CS4070 Random Signal Processing

Final Matlab Assignment

September 16, 2019

David M.J. Tax

1 Classification of Electrocardiograms (ECG)

1.1 Problem Setting

We are working in a hospital where we want to diagnose patients if they have some heart disease. In order to detect this disease, an electrocardiogram (ECG) is being recorded. In order to detect if a person is healthy, you are asked to implement a decision rule in Matlab.[1]

Please write a *short* report, 2-4 pages, that contains the answers to the questions. You don't need to supply all code, but give the core code, and the figures, in the text.

1.2 Modeling of the Classes

In this assignment we are going to make an automatic classification system to classify between ECGs of healthy and diseased persons. The classes are denoted by 'healthy' or 'diseased' and we will model it as a Bernoulli random variable `H`. The value `H=1` indicates that the object belongs to the 'healthy' class, and `H=0` indicates that the object belongs to the 'diseased' class.

The automatic classification is done as follows. A few healthy and diseased patients are taken. ECGs are taken from all of them, and each scan is considered a realization of a random process that is WSS. In order to classify the ECGs, we first compute the autocorrelation functions $R_X(k)$ for each of the scans, and normalise them in order to make them robust against amplitude variations between scans. We hope that the normalised autocorrelation functions are a bit different between healthy and diseased. By visual inspection we will select a value k^* where the the difference between healthy and diseased is the largest. Then we will use the values of the autocorrelation $R_X(k^*)$ to make a predictor of the health status.

1.3 Data Available on Brightspace

The data used for the exercise are downloadable from Brightspace. The file `trainecg.mat` contains the following 2 matrices:

- `healthy` contains the scans from normal people,

[1] Some people like to work in Python. That is no problem, but the provided datasets are stored in Matlab format, so then you have to convert it to Python data.

- `diseased` contains the scans from diseased people.

Remark: Note that when you load a file using the command

```
>> load trainecg.mat
```

the matrices will be present in your workspace. Use the command

```
> who
```

to see the variables in your workspace.

2 Extraction of the Autocorrelation function

Exercise 2.1 Compute (estimates of) the autocorrelation function $R(k)$ from the ecgs from the file `trainecg.mat`. Normalise them by dividing them by $R_X[0]$:

$$\tilde{R}_X[k] = \frac{R_X[k]}{R_X[0]} \tag{1}$$

Plot the estimates in overlay, and give the functions of the healthy scans a different color than the functions of the diseased ECGs.

Exercise 2.2 What value of k would you propose to use for the classification? In other words, what is k^* that can best distinguish between healthy and diseased?

3 Classification using the Autocorrelation function

Through the following questions you will develop a MAP (maximum a posteriori) classifier based on the value of the autocorrelation function $\tilde{R}_X(k^*)$ at k^*.

Exercise 3.1 Plot the histogram of $\tilde{R}_X(k^*)$ for the healthy and the diseased ECGs. Can you (visually) separate the two classes perfectly?

For the MAP estimator, we need the a priori probabilities of the two classes, i.e. $P[L = 0]$ and $P[L = 1]$. We estimate these probabilities by using the assumption that the scans are uniformly sampled, i.e. the fraction of normal and ILD samples is representative for the real a priori probabilities

Exercise 3.2 Estimate $P[H = 1]$ and $P[H = 0]$ from the relative frequencies of healthy and diseased ECGs that were given in the file `trainecg.mat`.

We now classify every scan on the basis of $\tilde{R}_X(k^*)$ as `H=0` or `H=1` using the following Maximum A Posteriori (MAP) decision rule:

$$D(r) = \begin{cases} 0 & \text{if} \quad P(H = 0|\tilde{R}_X(k^*) = r) \geq P(H = 1|\tilde{R}_X(k^*) = r) \\ 1 & \text{if} \quad P(H = 0|\tilde{R}_X(k^*) = r) < P(H = 1|\tilde{R}_X(k^*) = r) \end{cases}$$

We model the (conditional) random variables $\tilde{R}_X|H = 0$ and $\tilde{R}_X|H = 1$ as having a Gaussian PDF. **Hint**: to get the decision rule from the PDFs, you need to use Bayes' rule.

Exercise 3.3 Estimate the expected value and variance of the PDFs $f_{\tilde{R}_X|H=0}$ and $f_{\tilde{R}_X|H=1}$, and plot the resulting PDFs in a figure. Compare it to the histograms you made in 3.1.

Exercise 3.4 Classify the scans that are stored in `healthy` and in `diseased`. What fraction of the ECGs correctly classified?

Exercise 3.5 Classify the scans that are stored in the file `testecg.mat`. Show the predicted labels.