**Using Deep Learning Methods to Analyse the Minst Dataset**

In this project, the Minst data set was used to train a neural network with one hidden layer and the relevant statistics was graphed. The Adam optimizer was used.

The Minst data set is used to train a machine to detect handwritten digits. Given the problem of image detection, we feel a deep learning method will suit the best.

Various types of deep learning methods involve feed forward networks, recurrent networks, auto encoders, GANs, convolution networks etc. Among these, auto encoders are used to encode images only and GANs are used to create realistic false data points. For our purpose of training a feed forward or convolution NN will be best suited. Since the images we use are not sequential, a recurrent network may unnecessarily take up too much time and memory. We use the simplest form, a feed forward neural network.

**Data Exploration**

The Minst data set was obtained from Keras datasets. Initial observation showed the following data distributions for training and test datasets respectively. The size of each image was found to be 28X28 px.

number of unique items: {0: 5923, 1: 6742, 2: 5958, 3: 6131, 4: 5842, 5: 5421, 6: 5918, 7: 6265, 8: 5851, 9: 5949}

number of unique items: {0: 980, 1: 1135, 2: 1032, 3: 1010, 4: 982, 5: 892, 6: 958, 7: 1028, 8: 974, 9: 1009}

image size: 28

Thirty random data points were obtained and were visually identified as numbers. The dataset appears clean with all dictionary items mapped to images of numbers.

## Neural Net Construction

A sequential neural network with one hidden layer was created and plotted. We used the relu activation function,an input dimension of 28X 28, which is the image size and dropout functions corresponding to each layer. The initial dropout value was .45 and the size of the hidden layer was 256 units. The output dimension was equal to the number of labels, which is 10. The batch size was set to 128.

## Analysis and Statistics

The accuracy and loss functions were graphed and the model accuracy was found to be 98.0 percent.

We then played around with the batch size,and the dropout value. We found that setting the batch size to smaller numbers (20 and 56) did not improve the accuracy. Setting the dropout to 0 decreased the accuracy to 97.7 percent. Increasing the dropout value to .8 also decreased the accuracy to 97.7.

We thus concluded that while the accuracy did not change substantially, the initial parameters of batch size 128 and dropout .45 were optimal. We also observed that the elbow of the accuracy curve was reached after five epochs. Hence, we deduced that running the fit for more than ten epochs was unnecessary.

## Conclusion

We concluded that using the Adam optimizer in a neural network was an efficient way to train and detect the machine on the Minst dataset. However we observed that the randomly chosen data points were eligible by visual examination and very clean. Hence, using a convolution network would further reveal relations within the patterns and may turn out to be effective when the input numbers are more distorted and when main features will play a big role in detection.