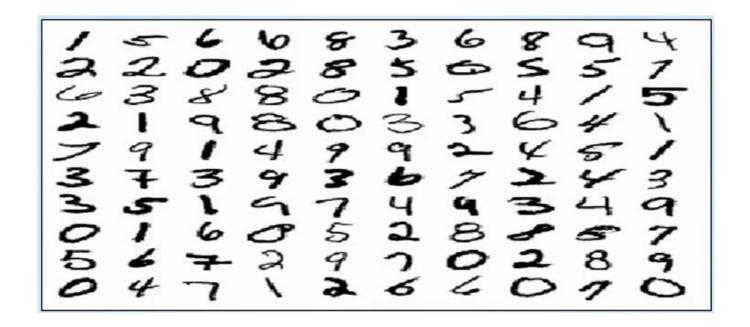
6. MNIST(필기체 숫자)

박경미

MNIST 개요

- MNIST(MNIST(Modified National Institute of Standards and Technology)
 - 손으로 직접 쓴 숫자(필기체 숫자)들로 이루어진 데이터 셋(Data Set)
 - 0~9까지 숫자 이미지로 구성되며 60,000개의 훈련데이터와 10,000개의 테스트 데이터로 이어져 있음



MNIST 다운로드

- > from internet
 - Training Data (csv format) http://www.pjreddie.com/media/files/mnist_train.csv
 - Test Data (csv format) http://www.pjreddie.com/media/files/mnist_test.csv



■ MNIST는 Keras, TensorFlow 등의 딥러닝 프레임웍을 통해서도 가져올 수 있음.

```
(C:\Program Files\Anaconda3) C:\Users\SungHoPark> pip install keras

Requirement already satisfied: keras in c:\Uprogram files\Anaconda3\Uprogram files\uprogram files
```

■ Keras 를 설치 한 후에, 다음과 같이 mnist.load_data() 를 통하여 MNIST 를 가져올 수 있음.

```
from keras.datasets import mnist
(x_train_data, t_train_data), (x_test_data, t_test_data) = mnist.load_data()
```

MNIST 구조

mnist_train.csv

⇒ mnist_train.csv 파일에는 학습에 이용될 수 있도록 정답(label)이 있는 총 60,000 개의 데이터 존재함. 1개의 데이터는 785개의 숫자가 콤마(,)로 분리되어 있는데, 정답을 나타내는 1개의 숫자와 실제 필기체 숫자 이미지를 나타내는 784 개의 숫자로 구성되어 있음

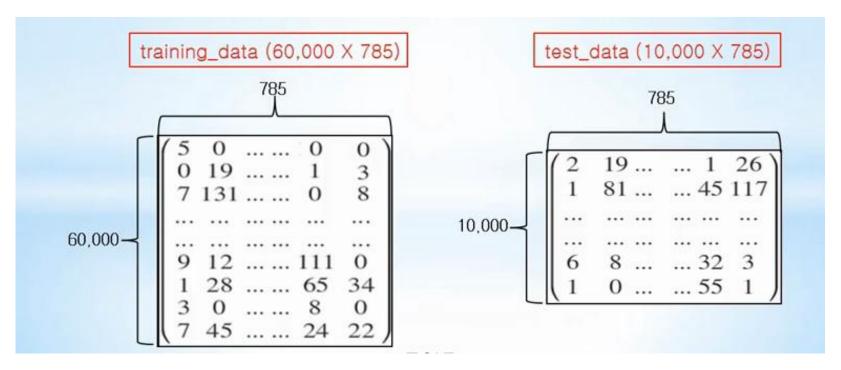
[예] 다음과 같은 데이터라면, 정답은 첫 번째 나오는 숫자 5 이며, 콤마로 분리되어 있는 나머지 784개의 숫자는 정답인 5라는 필기체 숫자의 이미지를 나타내고 있는 숫자들의 조합임

mnist_test.csv

⇒ mnist_test.csv 파일에는 총 10,000개의 데이터가 있으며, 학습을 마친 후에 구현된 딥러닝 아키텍처가 얼마나 잘 동작하는지 테스트 하기 위해 사용됨. 테스트 데이터 또한 정답(label)이 포함된 785 개의 숫자로 되어 있음

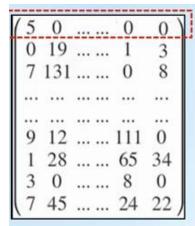
MNIS-구조

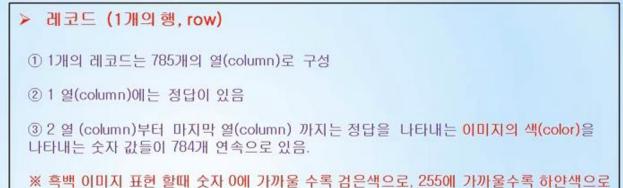
* training_data : 60000개, test_data : 10000개



MNIST-구조

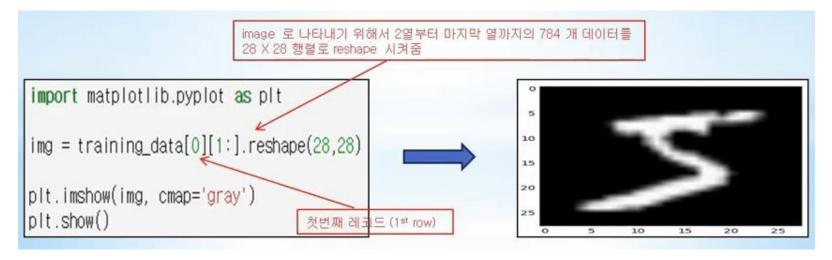
❖ training_data행렬



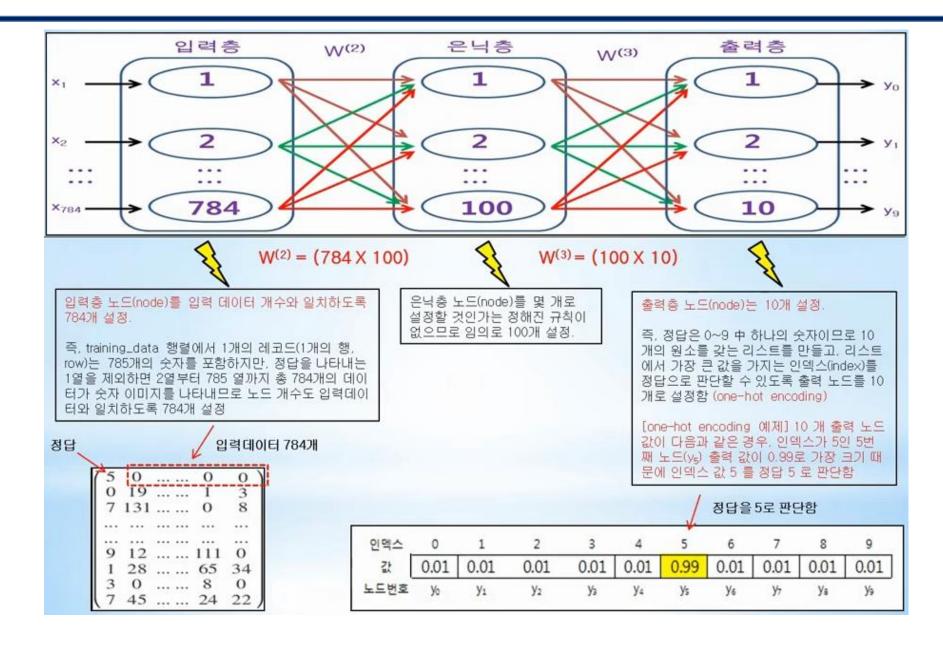


나타내는데 2열부터 마지막 열까지 나열된 숫자가 바로 이미지 색을 나타내는 정보임

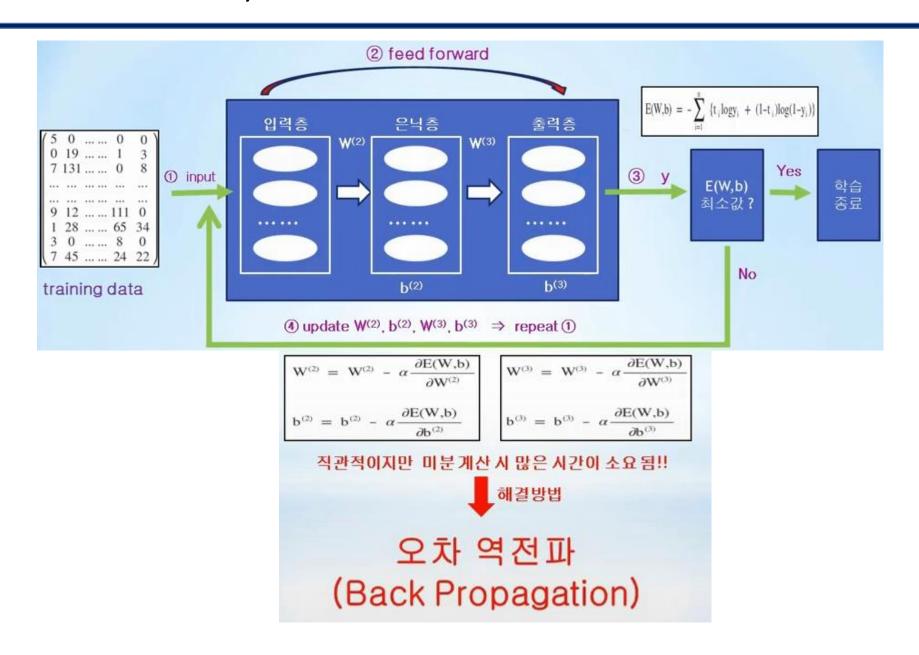
❖ training_data 이미지 표현



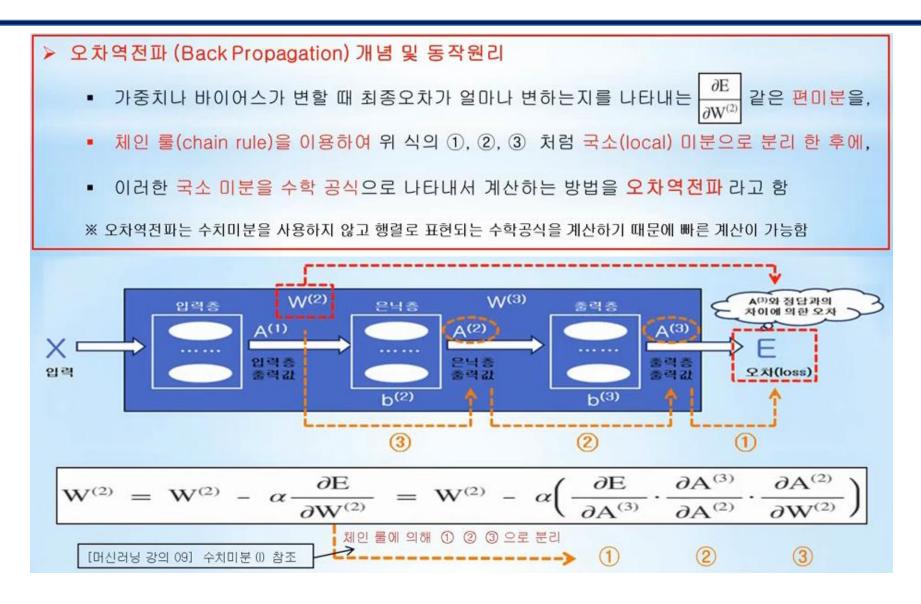
딥러닝 아키텍처(one-hot encoding)



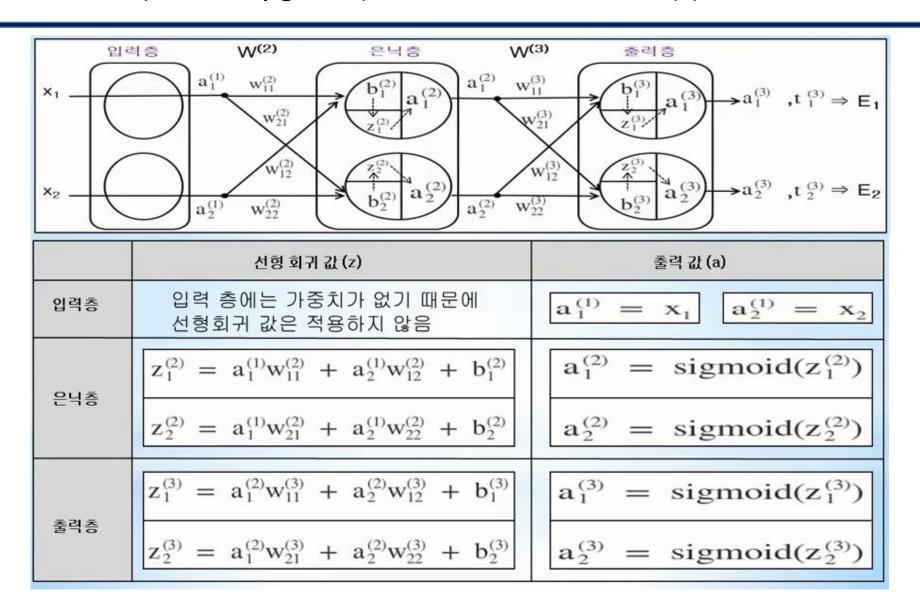
MNIST(필기체 숫자 인식) -수치미분 문제점



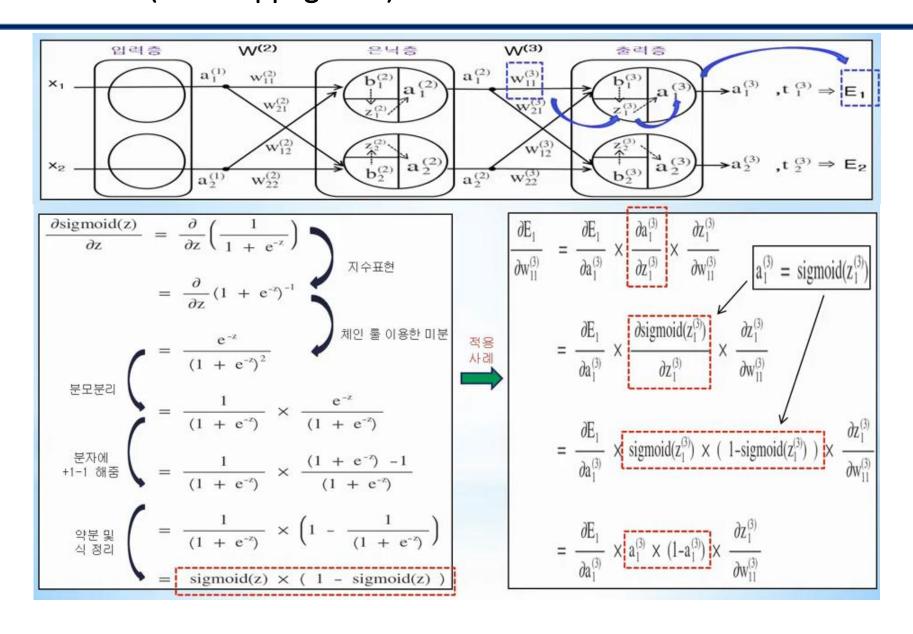
오차역전파(Back Propagation) 개념 및 원리



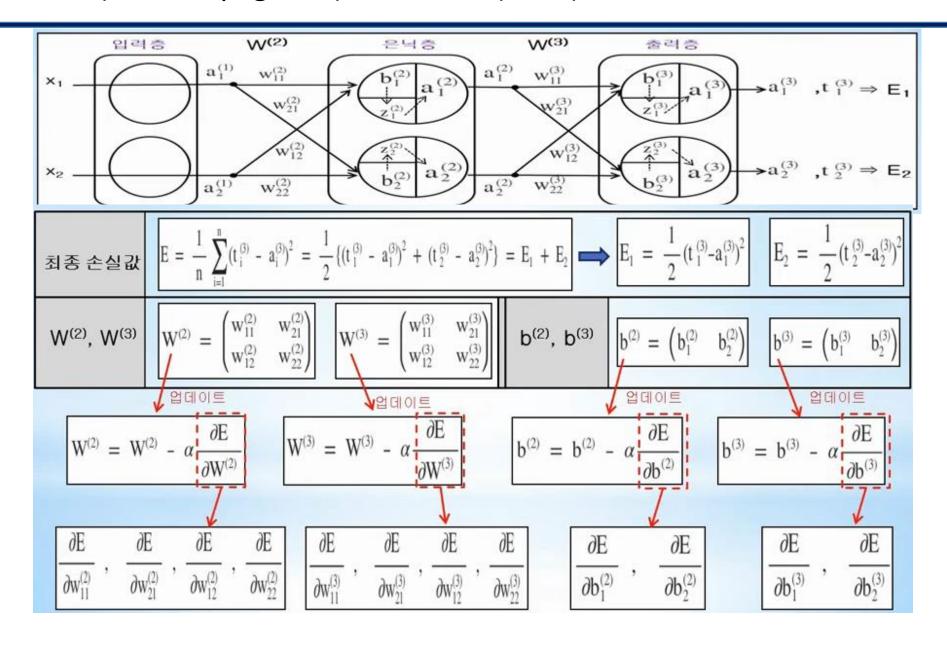
오차역전파(Back Propgation)-각 층의 선형회귀 값(z) / 각층의 출력값 (a)



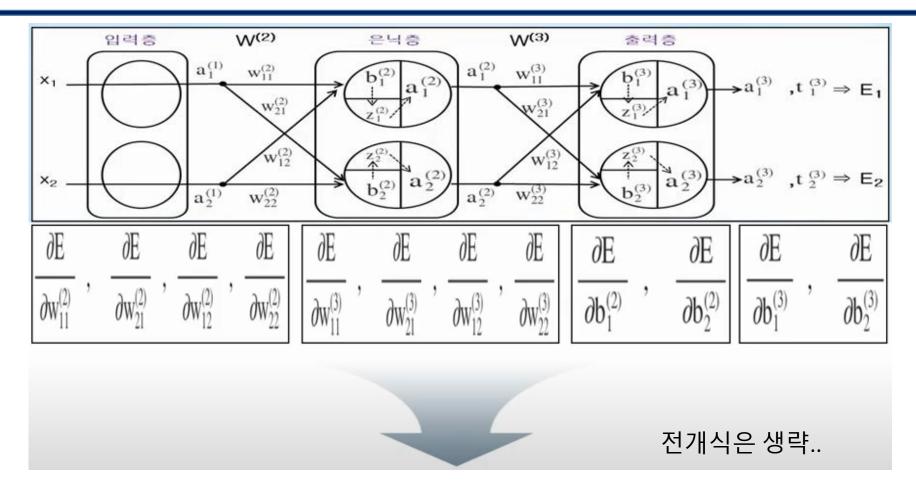
오차역전파(Back Prppagation)-시그모이드 함수 미분



오차역전파 (Back Propagation)-최종 손실(에러)값 E/가중치 W / 바이어스 b



오차역전파(Back Prppagation)-가중치/바이어스 업데이트



가중치(w(2)/ 가중치w(3) 바이어스 b(2) /바이어스 b(2) 미분 값을 수치 미분이 아 닌 행렬기반 수학공식을 이용하는 오차 역전파 구현

오차역전파 일반공식

