MNIST Digit Recognizer

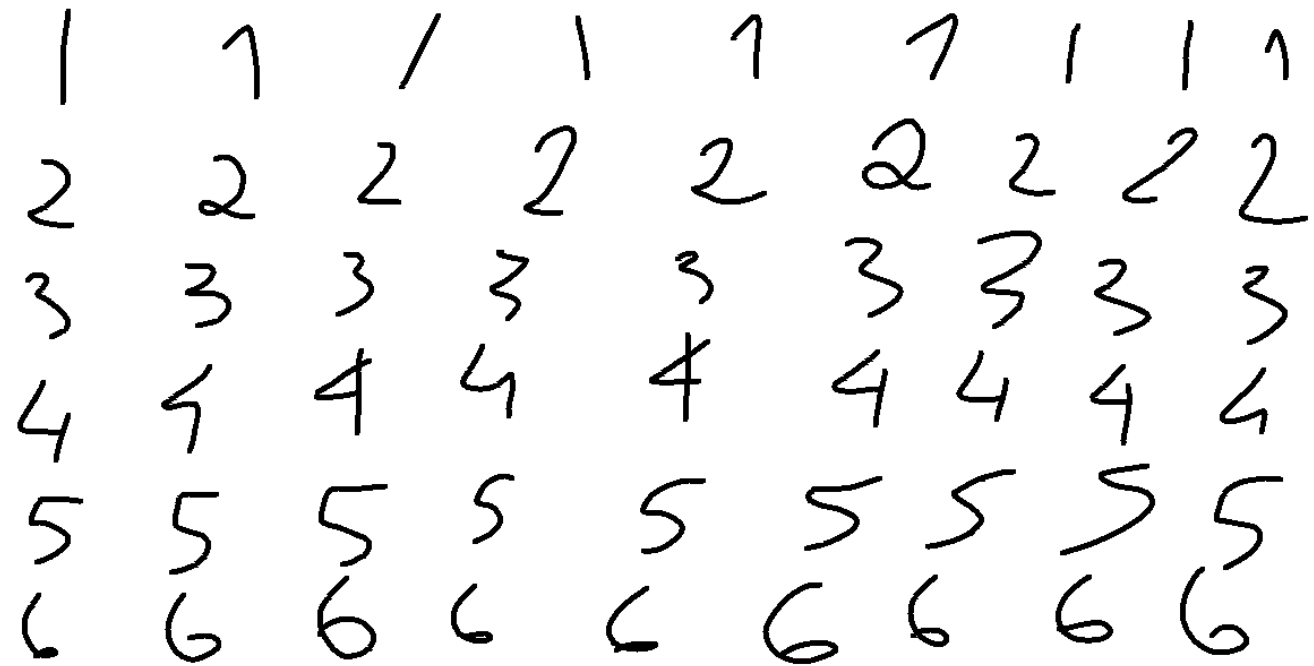MORGAN BAKER

Introduction

Handwritten digits

MNIST Dataset

A dataset created by the National Institute of Standards and Technology

There is a list of methods that have been used.

Concept

Taking the data and trying to get the lowest error

Background and Hypothesis

There have been many attempts to make an accurate model

http://yann.lecun.com/exdb/mnist/index.html

Methodology: K Nearest Neighbors

Ran a simple K nearest neighbors

Took about an hour or two to complete.

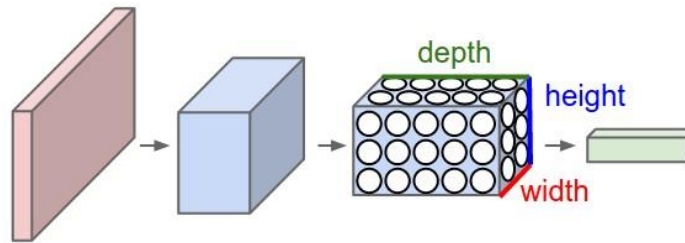94% accuracy.

Methodology: Convolutional Neural Network

Use a convolutional neural network to read the digits

Neural Networks for Images

Three layers
- Convolutional
- Pooling
- Fully Connected

About 96% accuracy

Methodology: Random Forest

The last algorithm on my list

Creates a bunch of different decision trees and takes the majority vote.

Haven't tested it yet

Testing and Hypothesis

The goal is to have a error rate of <2%

K Nearest Neighbors unfortunately does not meet that goal, but Random Forest hopefully could.

Sources

"Random Forests Leo Breiman and Adele Cutler." *Random Forests - Classification Description*. Berkeley.edu, n.d. Web. 07 Dec. 2016.

David Dalisay Junior Data Scientist at Booz Allen Hamilton, Michael McCarthy Cofounder, Kevin Beaulieu Software Engineer, and Rizwan Mallal Chief Operating Officer, Forum Systems. "Machine Learning Mastery." *Machine Learning Mastery*. Jason Brownlee, 14 Nov. 2016. Web. 07 Dec. 2016.

"Digit Recognizer." *Digit Recognizer | Kaggle*. Kaggle, 25 July 2012. Web. 07 Dec. 2016.

LeCun, Yann, Corrina Cortes, and Christopher J.C. Burges. "THE MNIST DATABASE." *MNIST Handwritten Digit Database, Yann LeCun, Corinna Cortes and Chris Burges*. NYU, n.d. Web. 07 Dec. 2016.