**Machine Learning HW3**

By: Roy Zohar 209896174

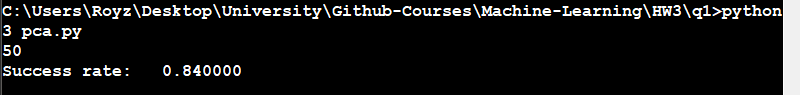
Roy Mezan 319042800

**Note**:

We implemented the entire project in python since we are more comfortable in it than in MATLAB. We implemented our own methods for reading the data, and also translated files into csv format (supplied to the zip folder)

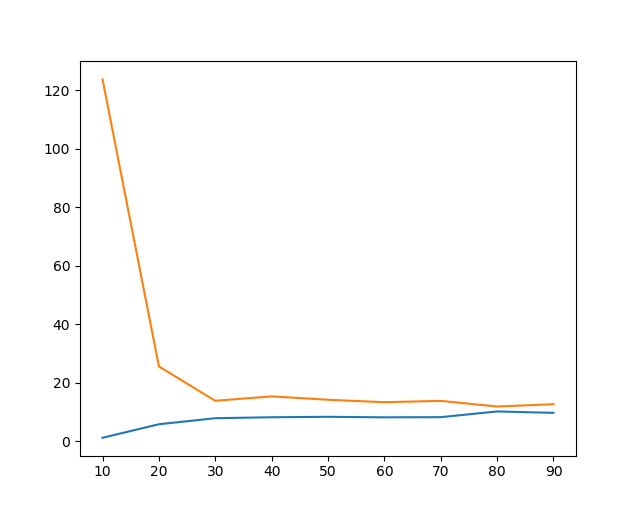
**Question 1:**

In this question, we implemented the PCA and MDA as described. We used numpy.linalg.eig() in order to find the eigenvectors which represent the data most accurately (i.e. largest corresponding eigenvalues), and projected the data onto them. For the MDA, the same method (numpy.linalg.eig()), also finds the solution to the generalized eigenvalue problem. We also used a KD-Tree in order to find the closest sample. We reached a success rate of 84% classification using this method:



**Question 2:**

We implemented our own linear regression to predict a city-cycle fuel consumption in Python. After we exported data to csv files we read them in Python and used a pre-prepared function from the sklearn library. We implemented our own squared error and prediction function.

The plot below, shows the error (distance) of the model from the true value, as a function of the training set size. The blue line represents our model's error from the training data, and the orange line is our models error from the testing data.

Note that for a sufficient training set, we get a pretty accurate model, with quite a small error.

**Question 3:**

In this question, we again copied the data into csv table so that our Python script could read them. We implemented the SVM using the sklearn library, and functions we wrote by ourselves. We trained the data chunks from our K-Fold method, and trained them in "One Against All" (train each class against all other classes combined), We repeat the process for all values of C and G. Below are our results:

