AMSC/MATH 420, Project 2A, SPRING 2016

Oral Presentation Due: TBA
Written Presentation Due: TBA

1) Download the dataset of handwritten digits collected by USPS and divide them into two sets: the training set consisting of 100 examples of each of the digits 0 through 9; and the testing set consisting of 1000 examples of each of the digits 0 through 9. The goals of this project are as follows:

- Develop and test methods for classification of the handwritten data (to be discussed further in class), which are optimized on the training set and then applied to the testing set to assess the performance of the developed methodology.
- Analyze the impact of the parameter optimization of the classifier on the training set, on the performance of the classification scheme as measured on the testing data.

Use the nearest neighbors classification scheme (to be introduced in class), or any other suitable classifier that you may already know, to verify the success rate of your compression scheme. Optimize its parameters and the selected metric to maximize the global success rate. Report your success percentages for each compressed digit separately, as well as globally. Analyze the performance differences between individual digits.