Assignment 1 (ML for TS) - MVA 2023/2024

Simon Queric simon.queric@telecom-paris.fr
Vincent Herfeld vincent.herfeld@telecom-paris.fr

November 7, 2023

1 Introduction

Objective. This assignment has three parts: questions about the convolutional dictionary learning, the spectral features and a data study using the DTW.

Warning and advice.

- Use code from the tutorials as well as from other sources. Do not code yourself well-known procedures (e.g. cross validation or k-means), use an existing implementation.
- The associated notebook contains some hints and several helper functions.
- Be concise. Answers are not expected to be longer than a few sentences (omitting calculations).

Instructions.

- Fill in your names and emails at the top of the document.
- Hand in your report (one per pair of students) by Tuesday 7$^{\text{th}}$ November 23:59 PM.
- Rename your report and notebook as follows:
`FirstnameLastname1_FirstnameLastname2.pdf` and
`FirstnameLastname1_FirstnameLastname2.ipynb`.
For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: docs.google.com/forms/d/e/1FAIpQLSdTwJEyc6QIoYTknjk12kJMtcKllFvPlWLk5LbyugW0YO7K6Q/viewform?usp=sf_link.

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 \quad + \quad \lambda \|\beta\|_1 \tag{1}$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ the design matrix, $\beta \in \mathbb{R}^p$ the vector of regressors and $\lambda > 0$ the smoothing parameter.

Show that there exists $\lambda_{\max}$ such that the minimizer of (1) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

We would like to show that we can find a λ_{max} that is too big and forces β to be $\mathbf{0}_p$. This means that for such $\lambda \geq \lambda_{\max}$ we have :

$$\forall \beta \in \mathbb{R}^p, \frac{1}{2}\|y\|_2^2 \leq \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

$\iff$

$$\langle y, X\beta\rangle - \frac{1}{2}\langle X\beta, X\beta\rangle \leqslant \lambda\|\beta\|_1$$

We can notice that we have the following inequalities :

$$\begin{cases} \|\beta\|_2 \leqslant \|\beta\|_1 \\ \langle y, X\beta\rangle - \frac{1}{2}\langle X\beta, X\beta\rangle \leqslant \langle y, X\beta\rangle \leqslant \|X^T y\|_2\|\beta\|_2 \text{ by Cauchy-Schwarz inequality} \end{cases}$$

Then, it's sufficient for our $\lambda_{\max}$ to satisfy :

$$\|X^T y\|_2\|\beta\|_2 \leqslant \lambda_{\max}\|\beta\|_2 \iff \|X^T y\|_2 \leqslant \lambda_{\max}$$

$$\lambda_{\max} = \|X^T y\|_2$$

Question 2

For a univariate signal $\mathbf{x} \in \mathbb{R}^n$ with n samples, the convolutional dictionary learning task amounts to solving the following optimization problem:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\|\mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k\right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{2}$$

where $\mathbf{d}_k \in \mathbb{R}^L$ are the K dictionary atoms (patterns), $\mathbf{z}_k \in \mathbb{R}^{N-L+1}$ are activations signals, and $\lambda > 0$ is the smoothing parameter.

Show that

- for a fixed dictionary, the sparse coding problem is a lasso regression (explicit the response vector and the design matrix);
- for a fixed dictionary, there exists $\lambda_{\max}$ (which depends on the dictionary) such that the sparse codes are only 0 for any $\lambda > \lambda_{\max}$.

Answer 2

Have in mind that the user chooses $L \leq n$ and K. We can consider the sparse coding problem as Lasso regression let's find the vector of regressors and the design matrix :

$$\forall k \in \{1, ..., K\},\ z_k * d_k = X_k z_k,$$

where : $X_k \in \mathbb{R}^{n\times(n-L+1)}$ and $X_k = \begin{bmatrix} d_k[1] & 0 & 0 & ... & 0 \\ d_k[2] & d_k[1] & 0 & . & 0 \\ . & d_k[2] & d_k[1] & \ddots & 0 \\ . & . & d_k[2] & \ddots & 0 \\ . & . & . & . & d_k[1] \\ . & . & . & . & . \\ d_k[L] & d_k[L-1] & d_k[L-2] & ... & d_k[2L-n-1] \\ 0 & d_k[L] & d_k[L-1] & ... & d_k[2L-n] \\ 0 & 0 & d_k[L] & ... & . \\ . & . & . & ... & . \\ 0 & 0 & 0 & ... & d_k[L] \end{bmatrix}$

And since we want to write this in the form $X\beta$ we just concatenate the different matrices X_k horizontally and vectors z_k vertically, to have $X \in \mathbb{R}^{n\times K(n-L+1)}$ and $\beta \in \mathbb{R}^{K(n-L+1)}$ and finally $X\beta \in \mathbb{R}^n$.

Now using the result of Question 1, we have that :

$$\lambda_{\max} = \left\| X^T x \right\|_2$$

3 Spectral feature

Let X_n $(n = 0, \dots, N-1)$ be a weakly stationary random process with zero mean and autocovariance function $\gamma(\tau) := \mathbb{E}(X_n X_{n+\tau})$. Assume the autocovariances are absolutely summable, i.e. $\sum_{\tau\in\mathbb{Z}} |\gamma(\tau)| < \infty$, and square summable, i.e. $\sum_{\tau\in\mathbb{Z}} \gamma^2(\tau) < \infty$. Denote by f_s the sampling frequency, meaning that the index n corresponds to the time instant n/f_s and for simplicity, let N be even.

The *power spectrum* S of the stationary random process X is defined as the Fourier transform of the autocovariance function:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f\tau/f_s}. \tag{3}$$

The power spectrum describes the distribution of power in the frequency space. Intuitively, large values of $S(f)$ indicates that the signal contains a sine wave at the frequency f. There are many estimation procedures to determine this important quantity, which can then be used in a machine learning pipeline. In the following, we discuss about the large sample properties of simple estimation procedures, and the relationship between the power spectrum and the autocorrelation.

(Hint: use the many results on quadratic forms of Gaussian random variables to limit the amount of calculations.)

Question 3

In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise.

- Calculate the associated autocovariance function and power spectrum. (By analogy with the light, this process is called "white" because of the particular form of its power spectrum.)

Answer 3

A Gaussian white noise is an iid sequence of 0 mean Gaussian variables with variance σ^2. Let's calculate the autocovariance function using the independance :

$$\gamma(\tau) = \mathbb{E}[X_n X_{n+\tau}] = \mathbb{E}[X_n]\mathbb{E}[X_{n+\tau}] = 0, \forall \tau \neq 0$$
$$\gamma(0) = \sigma^2$$

Now we can calculate the Fourier Transform of the autocovariance function :

$$S(f) = \sum_{\tau=-\infty}^{\tau=+\infty} \gamma(\tau) e^{-2\pi f \frac{\tau}{f_s}} = \sigma^2 e^{-2\pi f \frac{0}{f_s}} = \sigma^2$$

We can now see that the power spectrum of a Gaussian white noise is constant and thus the process contains all frequencies at intensity σ^2, in the same way white light contains all frequencies of the visible domain.

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{4}$$

for $\tau = 0, 1, \ldots, N-1$ and $\hat{\gamma}(\tau) := \hat{\gamma}(-\tau)$ for $\tau = -(N-1), \ldots, -1$.

- Show that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$ but asymptotically unbiased. What would be a simple way to de-bias this estimator?

Answer 4

Let's calculate the expectation of the estimator $\hat{\gamma}(\tau)$:

$$\mathbb{E}[\hat{\gamma}(\tau)] = \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right] = \frac{1}{N}\sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}] = \frac{1}{N}\sum_{n=0}^{N-\tau-1} \gamma(\tau) = \frac{N-\tau}{N}\gamma(\tau)$$

The bias is :

$$b(\hat{\gamma}(\tau), \gamma(\tau)) = \mathbb{E}[\hat{\gamma}(\tau) - \gamma(\tau)] = \frac{-\tau}{N}\gamma(\tau) \neq 0$$

As we can see the estimator is biased but when $N \to +\infty$ the bias goes to 0. A simple way to de-bias this estimator is to add a multiplicative coefficient such that $\mathbb{E}[\tilde{\gamma}(\tau)] = \gamma(\tau)$:

$$\tilde{\gamma}(\tau) = \frac{N}{N-\tau}\hat{\gamma}(\tau) = \frac{1}{N-\tau}\sum_{n=0}^{N-\tau} X_n X_{n+\tau}$$

Question 5

Define the discrete Fourier transform of the random process $\{X_n\}_n$ by

$$J(f) := (1/\sqrt{N})\sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{5}$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$ where $f_k = f_s k/N$. (They can be efficiently computed using the Fast Fourier Transform.)

- Write $|J(f_k)|^2$ as a function of the sample autocovariances.
- For a frequency f, define $f^{(N)}$ the closest Fourier frequency f_k to f. Show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Answer 5

$$\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\sum_{0\leqslant n,m\leqslant N-1} X_n X_m e^{-\frac{2i\pi f_s k(n-m)}{f_s N}} \\
&= \frac{1}{N}\sum_{m=0}^{N-1}\left(\sum_{\tau=0}^{N-1-m} X_m X_{m+\tau} e^{-\frac{2i\pi k\tau}{N}} + \sum_{\tau=-m}^{-1} X_m X_{m+\tau} e^{-\frac{2i\pi k\tau}{N}}\right) \\
&= \frac{1}{N}\sum_{\tau=0}^{N-1}\sum_{m=0}^{N-1-\tau} X_m X_{m+\tau} e^{-\frac{2i\pi k\tau}{N}} + \frac{1}{N}\sum_{\tau=-N+1}^{-1}\sum_{m=-\tau}^{N-1} X_m X_{m+\tau} e^{-\frac{2i\pi k\tau}{N}} \\
&= \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-\frac{2i\pi k\tau}{N}} \\
&= \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-\frac{2i\pi f_k\tau}{f_s}}
\end{aligned}$$

We recall that this is a real number thanks to the parity of $\hat{\gamma}$.

Finally, $|J(f_k)|^2$ is the discrete Fourier transform of the sample autocovariance $\hat{\gamma}$ evaluated in f_k. By taking the expectation and the limit (using the result of question 4 and an argument of density of $\mathbb{Q}$ in $\mathbb{R}$) we have :

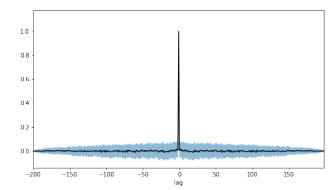
$$\sum_{\tau=-N+1}^{N-1} \mathbb{E}[\hat{\gamma}(\tau)] e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} \xrightarrow{N\to+\infty} \sum_{n=-\infty}^{+\infty} \gamma(\tau) e^{-\frac{2i\pi f\tau}{f_s}} = S(f) \tag{6}$$

Question 6

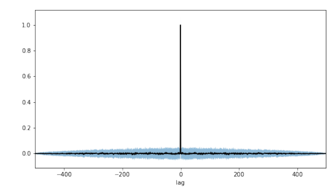
In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs τ) for 100 simulations of X. Plot the average value as well as the average $\pm$ the standard deviation. What do you observe?
- For $N \in \{200, 500, 1000\}$, compute the *periodogram* ($|J(f_k)|^2$ vs f_k) for 100 simulations of X. Plot the average value as well as the average $\pm$ the standard deviation. What do you observe?

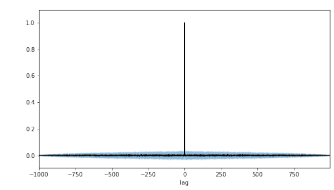
Add your plots to Figure 1.



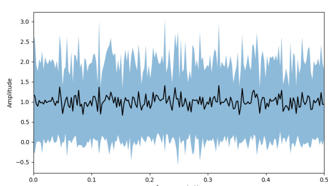
Autocovariance ($N = 200$)



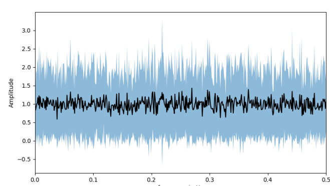
Autocovariance ($N = 500$)



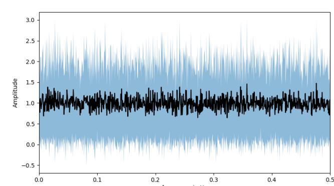
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

Answer 6

As we can see the longer the sequence the more points we have and the more our autocovariance estimator converges with precision to the real values the autocorrelation. This is the case for the autocovariance but in fact we notice that there seems to be even more noise / variance (or at least it doesn't decrease) in the periodogram when we add more points. At least we see that it is unbiased since it's mean is very close to 1 = σ^2.

Question 7

We want to show that the estimator $\hat{\gamma}(\tau)$ is consistent, i.e. it converges in probability when the number N of samples grows to ∞ to the true value $\gamma(\tau)$. In this question, assume that X is a wide-sense stationary *Gaussian* process.

- Show that for $\tau > 0$

$$\mathrm{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{7}$$

(Hint: if $\{Y_1, Y_2, Y_3, Y_4\}$ are four centered jointly Gaussian variables, then $\mathbb{E}[Y_1Y_2Y_3Y_4] = \mathbb{E}[Y_1Y_2]\mathbb{E}[Y_3Y_4] + \mathbb{E}[Y_1Y_3]\mathbb{E}[Y_2Y_4] + \mathbb{E}[Y_1Y_4]\mathbb{E}[Y_2Y_3]$.)

- Conclude that $\hat{\gamma}(\tau)$ is consistent.

Answer 7

We have $\mathrm{var}(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)^2] - \mathbb{E}[\hat{\gamma}(\tau)]^2$

We compute $\mathbb{E}[\hat{\gamma}(\tau)^2]$ first :

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\tau)^2] &= \mathbb{E}\left[\frac{1}{N^2} \sum_{0\leqslant n,m \leqslant N-\tau-1} X_n X_{n+\tau} X_m X_{m+\tau}\right] \\
&= \frac{1}{N^2} \sum_{0\leqslant n,m \leqslant N-\tau-1} \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}] + \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_m X_{m+\tau}] \\
&= \frac{1}{N^2} \sum_{0\leqslant n,m \leqslant N-\tau-1} \gamma(\tau)^2 + \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau) \\
&= \frac{(N-\tau)^2}{N^2}\gamma(\tau)^2 + \frac{1}{N^2} \sum_{0\leqslant n,m \leqslant N-\tau-1} \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau) \\
&= \mathbb{E}[\gamma(\tau)]^2 + \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|k|}{N}\right)\left(\gamma(k)^2 + \gamma(n+k)\gamma(n-k)\right)
\end{aligned}$$

using the change of variable $k = n - m$

Finally,

$$\mathrm{var}(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|k|}{N}\right)\left(\gamma(k)^2 + \gamma(n+k)\gamma(n-k)\right)$$

To show that $\hat{\gamma}$ is consistent, we need to prove that $\mathrm{var}(\hat{\gamma}(\tau))$ goes to 0 when N goes to $+\infty$.

We have

$$\begin{aligned}
|\mathrm{var}(\hat{\gamma}(\tau))| &\leqslant \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left|1 - \frac{\tau+|k|}{N}\right| \left[\gamma(k)^2 + |\gamma(n+k)\gamma(n-k)|\right] \\
&\leqslant \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left|\gamma(k)^2 + \gamma(n+k)\gamma(n-k)\right| \\
&\leqslant \frac{2}{N} \sum_{k\in\mathbb{Z}} \gamma(k)^2 \xrightarrow[N\to+\infty]{} 0
\end{aligned}$$

Now, using Markov's inequality, we have :

$$\begin{aligned}\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| \geqslant \varepsilon) &\leqslant \frac{\mathbb{E}[|\hat{\gamma}(\tau) - \gamma(\tau)|^2]}{\varepsilon^2} \\ &\leqslant \frac{\text{var}(\hat{\gamma}(\tau)) + \frac{\tau^2}{N^2}\gamma(\tau)^2}{\varepsilon^2} \xrightarrow[N\to+\infty]{} 0\end{aligned}$$

It concludes the proof.

Contrary to the correlogram, the periodogram is not consistent. It is one of the most well-known estimators that is asymptotically unbiased but not consistent. In the following question, this is proven for a Gaussian white noise but this holds for more general stationary processes.

Question 8

Assume that X is a Gaussian white noise (variance σ^2) and let $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s)$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n/f_s)$. Observe that $J(f) = (1/N)(A(f) + iB(f))$.

- Derive the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, \ldots, f_{N/2}$ where $f_k = f_s k/N$.
- What is the distribution of the periodogram values $|J(f_0)|^2, |J(f_1)|^2, \ldots, |J(f_{N/2})|^2$.
- What is the variance of the $|J(f_k)|^2$? Conclude that the periodogram is not consistent.
- Explain the erratic behavior of the periodogram in Question 6 by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

Recall that X being a Gaussian white noise it is a vector of iid Gaussian variables and thus constructs a Gaussian vector. Using the definition of a Gaussian vector we have that A(f) and B(f) are Gaussian variables since they are both linear combinations of the terms of X. Let's calculate their mean and variance and determine the distribution of the periodogram values.

$$\begin{aligned}\mathbb{E}[A(f)] &= \mathbb{E}\left[\sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s)\right] \\ &= \sum_{n=0}^{N-1} \mathbb{E}[X_n] \cos(-2\pi f n/f_s) = 0 = \sum_{n=0}^{N-1} \mathbb{E}[X_n] \sin(-2\pi f n/f_s) = \mathbb{E}[B(f)]\end{aligned}$$

By independance of the variables of X we have that their covariances are null :

$$\begin{aligned}Var(A(f)) &= Var\left(\sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s)\right) = \sum_{n=0}^{N-1} Var(X_n \cos(-2\pi f n/f_s)) \\ &= \sum_{n=0}^{N-1} \sigma^2 \cos^2(-2\pi f n/f_s)\end{aligned}$$

And thus,

$$Var(B(f)) = \sum_{n=0}^{N-1} \sigma^2 \sin^2(-2\pi f n/f_s)$$

Also, by noticing a geometrical sum we have :

$$Var(A(f_k)) - Var(B(f_k)) = \sigma^2 \sum_{n=0}^{N-1} cos\left(-4\pi \frac{kn}{N}\right) = \begin{cases} \sigma^2 N \text{ if } k \in \{0, \frac{N}{2}\} \\ 0 \text{ else} \end{cases}$$

So $A(f_k)$ and $B(f_k)$ have same variance (for the right k).

Let's show that $A(f_k)$ and $B(f_k)$ are independent. Notice that $(A(f_k), B(f_k))$ is a Gaussian vector since it's a linear combination of $(X_n)_{0 \leqslant n \leqslant N-1}$. Then, it's sufficient to prove that $\mathbb{E}(A(f_k)B(f_k)) = 0$.

$$\begin{aligned} \mathbb{E}[A(f_k)B(f_k)] &= \mathbb{E}\left[\sum_{0 \leqslant n,m \leqslant N-1} X_n X_m \cos(-2\pi nk/N) \sin(-2\pi mk/N)\right] \\ &= \sum_{n=0}^{N-1} \mathbb{E}[X_n^2] \sin(-2\pi kn/N) \\ &= \sum_{n=0}^{N-1} \sigma^2 \sin(-2\pi kn/N) \\ &= 0 \end{aligned}$$

$|J(f_k)|^2 = \left(\frac{A(f_k)}{\sqrt{N}}\right)^2 + \left(\frac{B(f_k)}{\sqrt{N}}\right)^2$ which is the squared sum of two independent Gaussian variables with same variance (for $k \neq 0$ and $k \neq N/2$) and mean 0, and thus $|J(f_k)|^2$ follows an exponential distribution of parameter $1/\sigma^2$. For $k \in \{0, N/2\}$ it follows a $chi2$ distribution with one degree of freedom.

Hence the variance of $|J(f_k)|^2$ is σ^4.

Clearly, $|J(f^{(N)})|^2$ is not consistent because $\mathbb{P}(|J(f^{(N)})|^2 - \sigma^2 \geqslant \varepsilon)$ is non-negative and constant for any $\varepsilon > 0$ and $N > 1$.

We want to explain the erratic behavior by the independance of the $J(f_k)$ variables. Notice that $(J(f_k), J(f_l))$ is a Gaussian vector since it's a linear combination of $(X_n)_{0 \leqslant n \leqslant N-1}$. Then, it's sufficient to prove that $\mathbb{E}(J(f_k)J(f_l)) = 0$ for $k \neq l$.

$$
\begin{aligned}
\mathbb{E}[J(f_k)J(f_l)] &= \mathbb{E}\left[\frac{1}{N}\sum_{n,m} X_n X_m e^{-\frac{2i\pi(nk+ml)}{N}}\right] \\
&= \frac{1}{N}\sum_{n,m} \mathbb{E}[X_n X_m] e^{-\frac{2i\pi(nk+ml)}{N}} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}[X_n^2] e^{-\frac{2i\pi n(k+l)}{N}} \\
&= \frac{\sigma^2}{N}\frac{e^{-\frac{2i\pi N(k+l)}{N}} - 1}{e^{-\frac{2i\pi(k+l)}{N}} - 1} \\
&= 0
\end{aligned}
$$

The last sum is a geometric sum with common ratio $e^{-\frac{2i\pi(k+l)}{N}}$ which is different of 1 because $k \neq l$ and $k, l \in \{0, \cdots, \frac{N}{2}\}$.

It shows that $J(f_k)$ and $J(f_l)$ are independant. Hence $|J(f_k)|^2$ and $|J(f_l)|^2$ are independant and their covariance is null. It explains the erratic behavior of the sequence $|J(f_k)|^2$ computed in Question 6.

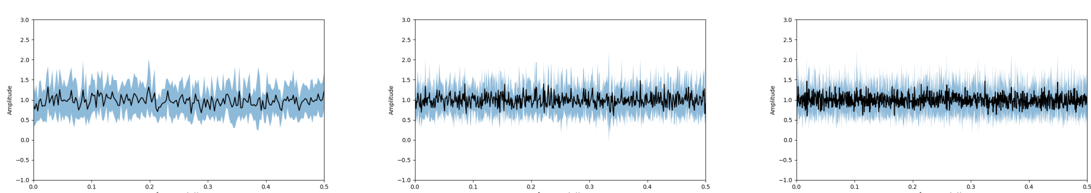
Question 9

As seen in the previous question, the problem with the periodogram is the fact that its variance does not decrease with the sample size. A simple procedure to obtain a consistent estimate is to divide the signal in K sections of equal durations, compute a periodogram on each section and average them. Provided the sections are independent, this has the effect of dividing the variance by K. This procedure is known as Bartlett's procedure.

- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe.

Add your plots to Figure 2.

Answer 9



Periodogram ($N = 200$) Periodogram ($N = 500$) Periodogram ($N = 1000$)

Figure 2: Barlett's periodograms of a Gaussian white noise (see Question 9).

Looking at the standard derivation (the blue part of the plots) we notice that it is a lot closer to the mean than in the standard Periodogram, this shows that we have reduced the variance of the estimator. We have also checked numerically. This can be explained because we have taken the mean of several periodograms and this has an effect of smoothing (variance reduction).

4 Data study

4.1 General information

Context. The study of human gait is a central problem in medical research with far-reaching consequences in the public health domain. This complex mechanism can be altered by a wide range of pathologies (such as Parkinson's disease, arthritis, stroke,...), often resulting in a significant loss of autonomy and an increased risk of fall. Understanding the influence of such medical disorders on a subject's gait would greatly facilitate early detection and prevention of those possibly harmful situations. To address these issues, clinical and bio-mechanical researchers have worked to objectively quantify gait characteristics.

Among the gait features that have proved their relevance in a medical context, several are linked to the notion of step (step duration, variation in step length, etc.), which can be seen as the core atom of the locomotion process. Many algorithms have therefore been developed to automatically (or semi-automatically) detect gait events (such as heel-strikes, heel-off, etc.) from accelerometer and gyrometer signals.

Data. Data are described in the associated notebook.

4.2 Step classification with the dynamic time warping (DTW) distance

Task. The objective is to classify footsteps then walk signals between healthy and non-healthy.

Performance metric. The performance of this binary classification task is measured by the F-score.

Question 10

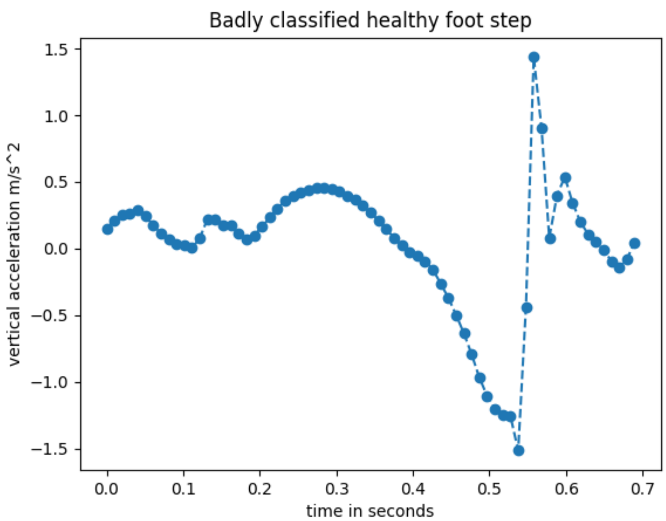
Combine the DTW and a k-neighbors classifier to classify each step. Find the optimal number of neighbors with 5-fold cross-validation and report the optimal number of neighbors and the associated F-score. Comment briefly.

Answer 10

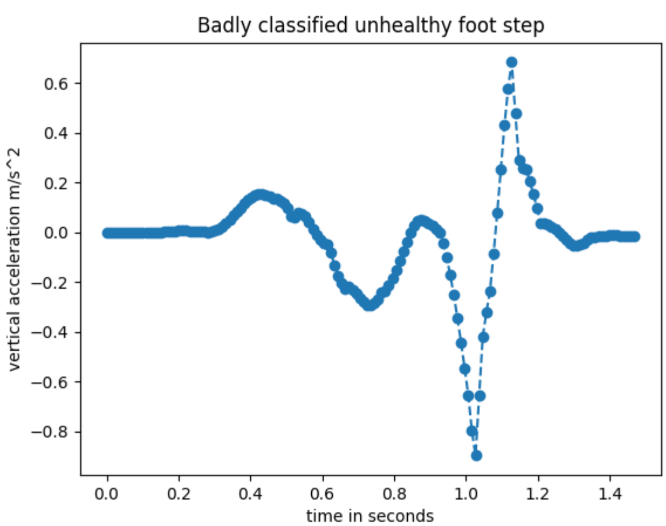
Question 11

Display on Figure 3 a badly classified step from each class (healthy/non-healthy).

Answer 11



Badly classified healthy step



Badly classified non-healthy step

Figure 3: Examples of badly classified steps (see Question 11).

By looking at the signals, we can notice that healthy foot steps have a smooth vertical acceleration, while unhealthy foot steps present discontinuities. Above, the two signals are quite smooth and have also some discontinuities. It could explain the misclassifications.