

## < Assignment 03 >

1. Write codes with detailed comments and present useful results at Jupyter Notebook.
2. Export the Jupyter Notebook file as a PDF file.
3. Submit the PDF file to Classroom.

[Visualize average images]

1. Load MNIST training dataset.
2. Compute the average images for each label (digit) based on L2-norm.
3. Visualize the average images.

### 1. Load MNIST training dataset.

In [43]:

```
import matplotlib.pyplot as plt
import numpy as np

file_data = "mnist_train.csv"
handle_file = open(file_data, "r")
data = handle_file.readlines()
handle_file.close()

size_row = 28    # height of the image
size_col = 28    # width of the image

num_image = len(data)

#
# normalize the values of the input data to be [0, 1]
#
def normalize(data):

    data_normalized = (data - min(data)) / (max(data) - min(data))

    return(data_normalized)

#
# example of distance function between two vectors x and y
#
def distance(x, y):

    d = (x - y) ** 2
    s = np.sum(d)
    # r = np.sqrt(s)

    return(s)
```

### 2. Compute the average images for each label (digit) based on L2-norm.

In [44]:

```
#
#norm을 구하기 위한 선언
#
from numpy import linalg as LA
#
#(size_row * size_col, num_image) 크기의 배열 생성
#
list_image0 = np.zeros((size_row * size_col, num_image), dtype=float)
#
#(size_row * size_col, 10) 크기의 배열 생성
#
list_image_avg = np.zeros((size_row * size_col, 10), dtype=float)
```

```

#
#(size_row * size_col, 10) 크기의 배열 생성
#
NA = np.zeros((size_row * size_col,10), dtype=float)
#
#인덱스 변수 생성
#
idx=0
#
#레이블을 기준으로 이미지 분류
#
for i in range(10):

    for j in range(num_image):

        if list_label[j]==i:
            list_image0[:,idx]=list_image[:,j]
            idx=idx+1

    num=idx
#
#인덱스 변수 0으로 초기화
#
    idx=0
#
#(size_row * size_col, num) 크기의 배열 생성
#
    list_image_temp=np.zeros((size_row * size_col,num))

    for k in range(num):
#
#레이블이 k일때의 이미지만 list_image_temp 배열에 저장
#
        list_image_temp[:,k]=list_image0[:,k]
#
#각 이미지들의 원소들끼리의 norm을 원소로 하는 벡터를 구한다.
#
        list_image_avg[:,i]=LA.norm(list_image_temp,axis=1)
#
#list_image_avg를 정규화한다.
#
        NA[:,i]=normalize(list_image_avg[:,i])

```

### 3. Visualize the average images.

In [45]:

```

#
#시각화
#
for i in range(10):

    label      = i
    im_vector  = NA[:,i]
    im_matrix   = im_vector.reshape((size_row, size_col))

    plt.subplot(1, 10, i+1)
    plt.title(label)
    plt.imshow(im_matrix, cmap='Greys', interpolation='None')

    frame      = plt.gca()
    frame.axes.get_xaxis().set_visible(False)
    frame.axes.get_yaxis().set_visible(False)

plt.show()

```



