**ML HW5**

**October 15, 2021**

**21500468 GyuSeok,Lee**

1. **Introduction**

In this homework, I have to calculate mean, variance, eigen values and eigen vectors about Mnist dataset (50000 x 784). By doing this, I can understand the dominant features of dataset and practice how to calculate descriptive statistics.

1. **Experiment**

In this homework, I used Mnist dataset (50000, 784). According to Mnist Database, the dataset consists of handwritten digits and the digits have been size-normalized and centered in a fixed-size image. The total mean of data is about 0.13044 and standard deviation is about 0.3072898.

|  |
| --- |
| 텍스트이(가) 표시된 사진  자동 생성된 설명 |

From now on, I will explain how the experiment was conducted for each question.

**Q1: Calculate mean, variance, eigen vectors and eigenvalues**

First, I calculate mean, variance, eigen vectors and eigenvalues by using NumPy like below.

|  |
| --- |
| 텍스트이(가) 표시된 사진  자동 생성된 설명 |

Then, I print out these results. You can check these results in attached python codes.

|  |
| --- |
|  |

**Q2: Print the mean and the first 10 eigenvector images**

To solve this problem, 784-vector size was changed to 28×28 then I can print out each image.

|  |
| --- |
| **Mean Image** |
|  |

In the same way, the 10 eigen vectors could be outputs as each image.

|  |
| --- |
| **Code** |
|  |

|  |
| --- |
| **Eigen Vector Images** |
|  |

**Q3: Plot the first 100 eigenvalues.**

This is a picture of up to 100 Eigen values obtained above.

|  |
| --- |
| **100 Eigen Values** |
|  |

1. **Results**

The results obtained through the task are as follows. By obtaining 784 means and variance, statistics for each pixel could be obtained. In addition, by obtaining the eigen vectors and the eigen values, the main features could be extracted from all images. Finally, it was confirmed that the eigen value gradually decreased.

1. **Conclusions**

I think it is important to get eigenvectors for extracting dominant features in each image because these features can be maintained unchanged. These features can greatly help reduce dimensions through methodologies such as PCA. Plus, through the experiment, it was found that the eigen value can be obtained through the covariance matrix. Such statistical knowledge seems indispensable for understanding machine learning algorithms. Above all, it will be essential to use basic statistics on what characteristics the data given to me has before modeling. I hope this realization will be well applied to performing a given task in the future.