

20152410 배형준 머신러닝 과제8

In [1]:

```
1 # library import
2
3 import numpy as np
4 import pandas as pd
5 import matplotlib
6 import matplotlib.pyplot as plt
```

In [2]:

```
1 # set my local working directory
2
3 import os
4
5 directory = 'C:\\Users\\WWgolds\\Desktop\\중앙대학교\\2020-1 4학년 1학기\\머신러닝'
6 os.chdir(directory)
```

In [3]:

```
1 # load dataset
2
3 filename = './과제8/mnist_test.csv'
4 mnist_test = pd.read_csv(filename, header=None)
5 mnist_test.head()
```

Out[3]:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... | 775 | 776 | 777 | 778 | 779 | 780 | 781 | 782 | 783 | 784 |
|---|---|---|---|---|---|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

5 rows × 785 columns

In [4]:

```
1 # make class 'minmaxscaler'
2
3 class minmaxscaler:
4     def __init__(self):
5         self.min_value = 0
6         self.max_value = 0
7
8     def fit(self, X):
9         self.min_value = np.min(X, axis=0)
10        self.max_value = np.where(np.max(X, axis=0) == 0, 1, np.max(X, axis=0))
11
12        return self
13
14    def transform(self, X):
15        scaled = np.zeros(X.shape)
16
17        for j in range(X.shape[1]):
18            scaled[:, j] = (X[:, j] - self.min_value[j]) / (self.max_value[j] - self.min_value[j])
19
20        return scaled
```

데이터 정리하고 정규화 시키기

In [5]:

```
1 # re-arrange data by label
2
3 label = np.array(mnist_test.iloc[:, 0])
4 data = np.array(mnist_test.iloc[:, 1:])
5
6 model_minmax = minmaxscaler()
7 model_minmax.fit(data)
8 data_scaled = model_minmax.transform(data)
```

숫자별로 평균 그림값 만들기

In [6]:

```
1 # make average of pixel matrix
2
3 K = len(np.unique(label))
4 n = 28
5 m = 28
6
7 index_list = []
8 label_list = []
9 data_scaled_list = []
10 data_scaled_reshape_list = [ [] for i in range(K) ]
11 data_scaled_array = []
12 data_scaled_array_mean = []
13
14 for i in range(K):
15     temp = (label == i) # label이 i인 row_number
16     index_list.append(temp) # index_list에 각 row_number 추가
17     label_list.append(label[temp]) # label_list에 해당 라벨들 추가
18     data_scaled_list.append(data_scaled[temp, :]) # data_scaled_list에 label별로 분류
19
20     temp_data_scaled = data_scaled[temp, :]
21     for j in range(np.sum(temp)):
22         data_scaled_reshape_list[i].append(temp_data_scaled[j, :].reshape(n, m))
23
24     data_scaled_array.append(np.array(data_scaled_reshape_list[i]))
25     data_scaled_array_mean.append(data_scaled_array[i].mean(axis=0))
```

랜덤 세타로 label에 따라 sigmoid 값의 평균 구하기

In [7]:

```
1 # define sigmoid function
2
3 def sigmoid(X, theta):
4     z = np.dot(X, theta)
5
6     return 1 / (1 + np.exp(-z))
```

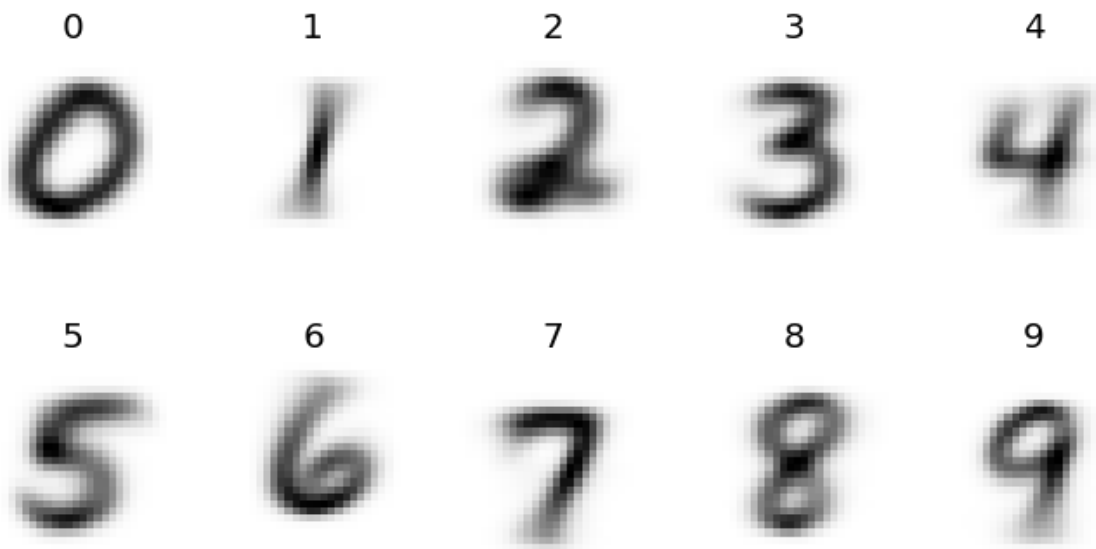
In [8]:

```
1 # calculate sigmoid values and its mean
2
3 sigmoid_list = []
4 sigmoid_mean = []
5
6 student = 20152410
7 np.random.seed(student)
8 theta = np.random.randn(n*m, 1)
9
10 for i in range(K):
11     temp_data_scaled = data_scaled_list[i]
12     temp_sigmoid = sigmoid(temp_data_scaled, theta)
13     sigmoid_list.append(temp_sigmoid)
14     sigmoid_mean.append(np.mean(temp_sigmoid))
15
```

1. Plot the average image

In [9]:

```
1 fig, axes = plt.subplots(2, 5, figsize=(12, 6))
2 axes = axes.ravel()
3
4 for i in range(K):
5     axes[i].imshow(data_scaled_array_mean[i], cmap='Greys', interpolation='None')
6     axes[i].set_title('{}' .format(i), fontsize=20)
7     axes[i].axis('off')
```



2. Present the output of the neural network with random weights

In [10]:

```
1 for i in range(K):
2     print('label이 {}일 때 sigmoid value의 평균은 {}이다.'.format(i, sigmoid_mean[i]), 'Wn')
```

label이 0일 때 sigmoid value의 평균은 0.9769882657694258이다.

label이 1일 때 sigmoid value의 평균은 0.36476027405708833이다.

label이 2일 때 sigmoid value의 평균은 0.7851240600292159이다.

label이 3일 때 sigmoid value의 평균은 0.7548720182540304이다.

label이 4일 때 sigmoid value의 평균은 0.9238865299022120이다.

label이 5일 때 sigmoid value의 평균은 0.8245111836539999이다.

label이 6일 때 sigmoid value의 평균은 0.9195193602958857이다.

label이 7일 때 sigmoid value의 평균은 0.7981214193222912이다.

label이 8일 때 sigmoid value의 평균은 0.8087193050794026이다.

label이 9일 때 sigmoid value의 평균은 0.8660368591934259이다.

