Digital Signal Processing
Instituto Superior Técnico
Lab assignment 4 – Anomaly detection in energy time series data
Authors: Margarida Silveira and Jorge Salvador Marques.

Introduction

In this assignment, we will address anomaly detection in energy production data. The goal is to model the behaviour of normal energy production along time and subsequently to identify anomalies based on deviations from the normal behaviour. We will model the normal energy signal $x(n)$ as the output of an autoregressive model (AR). For example, in long term prediction, the model can be used to predict each sample from a previous sample, plus a prediction error or residual:

$$x(n) = ax(n - N) + r(n) \tag{1}$$

Alternatively, in short-term prediction, this model can be used to predict each sample from the previous P samples, plus a prediction error.

$$x(n) = \sum_{k=1}^{P} a_k x(n - k) + e(n) \tag{2}$$

The coefficients of the AR model can be obtained by minimizing the sum of squared residuals, using the least squares method.

Notes

For this lab assignment, only short answers to the questions below and the Matlab code written to answer those questions are required. The deadline for submission in fenix is <u>one week after</u> the day of the respective lab session at 23:59. The items that you should address are marked in bold font, e.g. **R1.a)**.

Experimental work

1. **Prediction**

R1.a) We will start with a model of long term prediction. Write a Matlab function that, given a vector with time series data $x(n)$ and a delay N, computes coefficient a for the following AR model

$$x(n) = ax(n - N) + r(n) \tag{3}$$

Write your own code, do not use any Matlab ready made functions.

R1.b) Load the train data `energy_train.mat` which contains energy production data from a residential solar panel, where samples were obtained every 15 minutes and where there are no anomalies. Use your code to compute a prediction using N=96, which means you are predicting the current sample from the sample of the previous day, at the same time. Plot the training data, the prediction you obtained and the corresponding residual. Comment on what you observe.

R1.c) Indicate the coefficient and the energy of the residual $r(n)$. Comment.

R1.d) Next, we will use a short term model for the prediction of the residual you obtained with the long term model

$$r(n) = x(n) - ax(n - N) \tag{4}$$

Write a Matlab function that, given a vector with time series data $r(n)$ and a number of previous samples P, computes the coefficients of the following AR model

$$r(n) = \sum_{k=1}^{P} a_k r(n - k) + e(n) \tag{5}$$

Again, write your own code, do not use any Matlab ready made functions.

R1.e) Use your code to compute a prediction for $r(n)$ using P=6. Use the predicted residual to obtain a new prediction for $x(n)$. Plot the training data, the new prediction you obtained and the corresponding residual $e(n)$. Comment on what you observe and compare.

R1.f) Indicate the coefficients and the energy of residual $e(n)$. Comment.

2. **Anomaly Detection**

R2.a) Write a Matlab function that detects anomalies in time series data given a time series and its prediction.

R2.b) Load the test data `energy_test.mat` and use the previous two models you developed to predict normal behaviour for test data. Use your function to detect anomalous periods of energy production from predictions of both models. Comment on your results and compare both models in what regards their ability for anomaly detection.

R2.c) When an anomaly occurs, the detection of the next period of energy production, which can be normal, is affected. How could you solve this problem?