EECS 416 Alternative Case Study: Due 4/29/2019

Classifying Epilepsy EEG data using SVM

The goal of this study is to try develop a classifier to classify EEG signals to predict whether a particular patient will experience an onset seizure episode. The technique to be explored is the Support Vector Machine (SVM), both linear and nonlinear.

Read the paper "Epileptic Seizure Detection Using Genetically Programmed Artificial Features" ---particularly Sections 1, 2A and 2B---to learn the nature of EEG data and the classification problems to predict epilepsy seizures using EEG data. The classification method used in the paper is based on genetic algorithm, but you will be using SVM in this case study.

Read "Support Vector Machines for Mining Data" that summarizes the principle and techniques of SVM as a robust classifier.

Then formulate the EEG signal classification problem using a Linear SVM classifier. Solve the linear classifier problem using the revised simplex method (rsimplex) or the interior point method (or both and compare the speed and costs of the two methods for extra credits). Compare the performance of your linear classifier with the simple classifying scheme described below.

Then explore whether you can improve the classification performance using a nonlinear classifier. You may use FMINCON or an interior-point based system as an optimizer here.

Below is a description of the EEG data that you will be using, as well as the result of a simple classification scheme.

Original Data:

- EEG time series data obtained from epilepsy patients
- There are 2 groups of data:
  - Set N – EEG time series data obtained during non-seizure period (Epilepsy training_n set); and
  - Set S – EEG time series data obtained during seizure period (Epilepsy training_s set);
- There are 100 epochs for each data group ((Epilepsy testing set);
- The length of one epoch is 4096 samples
- The sampling rate of the data is 173.61 Hz
- Band-pass filter is 0.5-85 Hz

Given Data:

- Averaged power spectrum of 20 spectral sub-bands: 0-1, 1-2, …, 19-20 Hz, of the EEG time series data

- File training_n.dat contains 50 averaged power spectral sub-bands of the EEG time series data of the set N (in each column of the data matrix) to use for the training step
- File training_s.dat contains 50 averaged power spectral sub-bands of the EEG time series data of the set S (in each column of the data matrix) to use for the training step
- File data.dat contains 100 averaged power spectral sub-bands of the EEG time series data of both sets N and S (in each column of the data matrix) to use for the test
- Data format is ASCII

Preparation of Data:

- Remove the DC content of each EEG epoch
- Normalize each EEG epoch
- Apply the discrete Fourier transform (DFT) to each EEG epoch
- Compute the averaged power spectrum of each spectral sub-band

Illustrations

Figure 1: Examples of EEG time series data of both sets N and S.

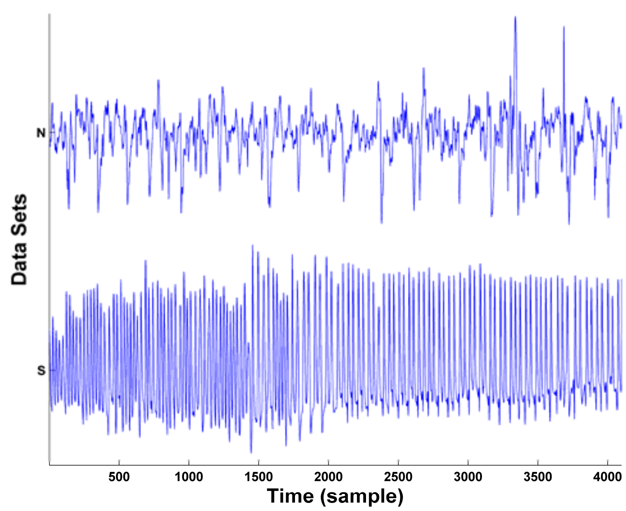


Figure 2: The corresponding power spectrum of the EEG time series data shown in Figure 1.

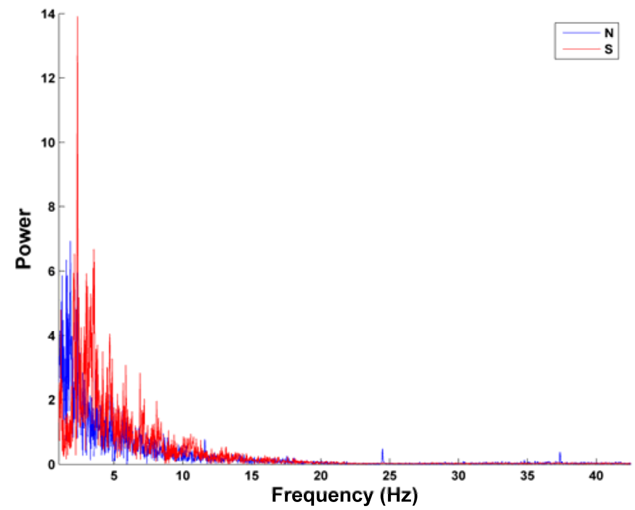
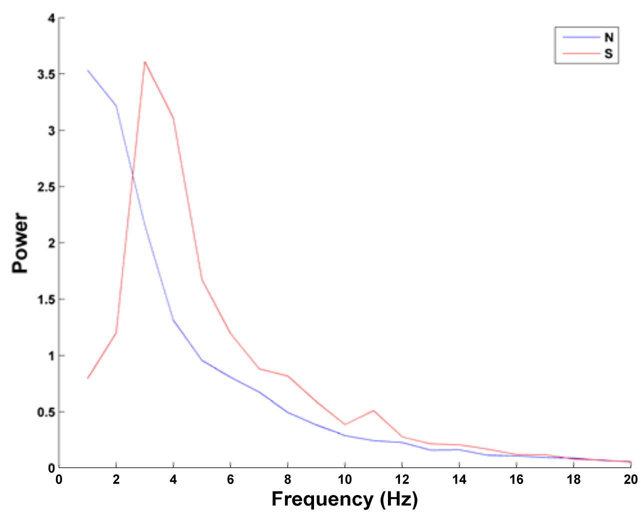
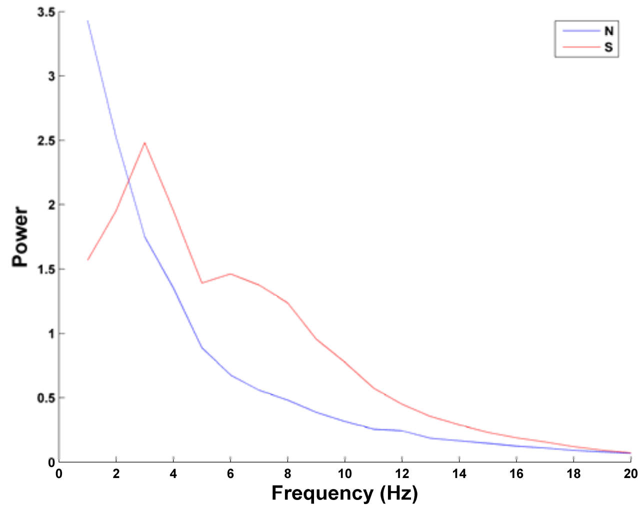


Figure 3: The averaged power spectrums of the corresponding power spectrum of the EEG time series data shown in Figure 2.

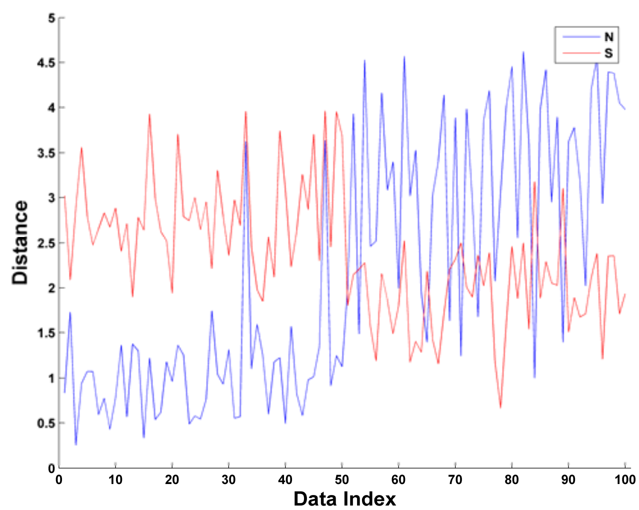


Experiment:

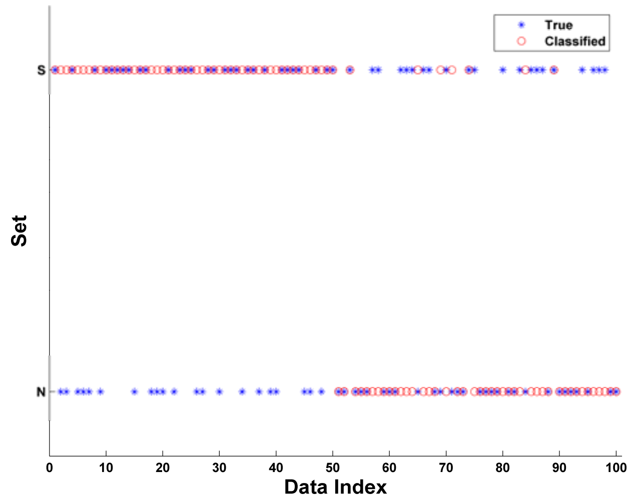
- Compute the average values of the feature (the averaged power spectral sub-bands) from the training sets of each data set (N and S). The results are shown in Figure 4.



- For each EEG time series of the test data, calculate the Euclidean norm distance between the feature of the test EEG time series and the average values of the feature of both sets N and S.
- The results of such distance are shown in Figure 5.



- The results of the classification are then shown in Figure 6. **(NOTE: Figure 6 cannot be disclosed because the students will know the true membership of the EEG time series data.)**



- The result of this approach: The true positive classification/detection is 57%.