Analysis of Approaches

In the end, we decided to use KNN, Decision Trees, and Random Forests (as our ensemble). Of the three, KNN was the most simple in nature, though all three were relatively simple. All three algorithms required no pre-processing on our part and we were able to pass in the data that was given to us.

As previously stated, FNN was very easy to use. We simply passed in the training data and produced a vector of predictions for the testing data. We were able to produce more iterations using this algorithm, as it was by far the fastest in producing predictions.

Decision Trees were also useful in predicting the labels, though they produced the worst models of the three. Again, we were able to pass in the data as it was given to us and produce a decision tree model. It took considerably longer than FNN, but less time to train than the random forest.

Our Random Forest model produced our best model of the three, but only slightly. It beat out FNN in terms of accuracy by a very slight percentage, but it took much longer to train (Kaggle results took about 2 hours to train). Like the previous models, our data did not need to be processed beforehand.

Of the three, I would argue that FNN was the best algorithm we used, both in accuracy and efficiency.

Conclusions

Going into this project, we were a bit nervous. We ended up choosing this project because the other 'simpler' projects had been taken. However, we soon became engrossed with the project. After the presentations, I might say that we had the simplest project, purely in terms of how little pre-processing we did. Other groups did a great deal of pre-processing in order to pass in their data to the algorithms, while we were able to produce good results with little to no pre-processing. That being said, our data had a large number of features, and our algorithm training speeds suffered as a result.

We had fun with this project, and we both learned a lot about image recognition. Future work on this project might include pre-processing and implemenation of neural networks and SVMs. Both of these algorithms require additional data manipulation.

All in all, we feel we succeeded in our attempts to produce good predictive models. We both definitely look forward to similar projects in the future!