Final Project Update 1

Kacie Salmon

Progress Since Proposal

Since the proposal, I have worked to manipulate the large dataset and build the least squares classifier model. I have developed it in the one-vs-all style and added ridge regression. I found that I needed ridge regression to make my matrices invertible and I also used SVD to reduce the dimensionality of the matrices. I ran into issues trying to run the dataset with the full training set of 60,000 datapoints, as I would try to run it in Jupyter notebook, and the kernel consistently died. I have tried using Python locally, with the Spyder IDE, but had issues with it running. For example, it cannot import NumPy. In the meantime, I have been able to run the least squares classifier program with a smaller training set of just 10,000 datapoints and evaluated it on the test dataset. It currently misclassifies roughly 2000 datapoints in the test dataset varying slightly dependent on the lambda used for regression. But it is working as designed as a one-vs-all classifier with 10 weight vectors, one for each label. They are all then applied and the argmax is assigned as the correct label.

Next Steps

The next major step I will take is finding a way to handle the large dataset of training data, either through local Python utilization or within Jupyter with methods we have learned. After this I will finish the least squares models and set up an automatic lambda regression selection. I'll then move on to the support vectors and neural networks algorithms and comparing the performance of each. I will then cross-validate each of these classifiers to find an average error rate and compare the general performance of each.

Link to Relevant Results

https://github.com/klsalmon/ECE532_FinalProject

https://github.com/klsalmon/ECE532_FinalProject/blob/gh-pages/Least%20Squares%20with%20Ridge%20Regression%20Model.py