Seizure Detection

Statistical digital signal processing assignment

Anjali Anand

Contents

1 Introduction

This report aims to understand how to model EEG signals, especially in the case of epileptic seizure activity. To tackle the given tasks, a brief literature study was done to identify methods and challenges. The given tasks were analysed and the best-suited approach was taken.

1.1 Seizure Detection

Seizure activity, recorded by EEG (Electroencephalography), is characterized as distinctive electrical disturbances in the brain's neural activity. EEG can capture these abnormal patterns, characterized by sharp spikes, high-frequency oscillations, and synchronous discharges. These features vary in intensity, duration, and location, providing valuable insights into the type and origin of seizures.

1.2 Autoregressive modeling

Autoregressive (AR) modelling (1) is a popular approach for analyzing EEG signals, as it allows for the estimation of spectral features and the characterization of underlying brain dynamics. However, determining the appropriate model order for AR modelling remains a crucial challenge. Francisco Vaz, et al. [1] explore the relationship between model order, segment length, and the characteristics of EEG segments. Furthermore, the study highlights the importance of using short segment lengths for AR modelling to ensure computational efficiency. It also suggests that the model order can be utilized to discriminate between rhythmic and non-rhythmic EEG segments, although challenges exist in selecting a constant model order for classification purposes. The AR(p) model is defined as:

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \ w_t \tag{1}$$

where $\mathrm{a_1,...,a_p}$ are the model parameters and $\mathrm{w_t}$ is white noise.

2 Motivation

Seizure detection using EEG is a critical component of diagnosing and managing epilepsy, a neurological disorder characterized by recurrent seizures. EEG can provide real-time monitoring of brain activity, allowing for the early detection of abnormal electrical patterns related to seizures. Detecting seizures involves analyzing EEG data for sudden, high-amplitude, and rhythmic activity. Timely identification of seizure activity through EEG enables medical professionals to provide relevant treatment strategies and improve the quality of life for individuals affected by epilepsy while reducing the risk of seizure-related injuries.

3 Proposed methodology

The data set provided contains a segment of the electrical activity in channel FP1 in which the seizure starts. It is to be noted that the data is corrupted by eye blinks at particular time instances. So, to start the analysis the noise needs to be removed as shown in Figure 1. To achieve this, the blink-related epochs were replaced with NaN values and then interpolated to fill in the missing values.

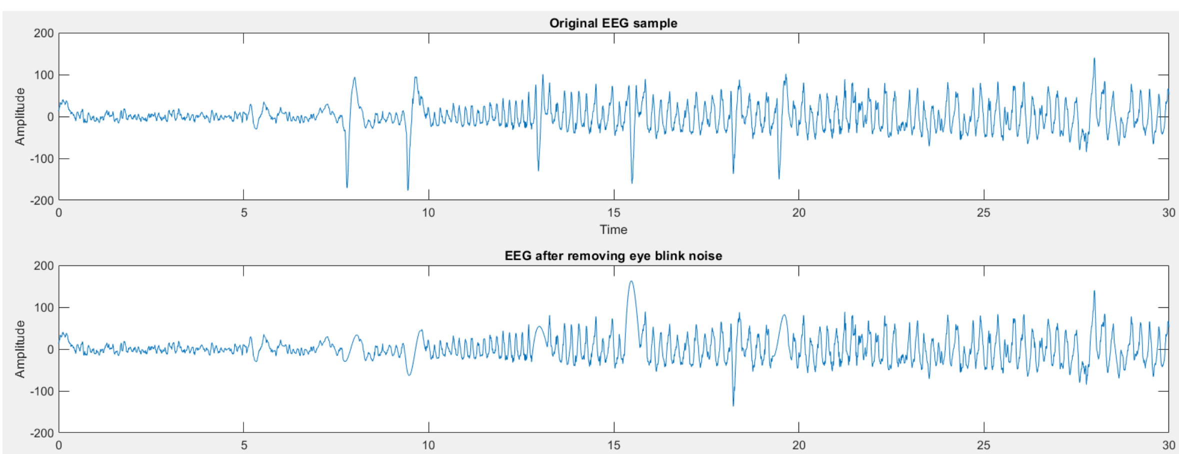

Figure 1: Before and after the removal of noise due to eye blinks

Five tasks were given and the following methodology was used to achieve them:

3.1 Show the signal is stationary

Since EEG represents a sum of localised electrical activity of neurons in the brain it cannot be considered stationary in time since a human being performs different activities. It is visible just by looking at the signal and can also be proved by sectioning the signal and analysing its statistical characteristics, like the mean. Another way to prove that the signal is not stationary is to model the signal segments and show that if the model parameters vary across the segments, the signal is not stationary.

3.2 Segment the signal

The choice of window size to segment can significantly impact the results and there is no one-size-fits-all protocol because the optimal window size depends on the specific goals of the analysis and the characteristics of the EEG data. Since the goal here is to detect epileptic seizures an appropriate window size must be selected that aligns with the duration of the start of the seizure attack. The International Federation of Clinical Neurophysiology (IFCN) Guidelines suggest a window size of 200-600ms to detect spikes, but a longer duration (1-2s) can be used to capture the context also. In this case, we use a rectangular window to segment the signal into segments of 2 seconds as it can be modelled sufficiently well using an autoregressive process [1]

3.3 Model the signal

To model the signal we use an autoregressive process. It is observed that a 5th-order AR model is generally sufficient for representing 1- or 2-second EEG segments, except for featureless background segments where higher-order models may be necessary [1][2]. For this case, we use the Yule-Walker method to estimate the model parameters. The equation to find the parameters $\mathrm{a_1,...,a_p}$ and b(0) are given below in (2) where $\mathrm{r_x}$ is the autocorrelation values.

$$r_x(k) + \sum_{l=1}^{p} a_p(l) r_x(k-l) = |b(0)|^2 \delta(k) \qquad ; \qquad k \geq 0 \tag{2}$$

3.4 Extract features that characterize the signal

To predict where the seizure starts the error between the original signal and the predicted signal using the AR model which was modeled for normal EEG characteristics can be used. For this case, we use the mean squared error which is given by (3) where n is the length of the segment, X is the original signal and $\hat{X}$ is the predicted signal

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X - \hat{X})^2 \tag{3}$$

3.5 Find where the seizure starts

To determine where the seizure starts the parametric changes between each windowed section can be analysed. It is evident that the segment with seizure activity will have more error compared to the original signal as it was modelled for a segment of EEG data without any seizure activity. So by considering a threshold and comparing the mean squared error of sections of the signal, we can essentially predict where the seizure starts if the error goes above the threshold.

4 Result

4.1 Show the signal is stationary

The signal was segmented with a segment length of 2 seconds and the mean of each segment was calculated. Figure 2 shows the means of each segment in contrast with the actual mean of the signal. From this, it is clear that the signal is non-stationary as the mean varies for each section.

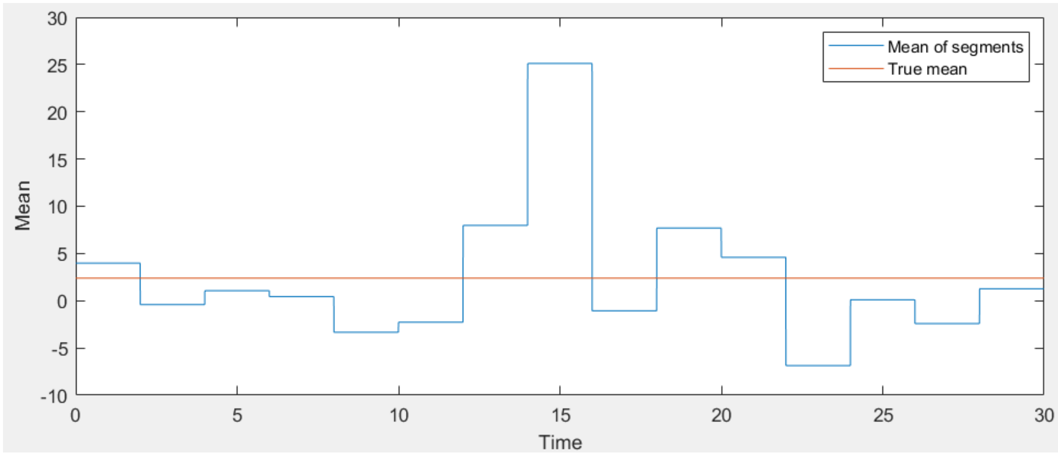

Figure 2: Mean of each segment generated

4.2 Segment the signal

The given signal was segmented with a length of 2 seconds using a rectangular window. Since the sampling frequency is 250Hz, each segment was 500 samples long. To visualise the segmentation, Figure 3 shows 6 segments with a length of 5 seconds.

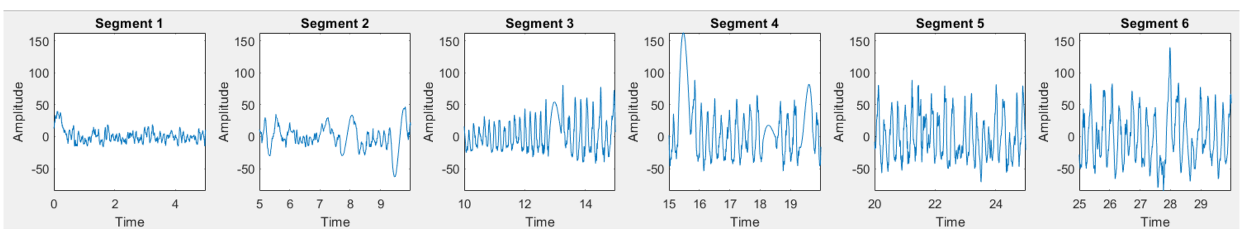

Figure 3: Segmented signal

4.3 Model the signal

The signal was modelled using a fifth-order autoregressive process using the data from the first two seconds of the given data. The model parameters were calculated using the Yule walker equation and were found to be as shown in Figure 4. The estimation fit was found to be 90.15%, and the mean square error was 1.745. We maintain the order of the system at 5 as increasing it further gives only a marginal increase in the fit percentage.

```
sys =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 1.732 z^-1 + 0.9896 z^-2 - 0.2204 z^-3 - 0.1837 z^-4 + 0.163 z^-5

Sample time: 0.004 seconds

Parameterization:
   Polynomial orders:   na=5
   Number of free coefficients: 5
   Use "polydata", "getpvec", "getcov" for parameters and their uncertainties.

Status:
Estimated using AR ('yw/ppw') on time domain data.
Fit to estimation data: 90.15%
FPE: 1.78, MSE: 1.745
```

Figure 4: Features of the model generated

4.4 Extract features that characterize the signal

The difference between the original signal and the prediction from the AR(5) model was found and its square was calculated and plotted in figure 5. From this, it is clear that the part of the signal related to the seizure activity will have a larger error as the prediction was made from the model which was calculated from the seizure-free EEG activity.

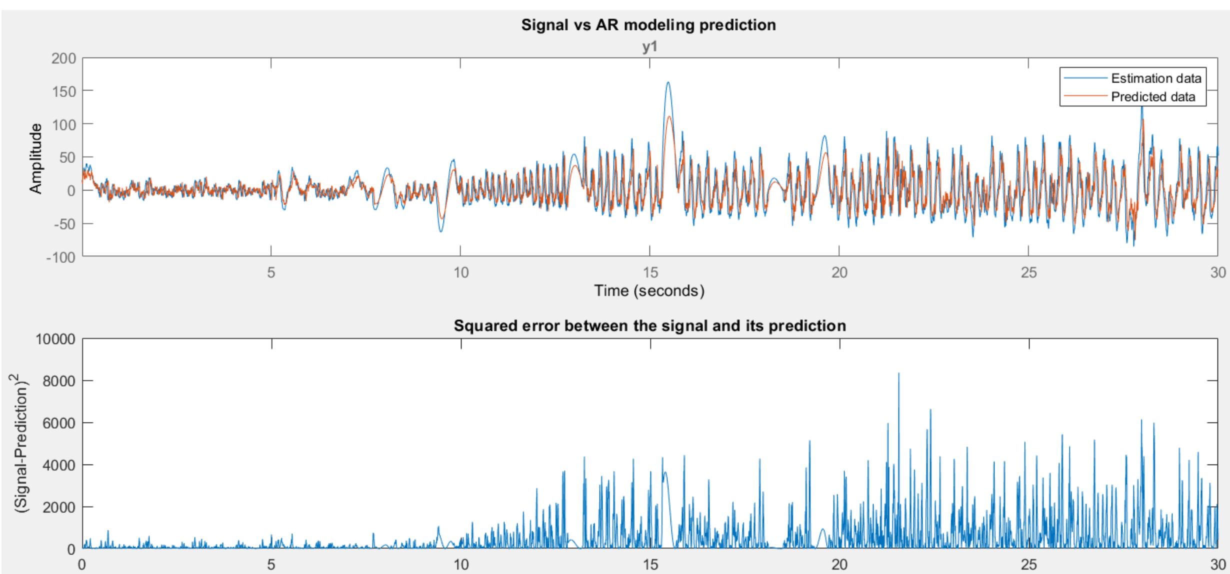

Figure 5: Error between the original signal and the predicted signal

4.5 Find where the seizure starts

To find where the seizure starts the mean squared error of segments of the error signal was found and was weighted against the largest error possible in the seizure-free section of the EEG signal. Figure 6 shows that the highest error in the normal signal is approximately 0.032. So we compare the rest of the signal against this number and it is found that the seizure starts at 10.504 seconds if we use a segment length of 0.5 seconds. By varying the segment length we find different answers for the given question, so the choice of segment length is crucial.

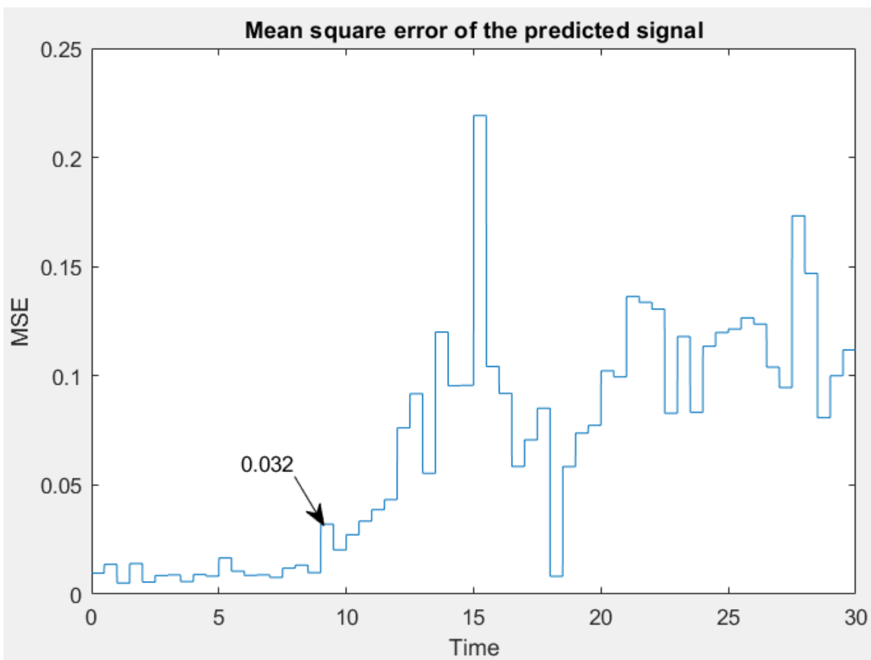

Figure 6: Error between the original signal and the predicted signal

5 Discussion

Another promising finding was that we can prove the signal is non-stationary by comparing the model parameters of different segments of the signal. Figure 7, 8 and 9 are the parameters generated using a fifth-order AR model. From this, it is clear the signal is non-stationary as the parameters vary across the signal.

```
sys =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 1.778 z^-1 + 1.084 z^-2 - 0.3427 z^-3 - 0.08928 z^-4 + 0.1378 z^-5
```

Figure 7: Parameters of the first segment

```
sys =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 2.155 z^-1 + 1.833 z^-2 - 0.9272 z^-3 + 0.1795 z^-4 + 0.08071 z^-5
```

Figure 8: Parameters of the second segment

```
sys =
Discrete-time AR model: A(z)y(t) = e(t)
  A(z) = 1 - 2.036 z^-1 + 1.6 z^-2 - 0.7413 z^-3 + 0.08107 z^-4 + 0.1167 z^-5
```

Figure 9: Parameters of the third segment

In this context of modelling EEG signals, there are two crucial aspects: the segment length used to model the signal and the model order. For a given AR(p) model, if we increase segment length the mean square error reduces but the fit accuracy also reduces, hence it is not viable to have a longer segment to model. As for the model order, values lower than 5 give lower accuracy, but higher than 5 only give a marginally higher accuracy, so for the sake of reducing complexity we maintain the order at 5.

Model order	**Segment length(sec)**	**Accuracy**	**MSE**
1	2	85.94%	3.555
1	4	82.95%	3.364
2	2	89.42%	2.015
2	4	86.78%	2.023
5	2	90.15%	1.745
5	4	87.78%	1.728
10	2	90.3%	1.693
10	4	87.82%	1.717
20	2	90.59%	1.591
20	4	88.08%	1.644

While looking at the origin of the seizure, the segment length to calculate the mean squared error also varies the result. If the segment length is too large there is a chance of missing the true start of the seizure as shown in Figure 10, where it states that the seizure starts at 11.004s. Also if the segment length is too small, there is a chance we are stuck in the errors in the normal signal as it cannot be modelled perfectly as shown in Figure 11 where it estimates the seizure starts at 9.304s which is actually an error due to the eyeblink.

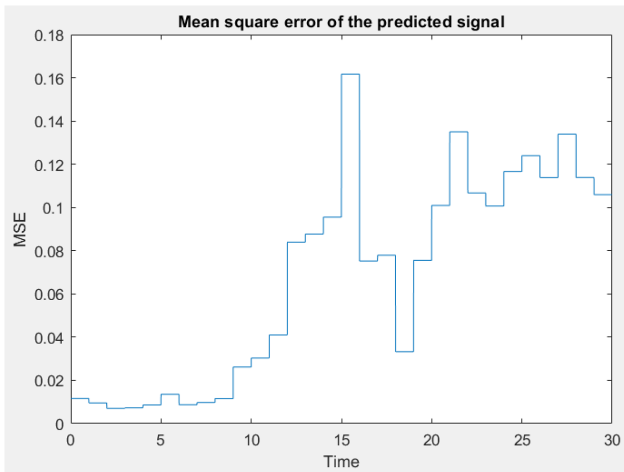

Figure 10: MSE when segment length is 1 second

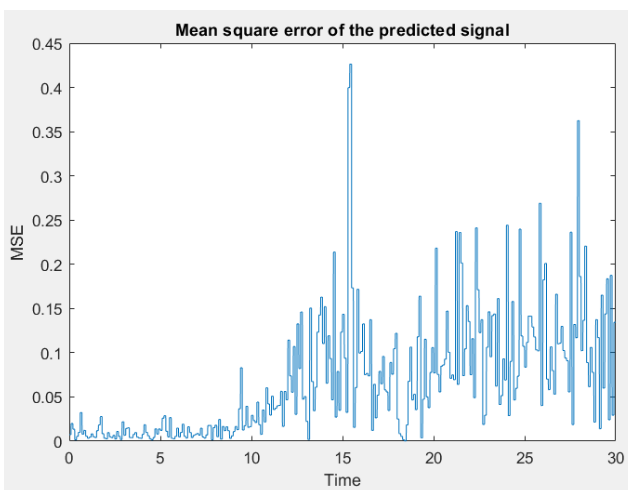

Figure 11: MSE when segment length is 0.3 second

6 Conclusion

The five given tasks have been successfully completed, and the results have been computed. These tasks encompassed data preprocessing to enhance the quality of EEG recordings by removing eye blinks, segmentation, feature extraction to capture relevant information, AR model application and thresholding to identify potential seizure events. The execution of these tasks forms a foundation for reliable seizure detection in EEG data using a signal modelling approach. These results provide essential insights for understanding, diagnosing, and managing epileptic conditions, potentially leading to improved patient care.

References

[1] Francisco Vaz, Pedro Guedes de Oliveira, and JoséC. Principe. "A study on the best order for autoregressive EEG modelling". In: *International Journal of Bio-Medical Computing* 20.1 (1987), pp. 41–50. ISSN: 0020-7101. DOI: `https://doi.org/10.1016/0020-7101(87)90013-4`. URL: `https://www.sciencedirect.com/science/article/pii/0020710187900134`.

[2] D Michael and J Houchin. "Automatic EEG analysis: A segmentation procedure based on the autocorrelation function". In: *Electroencephalography and Clinical Neurophysiology* 46.2 (1979), pp. 232–235. ISSN: 0013-4694. DOI: `https://doi.org/10.1016/0013-4694(79)90075-0`. URL: `https://www.sciencedirect.com/science/article/pii/0013469479900750`.