**2)**

first imported needed libraries, then read data from csv files similar to the lab sessions.

Graphical user interface, text, application

Description automatically generated

**3)**

In this part, I divide data in two part. First one is train data which from 0 to 30000th data. remaining 5000 data is the test data for us. Test data will be used for comparing with results.

Text

Description automatically generated

**4)**

In calculating parameters, first I got K(number of classes),N(number of data for train) values. Then I calculated sample means from given data with corresponding formula below(similar to the hw1).

Text

Description automatically generated with medium confidence

Graphical user interface, text, email

Description automatically generated

I calculated sample deviations by using np.std. Calculated sample deviations is a matrix with shape (5,784)

Graphical user interface, text, application

Description automatically generated

Then calculated class priors with given formula, similar to the hw1.

Text

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

**5)**

Using estimated parameters, I used naive bayes classifier as parametric classification rule. Used this g score formula to classify data.

Text, letter

Description automatically generated

Created g score function for the train and test data with given formula above.

Graphical user interface, text, application

Description automatically generated

After finding g scores of the data, to find their class I used this function. I could use np.argmax(g\_scores, axis=1) where axis=1 means for every row.

Graphical user interface, text, application

Description automatically generated

Finally, I found confusion matrices for data points in train data and test data. confusion matrix for train data is similar to what we expected. However, in test data, we are not good at predictions.

A picture containing text

Description automatically generated