Homework 6
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 "Data" button we can find several datasets. The "Zip code" dataset has a database of handwritten digits. The button "Info" describes the format. Each digit is represented by 256 gray values. There is a training set and a test set that you can use for Problem 1.

Problem 1

a) Find the mean of each of the ten classes and also the first 3 principal components for each class. You will have to add some random noise to the covariance matrices to make them have determinant different than zero. You can use built-in functions in your programming language for finding eigenvectors and eigenvalues.

Show the mean of each class as a little picture 16x16 and also each of the tree principal components as a 16x16 picture.

(10 points)

Problem 2 (10 points)

This problem refers to Chapter 4 of Neural Networks, Springer-Verlag (Perceptron Learning).

Create a linearly separable training set with two classes P and N (in two dimensions, ten points each class). Find the separating line using the perceptron learning algorithm. Stop the algorithm when the classes have been separated. Plot the result.

Train the same set using gradient descent on the error function for a sigmoidal activation function. Plot the separating boundary for threshold 0.5.