1. For this problem, use the data contained in the file `'moonDataset.csv'`. The first three columns contain the values of three features, and the fourth column has the label (either 0 or 1). It consists of 200 samples. You will need to partition the dataset into two datasets. The first 150 samples will be used for training and the remaining 50 samples will be used for testing.

   (a) Starting with the training dataset with 150 samples, construct 50 training datasets (with 150 samples in each) using bootstrap method.

   Ans: See the Matlab code provided at the end.

   (b) For each of the 50 datasets from part (a), construct a feedforward network with one hidden layer consisting of 10 hidden nodes, and train it using the dataset for binary classification. Compute the error rate of the neural network using the test dataset with 50 samples, and plot the histogram of the error rates for the 50 neural networks. Each dataset is used to train a model only once.

   Ans: The histogram of the error rates for 500 datasets generated from the original dataset is shown in Figure 1. As you can see, although the error rate is small most of the time (about 56 percent of them), it can vary quite a bit and is as high as 25 percent in some case.

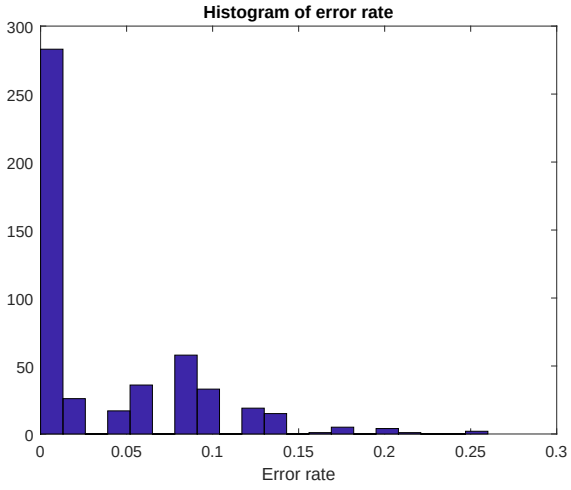

Figure 1: Histogram of 500 error rates.

   (c) Use bagging with ensemble size $m \in \{5, 10, 15, 20\}$ to design new binary classifiers. Plot the error rate of the new classifiers as a function of ensemble size m.

   Ans: The average from 200 runs is plotted in Figure 2. As you can see, the average error rate decreases with the ensemble size.

2. **[Optional Problem: Will not be collected]** The file `'DailyDelhiClimateTrain.csv'` contains the climate information over a period of 4 years. The first column contains the date, and the columns 2 through 5 include the values of `mean temperature`, `humidity`, `wind speed`, and `mean pressure`.

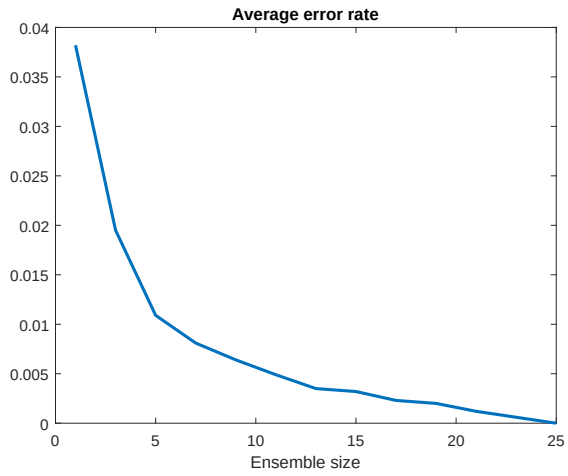

Figure 2: Error rates as a function of ensemble size (average of 200 runs.

Construct 3 training sequences using the data for the first 3 years (one sequence for each year), and use the data for the fourth year for testing and prediction. Using the training sequences, train a long short-term memory (LSTM) network with 128 nodes, and forecast the future values following the testing data. (The prediction would not be very impressive given that we are using only three sequences for training, but you will see a trend in the forecast.)

Ans: The plots of forecasting are shown in Figure 3.

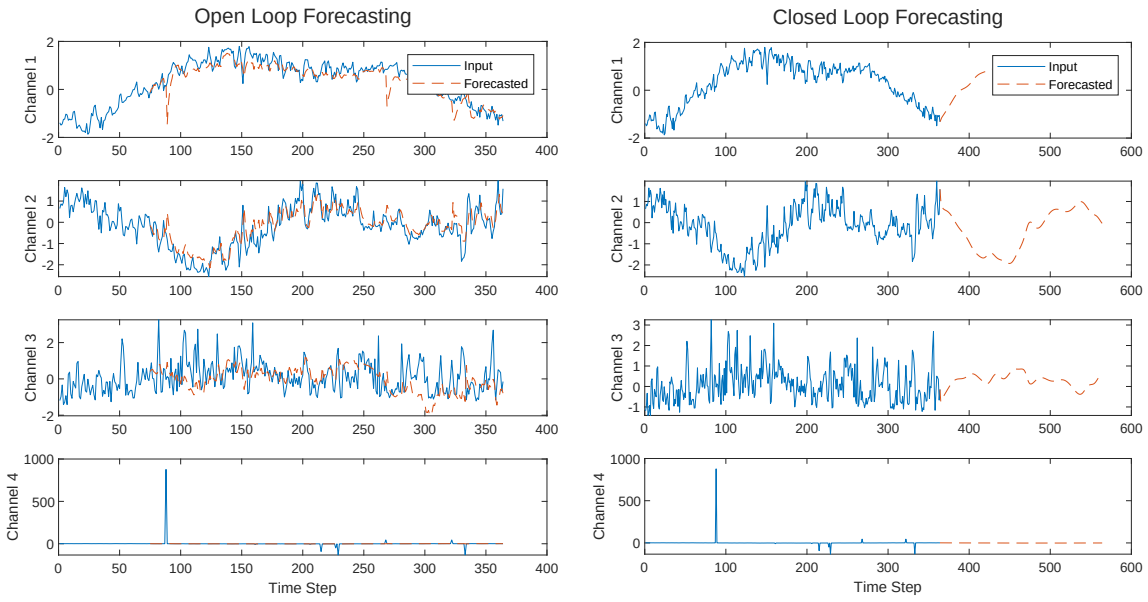

Figure 3: Plots of forecast data.