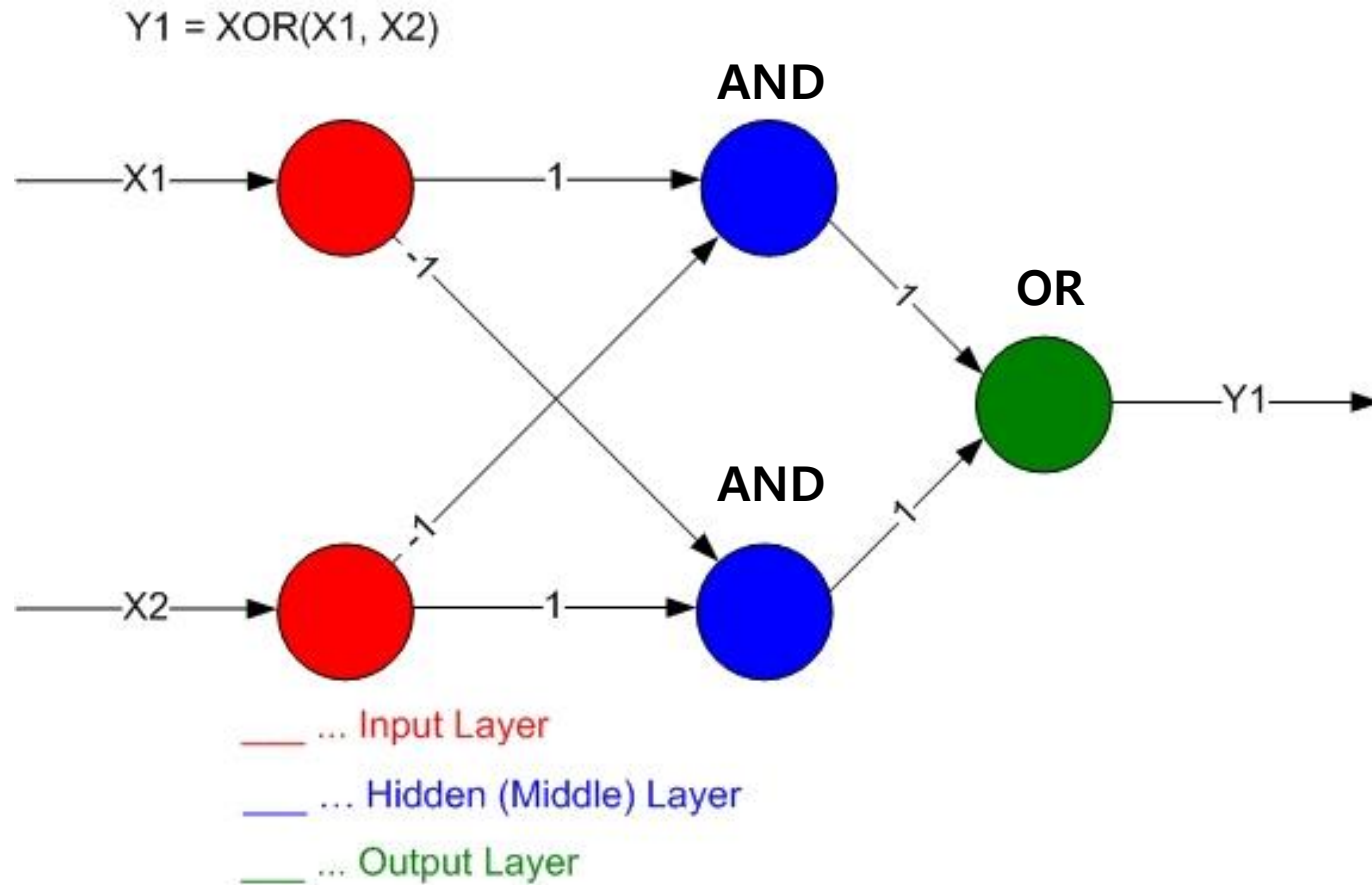
TAVE Research Group
Lee Moongi

01 텐서플로우 플레이그라운드

<https://playground.tensorflow.org/>



02 A(+)B (XOR 연산) 인공신경망



03 퍼셉트론<로지스틱? 로지스틱으로 만들려면?

■ 책 360쪽

→ 퍼셉트론은 클래스 확률을 제공하지 않으며 고정된 임계값을 기준으로 예측을 만듦

■ 책 362 ~ 4쪽

→ 활성화 함수를 계단함수에서 로지스틱 활성화 함수로 변환

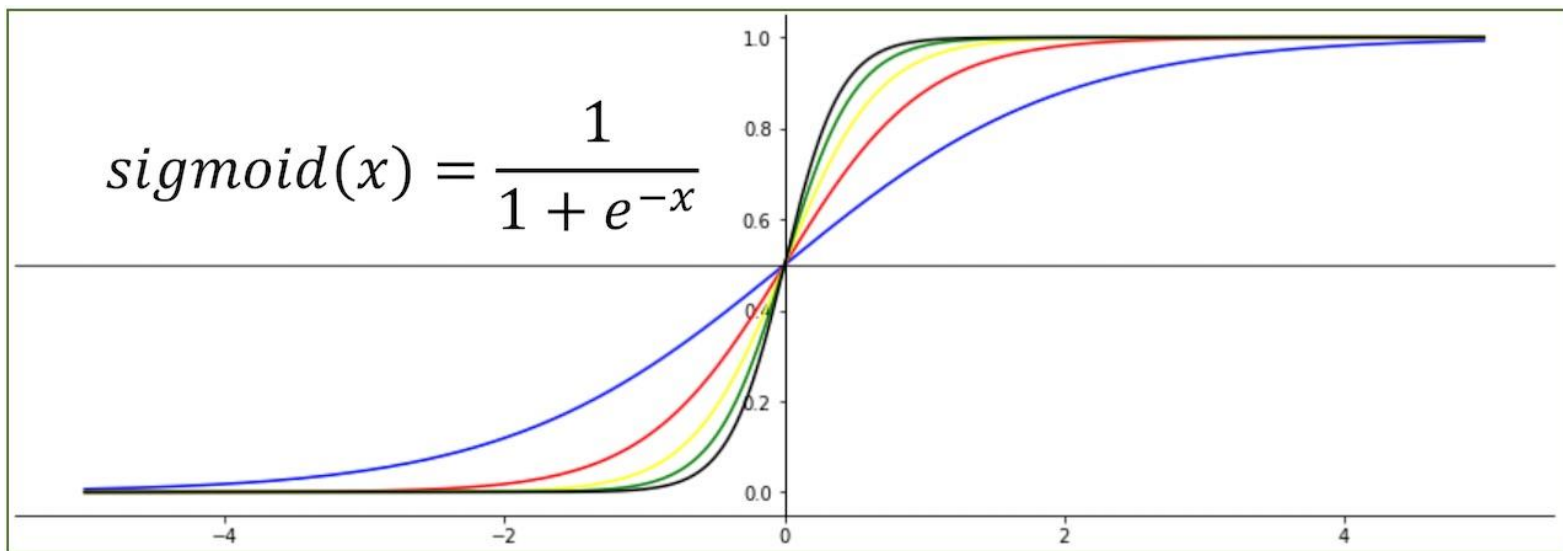
$$\sigma(z) = 1 / (1 + e^{-z})$$

→ Backpropagation을 위해 경사하강법 사용

04 왜 초창기 MLP는 로지스틱이 핵심?

■ 책 364쪽

- 계단함수는 수평선밖에 없으니 Backpropagation 불가
- 그러나 로지스틱 함수는 어디서든지 0이 아닌 Gradient 계산 가능



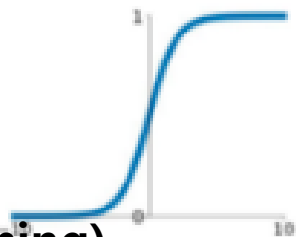
05 인기 많은 활성화 함수 세가지!

Activation Functions

Sigmoid

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

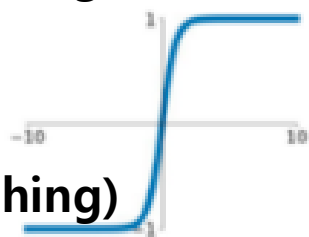
안씀(Gradient Vanishing)



tanh

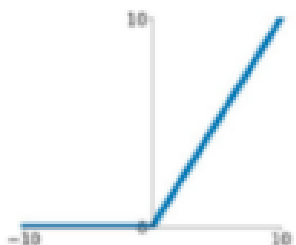
$$\tanh(x)$$

안씀(Gradient Vanishing)



ReLU

$$\max(0, x)$$

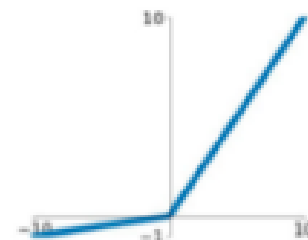


1순위

2순위 Relu에서 가중합 0이하 비활성화가 아쉬울 때

Leaky ReLU

$$\max(0.1x, x)$$

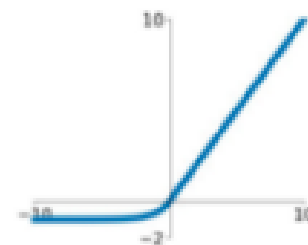


Maxout

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

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Different Activation Functions and their Graphs

06 입력층(10) / 은닉층 (50) / 출력층 (5), Relu

$$X_{m \times n} \times W_{n \times j} + B_{m \times j} = Y_{m \times j}$$

입력층 크기 $\rightarrow m$ (배치 크기) $\times n$ (입력 뉴런)

은닉층 가중치 $W_h \rightarrow n$ (입력 뉴런) $\times j$ (은닉 뉴런)

은닉층 편향 벡터 $b_h \rightarrow j$ (은닉 뉴런) ※ 편향 행렬로 원했다면 $m \times n$

$$X_{m \times n} \times W_{n \times j} + B_{m \times j} = Y_{m \times j}$$

은닉층 크기 $\rightarrow m$ (배치 크기) $\times j$ (은닉 뉴런)

출력층 가중치 $W_o \rightarrow j$ (은닉 뉴런) $\times k$ (출력 뉴런)

출력층 편향 벡터 $b_h \rightarrow k$ (출력 뉴런) ※ 편향 행렬로 원했다면 $m \times k$

$\rightarrow Y = \text{Relu}(\text{Relu}(X * W_h + b_h) W_o + b_o)$

07 스팸메일? MNIST?

■ 스팸메일

출력층 뉴런 개수 : 1개 (1 or 0)

활성화 함수 : Relu나 로지스틱 함수...

■ MNIST

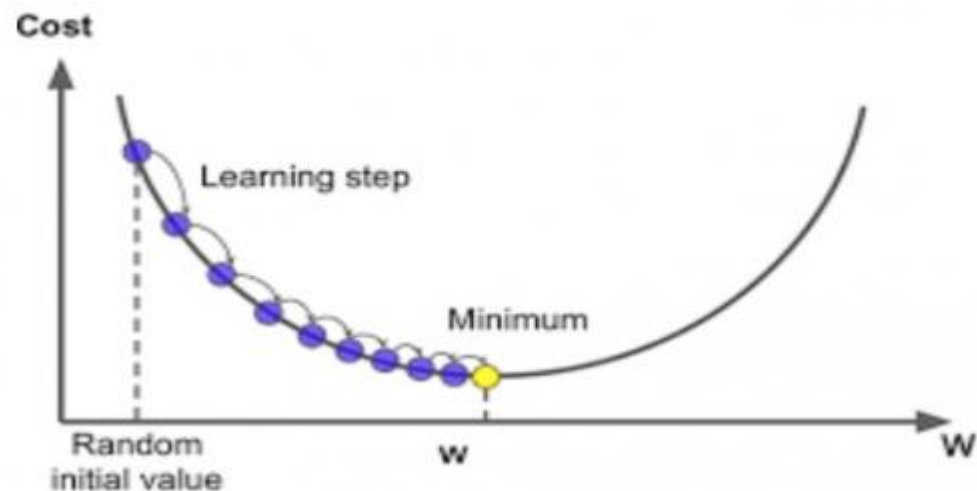
출력층 뉴런 개수 : 10개 (10개의 카테고리)

활성화 함수 : Softmax (다중분류에는 각 예측확률을 계산할 수 있는 Softmax를!)

08 역전파 vs 후진모드 자동미분

■ 역전파 (Chapter 4.2 경사하강법, 362쪽 참고)

그레디언트를 자동으로 계산하는 ‘경사하강법’



■ 역전파 vs 후진모드 자동미분 (362쪽, 부록D ‘자동미분’ 참고)

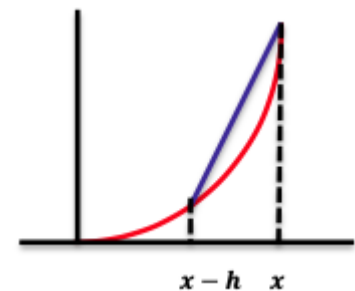
수동 미분 (진짜 인간이 계산)

/

유한 차분 근사

전진모드 자동미분 (함수 덧셈의 도함수는 각 함수를 더한 것과 같으므로)

후진모드 자동미분 (우리가 역전파 때 쓰는 미분 - 연쇄 법칙 이용)



09 MLP에서 조정 가능한 하이퍼파라미터

■ MLP에서 튜닝 가능한 하이퍼파라미터 (10.3 참고)

은닉층 개수, 뉴런 개수, 학습률, 옵티마이저, 배치 크기, 활성화 함수, 반복 횟수

■ 과적합 시?

은닉층/뉴런 개수 조정

조기 종료 사용