Assignment_03

March 28, 2019

20145271 Geum Kang Hyeon

0.1 Norm

Norm is a method of measuring the length or size of the vector.

$$L_p = (\sum_{i}^{n} |x_i|^p)^{\frac{1}{p}}$$

0.2 L2 Norm

L2 Norm means when p is 2. Also it called the Euclidean norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^t x}$$

The data file contains a total of 60,000 pixels of information, consisting of numbers from 0 to 9.

First, the data entered in a row was converted to a 28x28 pixel array.
Second, For each number, the L2 Norm value was calculated by squaring each pixel data.
Third, I calculated the root value of the result.

In accordance with the above sequence, I calculated the L2 Norm by accumulating this data for each number.
Below code is my Computation Result.

```
In [24]: 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)
count           = 0     # count for the number of images

#
# 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)

#
# make a matrix each column of which represents an images in a vector form
#
list_image  = np.empty((size_row * size_col, num_image), dtype=float)
list_label  = np.empty(num_image, dtype=int)
list_labeled = np.zeros((size_row * size_col, 10), dtype=float)
label_count = np.zeros(10, dtype=int)

for line in data:

    line_data   = line.split(',')
    label       = line_data[0]
    im_vector   = np.asfarray(line_data[1:])
    im_vector   = normalize(im_vector)

    list_labeled[:, int(label)] += im_vector[:]**2
    label_count[int(label)] += 1

    list_label[count]       = label
    list_image[:, count]    = im_vector
```

```
    count += 1

for i in range(10):
    list_labeled[:, i] = np.sqrt(list_labeled[:, i])
#
# plot first 100 images out of 10,000 with their labels
#
f1 = plt.figure(1)

for i in range(10):

    label       = i
    im_vector   = list_labeled[:, 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()
```