Oya İlayda Yalçın

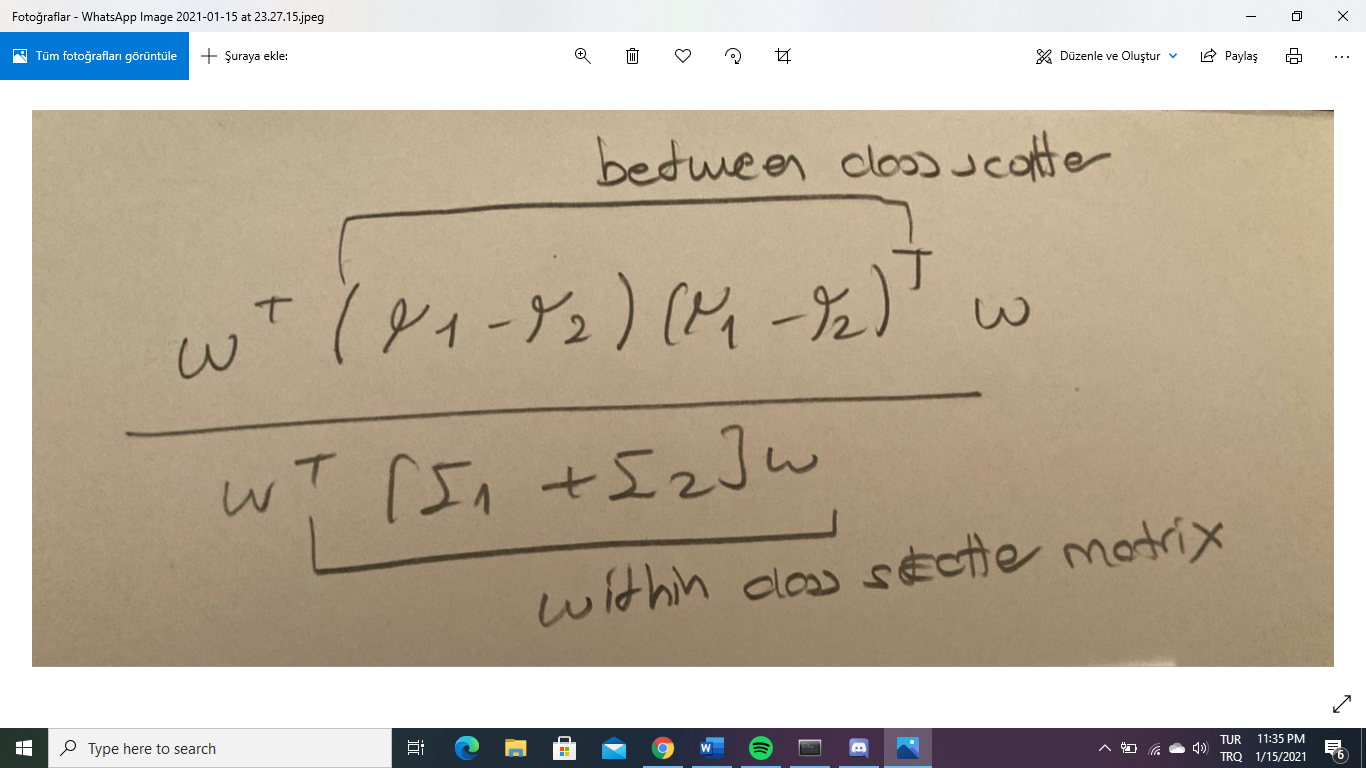
60692

ENGR 421 HW#7

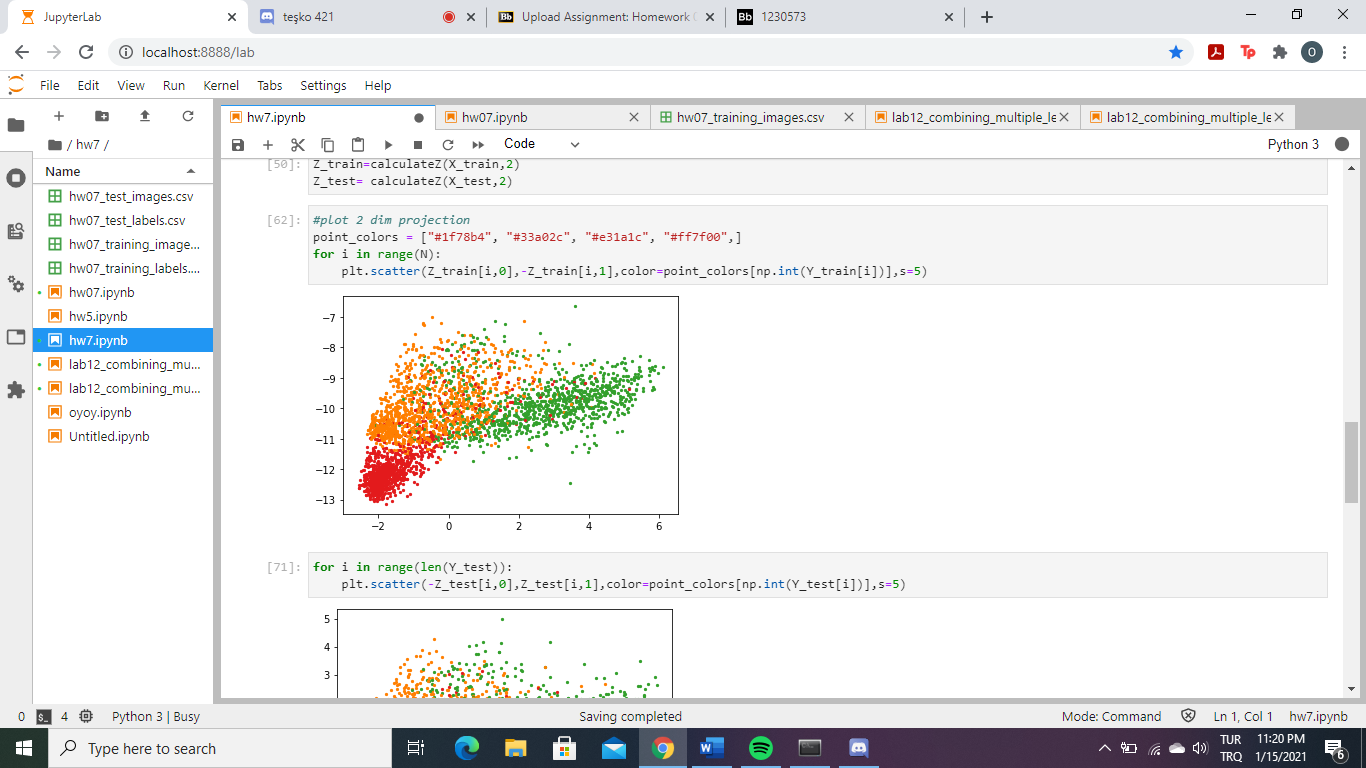
In this project we were asked to implement a linear disciriminant analysis on a data set about images with 28x28 pixels.

For this code first the class means are calculated and these total means and class means are used in order to find the between scatter and the within scatter matrixes. We wanted difference between means to be as large as possible and the sum off covarianes to be small as possible.

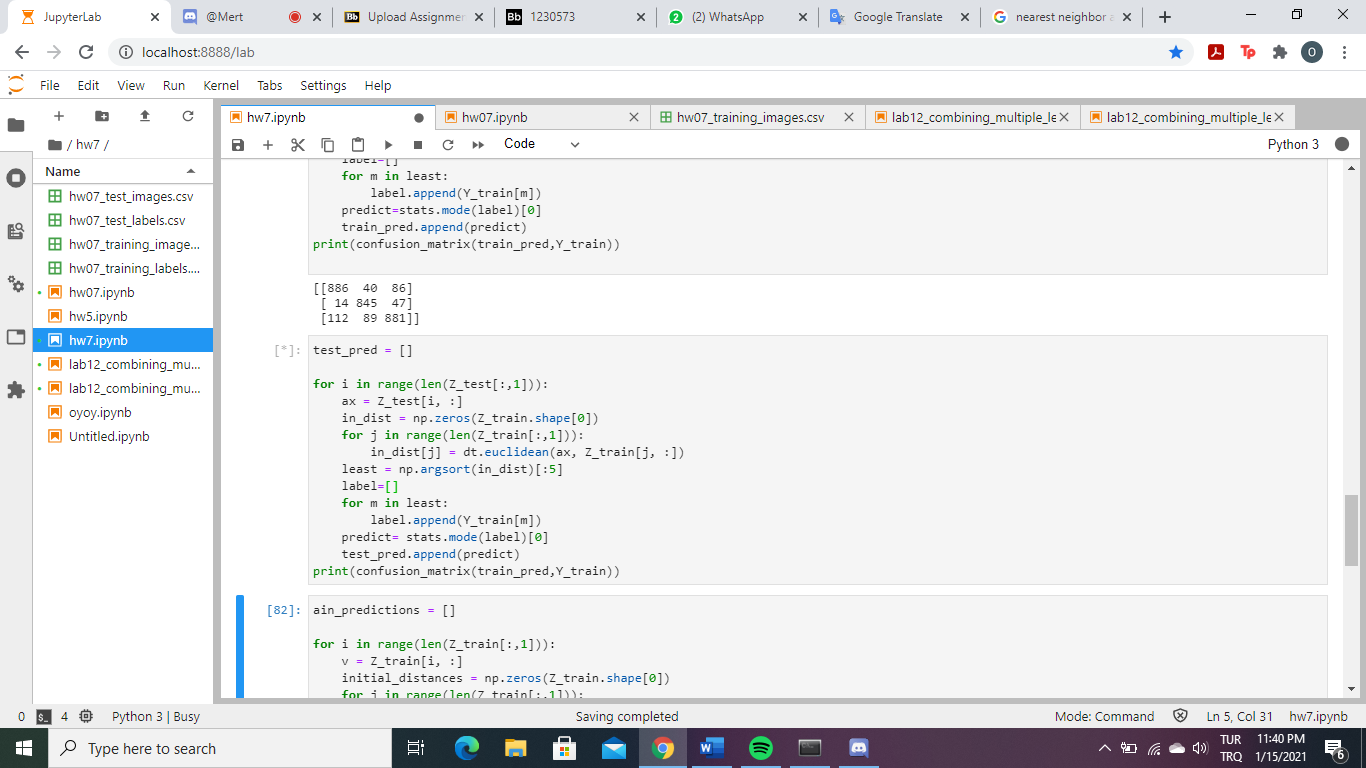
For this I coded two matrixes called within scatter and between scatter. Within class scatter is calculated by summing the covariences of classes. So, in the process a covariance calculation is made for all the classes. Between scatter matrix consisted of scale covariance formula. So for all the data points, class means are taken out. To be clear used formulas are given below:



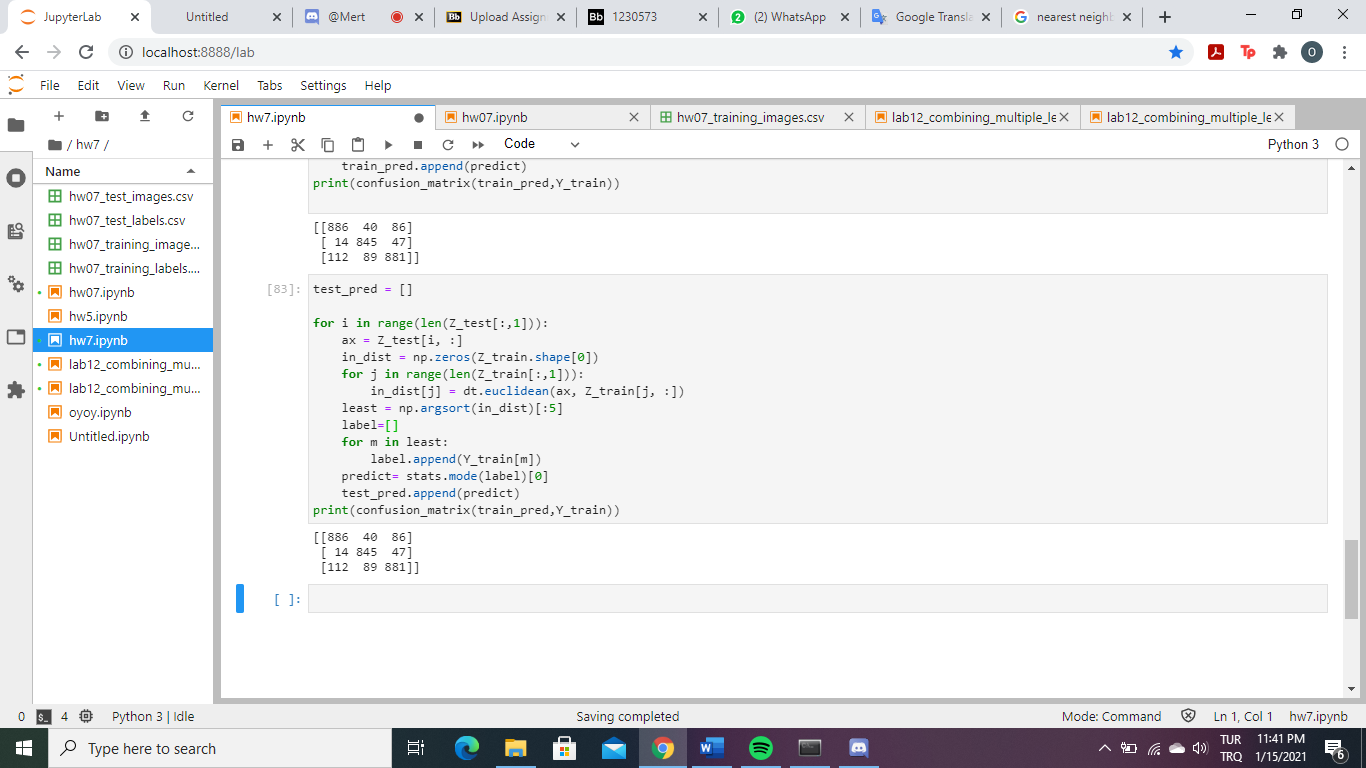
In the end, taking the eigen decomposition of Sw-1 \* Sb, I took the first two eigenvectors with largest eigen values to determine w. Matrix multiplying this w with data points I was able to project data to another Z plane which has two dimensions. The graphs I found are given:



Later on, for two-dimensional representation, I learned a five-nearest neighbor classifier using the projections of training data points and calculated the confusion matrices on the projections of training and test data points, respectively. To do this, I used Euclidian distances. Then from the nearest 5 neighbor I classified each data point. In the end, I created a confusion matrix for these and result was pretty good.



For training data



For test data