Project title: "Recognition of Handwritten Digits"

Data set: We will be using the "MNIST Database of Handwritten Digits" dataset referenced on the course website (http://www.kaggle.com/c/digit-recognizer/data).

Idea: In this project, we will implement the recognition of handwritten digits. Handwritten digit samples that we will train, cross-validate, and test our algorithm against will be stored as pixel data. For each sample, our intended output is an accurate classification of what that digit is.

Possible techniques that we could use in our implementation of an algorithm include neural networks, linear classification via logistic regression, and k-Nearest Neighbor regression. Based off the LeCun paper, it seems that the Neural Networking technique had the best performance, so that's an option we will look heavily into. We're also considering using a support vector machine.

Software to Write: We won't need to write software that will extract pixel data since it is already given in a CSV. If we decide to use a neural network, we will have to use TensorFlow. We will have to implement cross-validation and data visualization. If we don't use a neural network, we'll have to implement the algorithms that we first mathematically formalize.

Papers to Read: http://yann.lecun.com/exdb/publis/pdf/lecun-95b.pdf

https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-159.pdf

http://cs231n.stanford.edu/reports/vishnu_final.pdf

Milestone: By the 24th, we expect to have read all three papers and formally determined which algorithm to use. We also intend to have a naïve implementation completed. By naïve, we mean that we will have some sort of algorithm implemented (not necessarily the most efficient/complex) and will have test results after training on at least a subset of the data.