**Homework Assignment #6**

**21600004**

**Kang Seok-Un**

**Introduction**

This homework uses the famous public dataset, the MNIST dataset. At this time, entire data are used. The algorithm used in this homework is principal component analysis(PCA) and linear discriminant analysis(LDA). When visualize the dataset, dimension will be two.

**Experiments**

1. **PCA**

First, calculate the mean of the entire train dataset. Then, calculate the Covariant matrix ‘cov’. After that, by using eigendecomposition, find the eigenvalues and eigenvectors. At this time, the important work is that by sorting eigenvalues in descending order, sort eigenvectors corresponding to eigenvalues in descending order.

Multiply the front two vectors of the sorted eigenvector to project the train data into the two-dimensional eigenspace by the train data (Fig 1.). After all these processes, plotting the ‘scatter()’ is the same as Figure 2.

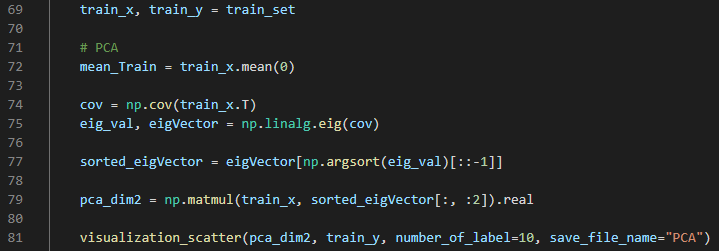


Figure 1. The implementation of PCA

1. **LDA**

To use LDA, should follow the processes below.

* 1. Calculate Mean of Global and each classes

For Calculate and , should calculate the mean of entire dataset and each classes as figure 2.

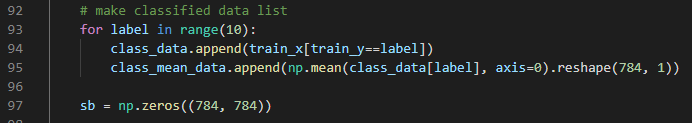


Figure 2. Calculating the mean of each classes

* 1. Calculate

At section 2.1., we know the mean of entire dataset and each classes. First, calculate the differences between mean of each classes and global mean (figure 3. Line 102). And then multiply number of dataset in each class, differences and that’s transposed matrix. If add all the results of the previous multiplication, will get .

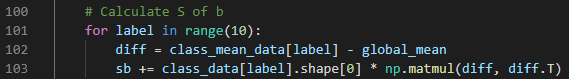
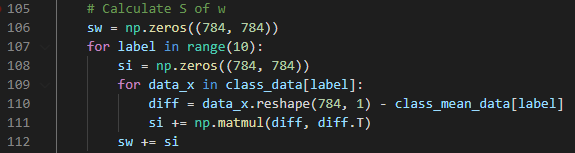


Figure 3. Calculating

* 1. Calculate

This section has a similar process to section 2.2. However, this section finds the mean of the class to which the data belongs and the difference between that data.



**Figure 4. Calculate**

* 1. Projection onto 2-dim eigenspace

If multiply the inverse matrix of by and then perform eigenvalue composition, the eigenvalue and eigenvector want can be obtained. Next, the sorted eigenvector can be projected into 2-dim eigenspace by multiplying the dataset as did in PCA.

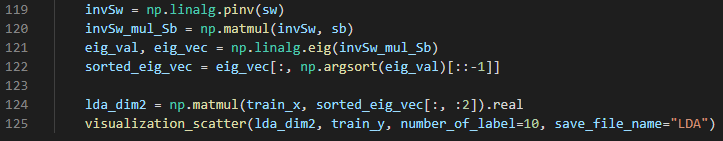


Figure 5. Projection into 2-dim eigenspace

**Results**

As shown in the Experience section, the following results can be obtained by implementing and operating the code.

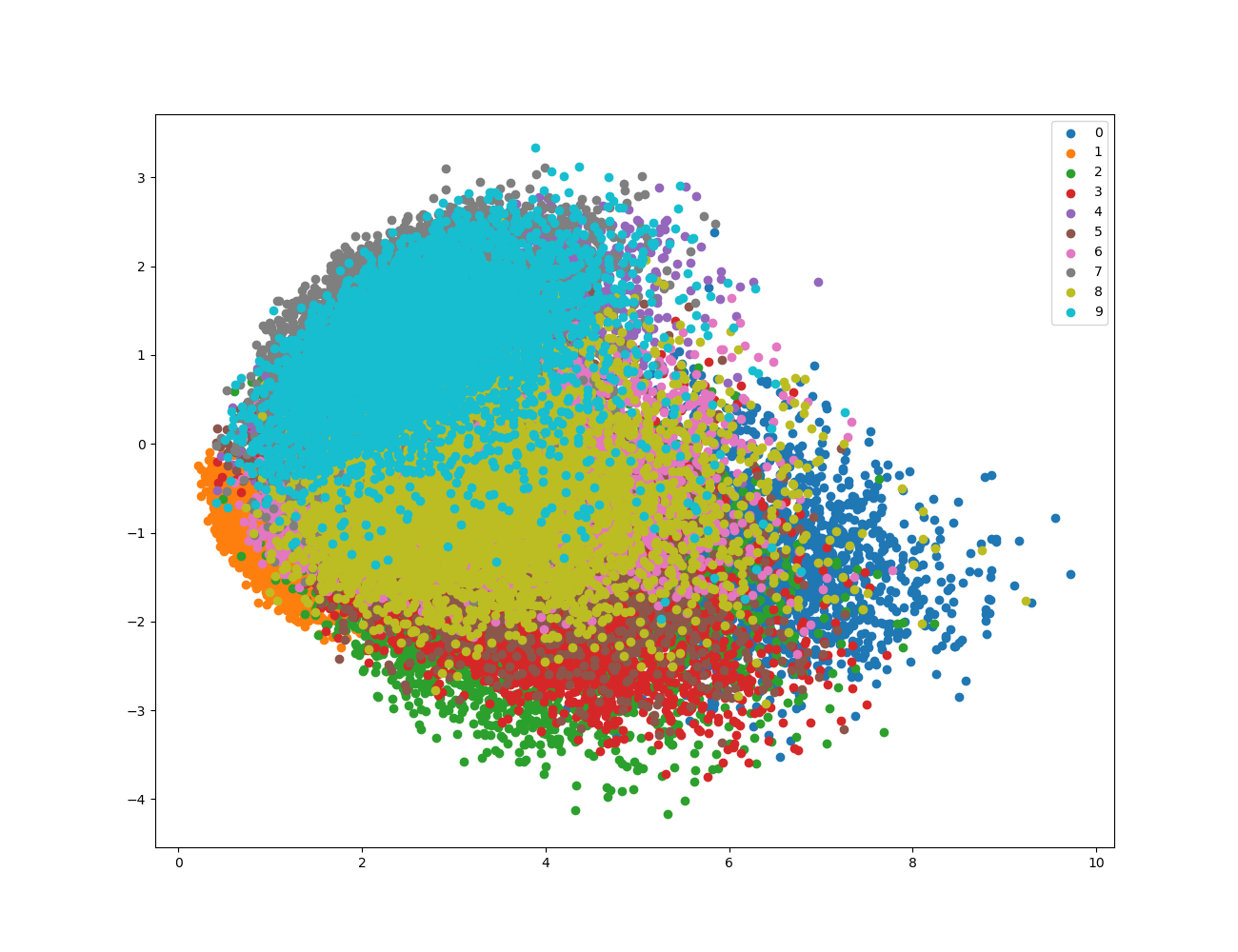


Figure 6. The result of PCA into 2-dim eigenspace

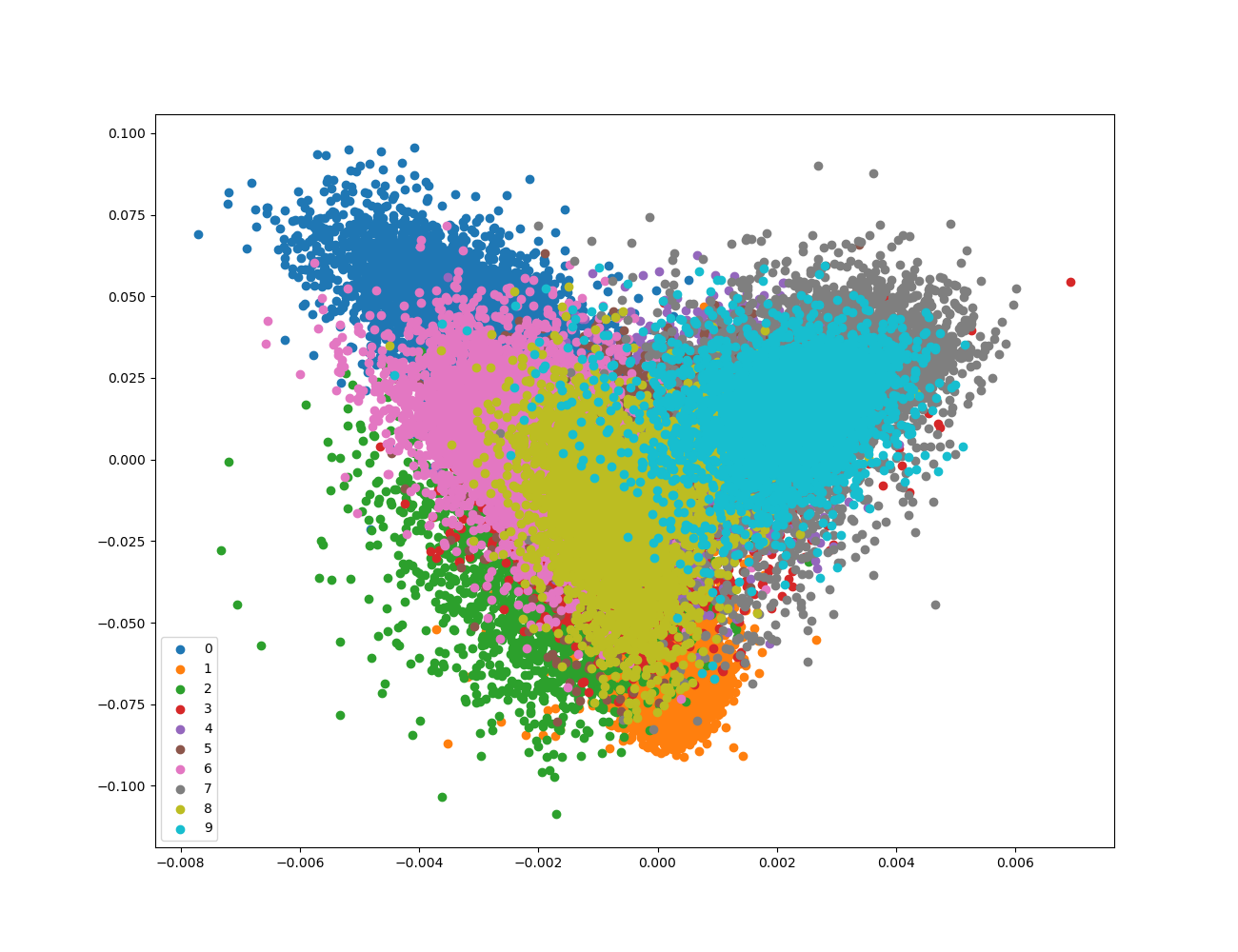


Figure 7. The result of LDA into 2-dim eigenspace

**Conclusion**

The most difficult thing in carrying out this task was to check and use the dimension of the data. In the case of coding to solve problems by using a specific algorithm, At most of them used a two-dimensional array. However, in this task, which uses a NumPy array and results vary depending on each axis, there were many difficulties in understanding the information on each axis.

LDA says it is easier in aspects of classification than PCA due to using labels. When comparing PCA and LDA, it can be seen that in class 0, LDA is a little denser. However, when comparing Figures 6 and 7, it is still questionable whether an answer can be obtained to the question, "Is it really easy to do classification?". In addition, the results were worse than those using deep learning in the textbook pdf, so I look forward to implementing more advanced technology in the next task.