Homework 3
Statistical Learning, Spring Term 2019
STAT760

The website for the Statistical Learning book has a repository of data:

https://web.stanford.edu/~hastie/ElemStatLearn/

Under the "Zip code category" we can find a training data set for handwritten digits.

Problem 1

a) Compute the mean and the covariance matrix for the features of each digit in the training set. Compute the determinant of each covariance matrix. Report the values of the means and these determinants.
(10 points)

Program using Matlab, R, Python, or any other language of your choice. Do not use a library function for obtaining the mean and covariance matrix.

Problem 2 (15 points)

This problem is about the Fisher discriminator. Generate randomly two 2-dimensional datasets with Gaussian distributions of known mean and known covariance matrix (different from each other using, for example, 25 data points each). Plot your datasets in the plane. Compute the optimal projection direction using the Fisher criterion. Plot that line and compare with the separating plane obtained using linear regression as a classifier. Provide the plot.

Paper to read: I uploaded a paper about face recognition to the cloud (the link you have). Read that paper for discussion, maybe next week. It is a nice example of what people in vision science call visual features.

The link to the repository is
https://drive.google.com/drive/folders/19kdVsrUPBT6NWIAba2WjiAzpL0R0XjG9?usp=sharing