Team

Akhil Surya Vamshi(120050067)
Jaseem Umar(120050081)
Zubin Arya(120050036)
Haren Saga(120050072)

Progress

We have parsed the data, understood the sources of data for making a better sense of result and have used the Linear SVM with Stochastic Gradient Descent optimization from sklearn library to obtain a error rate of 6%.

Plans

Performance

The algorithm currently takes about 5 minutes to learn from thousand handwritten images which is considerably high. Hence we plan on reducing the feature vectors used in classification as preprocessing and hence reducing the time on learning.

Error rate

The paper we are using goes on to improve the error rate to 0.6%. We plan to understand and reduce our error rate through techniques discussed in the paper which include constructing neural networks for classifying the data. We then plan to compare the results obtained with our current condition. We also were hoping to explore preprocessing techniques like deskewing to improve the accuracy

Pointers to literature

1. http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf
2. http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html
3. http://yann.lecun.com/exdb/mnist/t10k-labels-idx1-ubyte.gz
4. http://yann.lecun.com/exdb/mnist/t10k-images-idx3-ubyte.gz