Assignment03

March 24, 2019

20163228 Yuseon Nam

```
In [1]: import matplotlib.pyplot as plt
        import numpy as np

        #
        # 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)

        #
        # Sum of square of image matrix (L2-Norm)
        # new_sum = x1^2 + x2^2 + ... + xn^2 + xnew^2
        #
        def sum_img_matrix(new_img_matrix, sum_2):

            new_sum = sum_2 + new_img_matrix ** 2

            return (new_sum)

        #
        # Read Train File
        #
        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)
        count        = 0       # count for the number of images
```

```python
#
# Make a matrix each column of which represents an images in a vector form
# and save label in another vector
#

# image matrix for all images (28, 28, 60000)
whole_img_matrix = np.empty((size_row, size_col, num_image), dtype=float)
list_label       = np.empty(num_image, dtype=int)    # label for each image

for line in data:

    line_data         = line.split(',')     # len(line_data) = 784
    label             = line_data[0]
    list_label[count] = label

    im_vector = np.asfarray(line_data[1:])
    im_vector = normalize(im_vector)
    im_matrix = im_vector.reshape((size_row, size_col))  # one image (28*28)
    whole_img_matrix[:, :, count] = im_matrix

    count += 1




#
# Get sum of image matrix and count for each label
#

# average image matrix of each label(0-9) in array form (label, 28, 28)
avg_img_matrix = np.empty((10, size_row, size_col), dtype=float)
num_count      = np.empty(10, dtype=int)          # count for each label(0-9)

for i in range(num_image):
    label = list_label[i]
    num_count[label] += 1
    avg_img_matrix[label, :, :] = sum_img_matrix(whole_img_matrix[:, :, i],
                                                 avg_img_matrix[label, :, :])




#
# Get average image matrix of each label by calculating sqrt(sum / count)
#
for i in range(10) :
    avg_img_matrix[i, :, :] = (avg_img_matrix[i, :, :] / num_count[i]) ** 0.5
```

2

```
In [2]: #
        # Plot average image of each number (0-9)
        #
        f2 = plt.figure(figsize=(20,2))

        for i in range(10):
            plt.subplot(1, 10, i+1)
            plt.title(i)
            plt.imshow(avg_img_matrix[i, : , :], cmap='Greys', interpolation='None')

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

        plt.show()
```