Exploration of Digit Recognition Algorithms

Harnoor Singh
Department of Computer Science & Engineering
University of Washington
hsingh@cs.washington.edu

Brian Walker
Department of Computer Science & Engineering
University of Washington
bdwalker@cs.washington.edu

1 Objective

As described in our project proposal, we are using the Kaggle digit dataset to classify written numbers.

Our goal is to implement digit recognition using K Nearest Neighbors (KNN) and a Support Vector Machine (SVM) and compare the two approaches. A stretch objective is to enable data that isn't a part of the Kaggle dataset to work with our algorithms. This would require some form of normalization in the input image for consistency.

2 Progress

2.1 KNN

We have made significant progress in both the KNN algorithm and the SVM algorithm. Our KNN algorithm is completely implemented as is cross validation for it.

To compute the K nearest neighbors, we take the euclidean difference between the feature we are classifying, X_1, and every feature in our training set.

$$\sum_{X_2 \exists N} (X1 - X2)^2$$

Once we have calculated the K closest features, a simple voting method is used to classify X_1.

In order to select a value of K we do cross validation on a subset of the training set. We begin by passing a list of potential K values we are considering. For each item, K_i in this list, we hold out K_i samples from our training set and construct our classifier. We then attempt to classify the K_i samples and record the error rate.

The K value that led to the lowest error rate is used to build our final classifier.

2.2 SVM

TODO