CS 760

Homework 3

Submitted by: Sneha Rudra

Part 1

- The file 'nnet.py' implements a program to learn a neural network using stochastic gradient descent with online training. 'nnet.py' is callable from the command line and accepts the following command line arguments- l h e <train-set-file> <test-set-file>

  where l is the learning rate, h is the number of hidden units, e is he number of training epochs, <train-set-file> is training set file name and <test-set-file> is the test set file name.

- Script file 'nnet' can be used to accept the above command line arguments, to invoke nnet.py and the python interpreter. For example: If nnet.py needs to be executed for the heart dataset, l= 0.1, h=20 and e=10 using the nnet script, then the following command should be executed from the terminal (in the appropriate path where nnet.py, the training set arff file and test set arff file are located):

  [sneha@royal-18] (25)$ nnet 0.1 20 10 heart_train.arff heart_test.arff

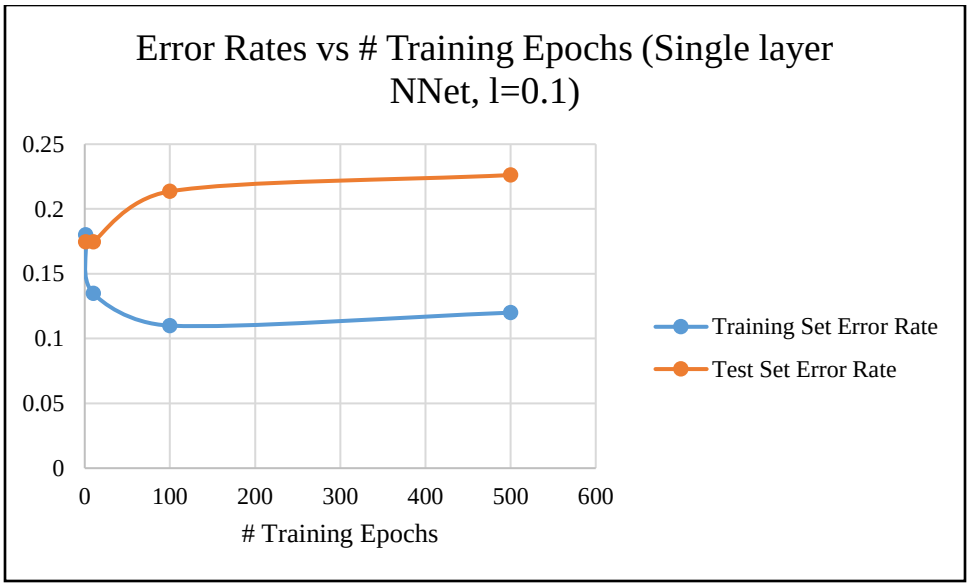
nnet.py reads the training set and test set files in the ARFF format. They can parse ARFF files that are similar in format to the example ARFF files provided (heart_train.arff, heart_test.arff, lymph_train.arff, lymph_test.arff).

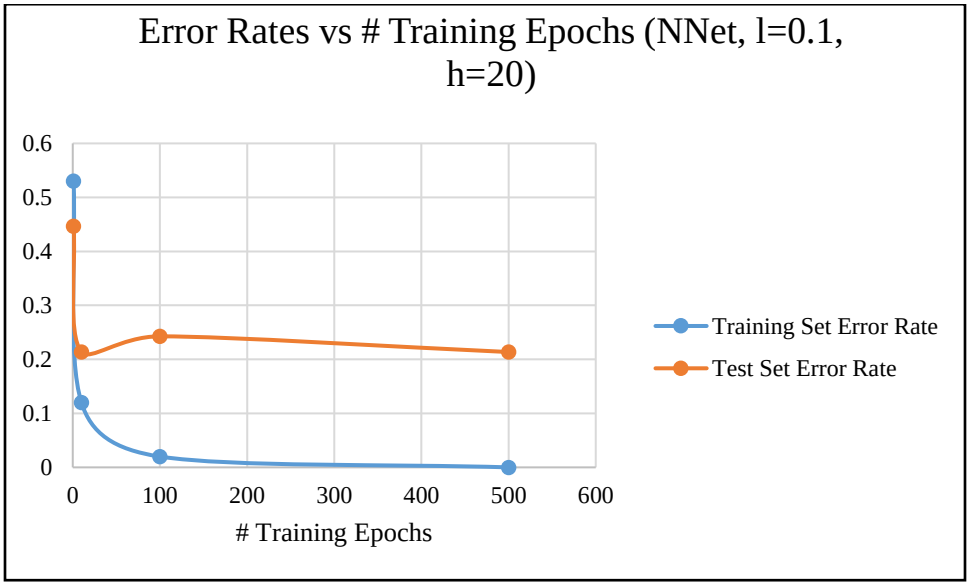
Part 2

Using heart data set you should make two graphs showing error-rates versus the number of training epochs. For the first graph, plot training and testing error rates for a single-layer network trained for 1, 10, 100 and 500 epochs, using a learning rate of 0.1. For the second graph, plot similar curves for a network with 20 hidden units. Be sure to label the axes of your plots.

$$Test\ Set\ Error\ Rate = \frac{\#\ Test\ instances\ incorrectly\ classified}{Total\ \#\ of\ test\ instances}$$

$$Training\ Set\ Error\ Rate = \frac{\#\ Training\ instances\ incorrectly\ classified\ after\ the\ last\ epoch}{Total\ \#\ of\ training\ instances}$$





Part 3

Use the activation of the output unit as the measure of confidence that a given test instance is positive, and plot ROC curves for both the heart data set and the lymphography data set. Be sure to label the axes of your plots.

$$True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$False\ Positve\ Rate = \frac{False\ Positive}{False\ Positive + True\ Negative}$$