Machine Learning Milestone

Deric Pang
dericp@cs.washington.edu

Saidutt Nimmagadda
nimmas@cs.washington.edu

Abstract

Our project investigates the task of recognizing handwritten digits. Our initial project proposal involved reading three papers [1]-[3] and determining which techniques to use to achieve our task.

1 Project Description

1.1 yolo

swag swag swag

2 Progress

We have implemented a classifier using k-nearest neighbors regression. We found the k point(s) with the smallest Euclidean distance(s). Given runtime constraints, we ran the algorithm on a fraction of the dataset (1000 rows) with validation blocks of size 200 rows. For each row of the validation set, we found the k nearest neighbors. Then, for each of these neighbors' classifications, amongst the k nearest neighbors, we find the classification with the lowest average distance. This classification will be our prediction $\hat{y}$ for the given x_i.

We then compare our $\hat{y}$ to the true classification. The number of incorrect predictions divided by the number of total predictions is validation error. As we iterate over our validation blocks, we try different values of k (1, 5, 10, 50, and 100), and then graph the corresponding validation errors.

References

[1] LeCun, Yann, et al. "Comparison of learning algorithms for handwritten digit recognition." International conference on artificial neural networks. Vol. 60. 1995.

[2] Maji, Subhransu, and Jitendra Malik. "Fast and accurate digit classification." EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-159 (2009).

[3] Sundaresan, Vishnu, and Jasper Lin. "Recognizing Handwritten Digits and Characters." (1998).