

AMSC/MATH 420, Project 1A, 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 N examples of each of the digits 0 through 9, and the testing set consisting of $1100 - N$ 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 size of the training set on the performance of the classification scheme.

Use the nearest neighbors classification scheme in the standard Euclidean metric with $k = 20$ (to be introduced in class), or any other suitable classifier that you may already know, to verify the success rate of your classifications. Analyze the role of the split of the training vs testing data in your classification results.