Project Milestone

Morgan Baker

11/8/16

1 Abstract

The project I am working on is an old machine learning problem. It is the problem of teaching a machine to recognize handwritten numbers. I am trying to improve on the problem, first by creating

2 Intro

For the sake of data, I am using the dataset from the Mixed National Institute of Standards and Technology (MNIST). There are two datasets, one of which is test, and the other is to train. The training set has 785 columns, one of which being the actual digit represented. and the other 784 columns are to show values of a 28*28 grid. the other file, test.csv, forggoes the digit column, making the machine create an educated guess.

3 Background

Due to being one of the oldest problems in machine history, there is a lot of background data to go off on. In the future, I will find suitable papers that have already covered this topic.

4 Methodology

In order to hopefully figure out this problem, I will be testing the dataset using Python code. I have already created a script to figure out how the average number in train.csv looks. This takes the label digit, and maps every pixel from those columns onto the same plot. I also pplan on normalizing the values, seeing as 784 columns each with a possible number from 1-255 might be too big for the interpreter to handle. I then plan on creating multipple leraning algoithms, and taking the best one. The algorithms I am thinking of are KNN, Linear Regression (as a classifier), and a Support Vector Classifier (SVC).

5 Experiments

None of the algorithms have been implemented, but I have gotten the averages of all numbers, and can point out outliers in the different numbers.

6 Discussion

Once there is something to be discussed, that part will go here. Also, analysis of the algorithms will be here too.

7 Conclusion

I'm not able to make any conclusions yet, seeing as I haven't really done much on the project.

8 References

References will go here at a later date.