Machine Learning
Assignment 5

Name: Sai Saketh Aluru

Roll No. 16CS30030

Part - 1:

Neural Network with one hidden layer, having ReLU as its activation function.

1. The given data is preprocess and converted into one hot encoded vectors. Here, I have considered all the distinct words appearing in the text as input features, i.e full vocabulary.
2. The loss function is cross entropy loss function. Here, since the cross entropy loss function requires the output predicted by the classifier lying between 0 and 1, a sigmoid layer is used to squish the output of the final layer into the required range.
3. Mini Batch gradient descent is used, with number of batch sizes taken as 8. The input data is first split into an 80-20 ratio of train data and test data. The train data is further split into small batches of roughly equal size. These batches are used in training and updating the weights
4. The threshold value is taken as 0.5, the inputs for which output of the classifier is greater than this is taken as spam and less than this is considered to be as ham. Since the final layer is a sigmoid, the outputs are always between 0 and 1.

The plot obtained for this part is:
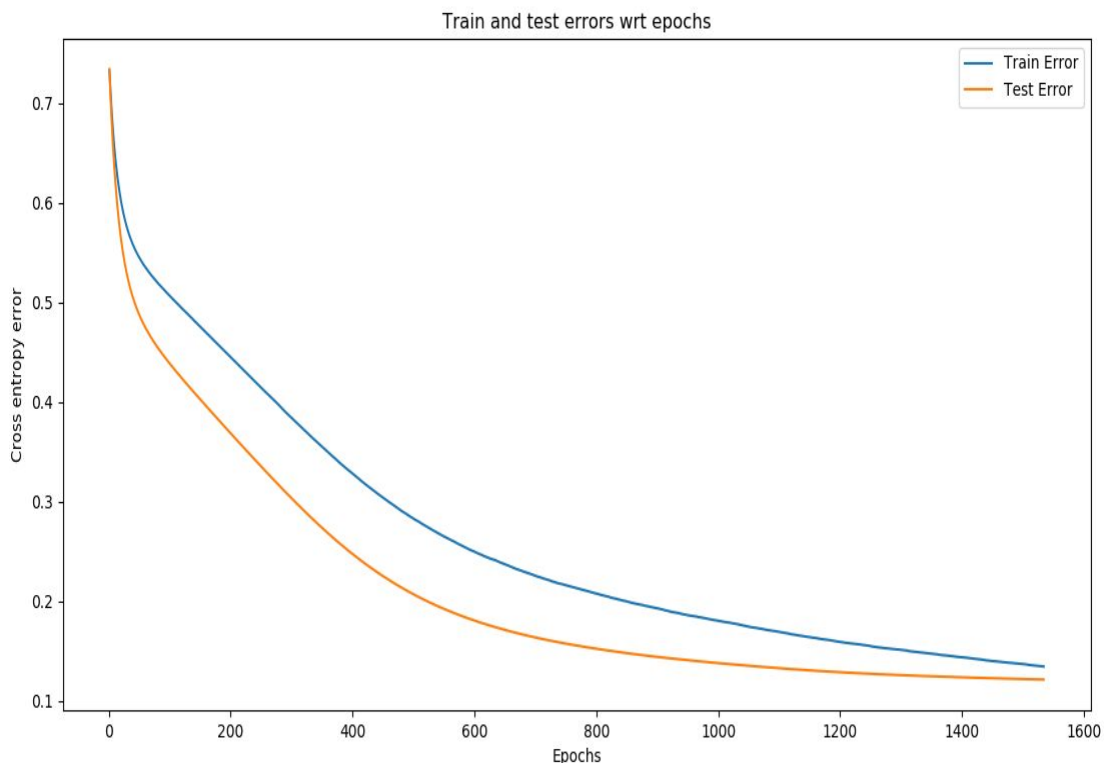
The above plot corresponds to the following hyperparameters:
1. Learning rate alpha = 0.1

2. Number of batches = 8
3. Number of neurons in the hidden layer - 100
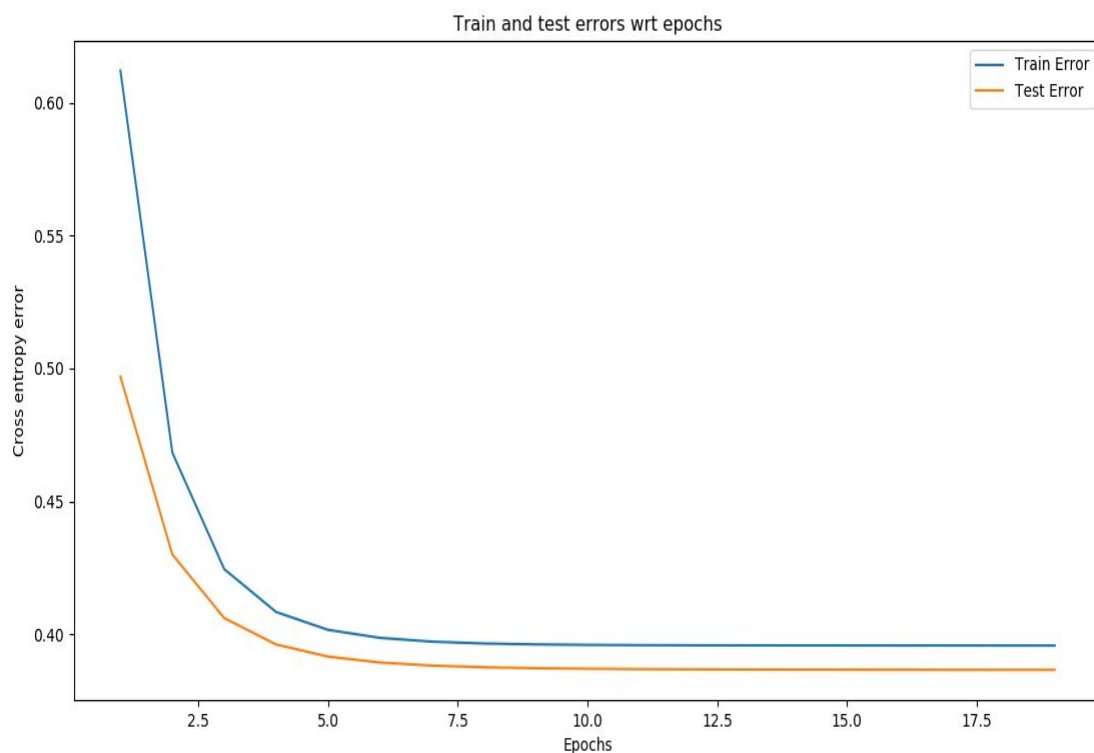4. Optimisation - SGD with cross entropy loss

The final test set accuracy obtained after roughly 1500 epochs is -- 97.39910313901345

Part 2:
Neural Network with two hidden layers, both having sigmoid activation and with a softmax layer on top of the output layer

1. The given data is processed in the same way, except now the labels are maintained as a one hot vector of 2 dimensions, [0 1] for ham [1 0] for spam.
2. Here also, categorical loss function is considered, and mini batch gradient descent is used for training the model.
3. The model contains two hidden layers, the sizes of which are taken as command line arguments.

The obtained plot for this part is:

The hyperparameters corresponding to this plot are:
1. Learning rate -- 0.1
2. Number of batches -- 5
3. Number of hidden layers -- 2, first one of size 100, second one of size 50
4. Optimisation - SGD with categorical cross entropy loss

The final test set accuracy obtained using this model is -- 86.99551569506727