Name: Andre Freeman

Student Id: 410521320

Introduction to Machine Learning

Assignment2:

Handwritten Character Recognition

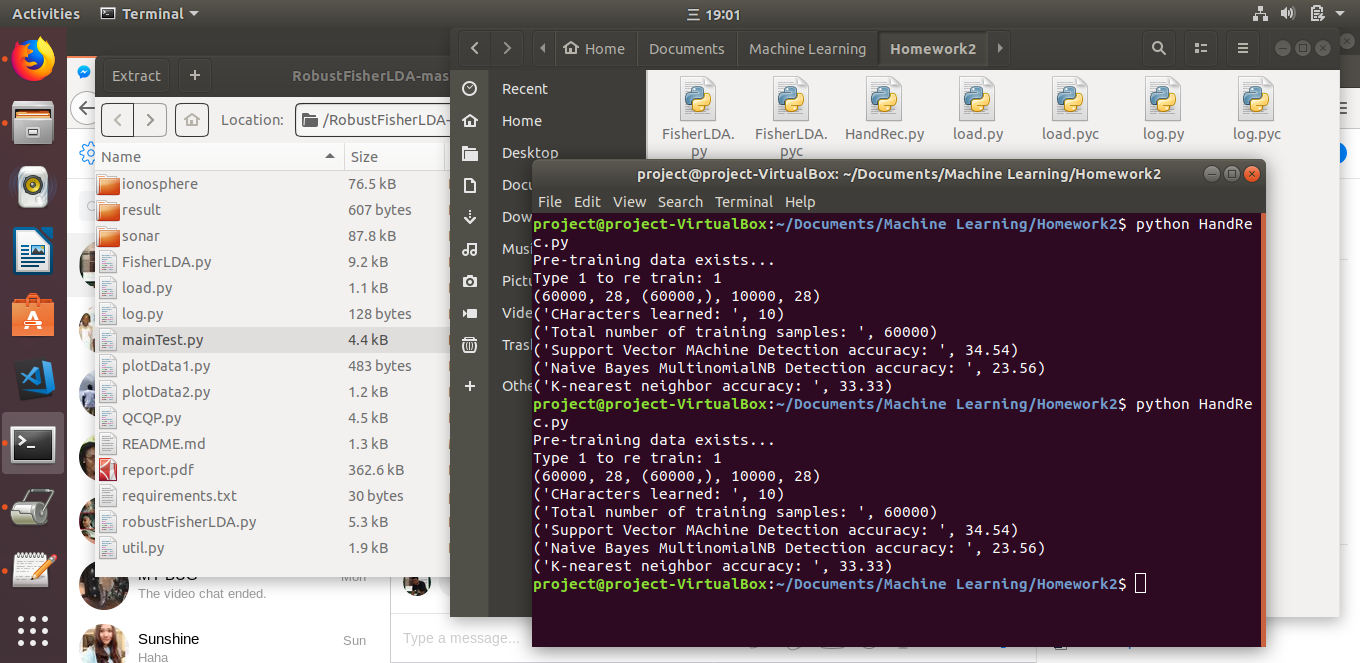
Data Preparation

The data utilized for this project was obtain through sklearn datasets fetch\_mldata function. The data contains 70000 flatter images of dimension 768. Each image is basically a flatten representation of a 28x28 image photo. We first attempt to normalize by dividing each pixel by 255. This then caused all the data to be in the range [0,1]. We then crop each photo but also resize it back to a 28x28 dimension. After normalization and cropping, due to the fact that 768 is a dimension that is too vast, we apply dimensional reduction. To achieve this we utilize the Principal Component Analysis (PCA) to reduce 768 dimension to a 28 dimension problem. For trial and testing purposes we decide to utilize 60000 samples for training data and 10000 for testing data

Data Training

As suggested by the Homework assignment we are ask to utilize different learning algorithm to computer and contrast their performance. In this project we utilized the support vector machine (SVM), Naive-Bayes multinomialNB and the Kth Nearest Neighbor (KNN). Each training algorithm was train tested and their accuracy percentage presented.

Results



The results above indicates PCA alone is not sufficient. Perhaps an implementation of the Fisher LDA could have contributed to better accuracy.

Challenges and Learned Experience

Understand how Fisher LDA works and how to implement it. As the can see the Fisher LDA is missing from this project due to my inability to fully understand its functionality with the given mandate.