Assignment 1 (ML for TS) - MVA

Joanna Peloso pelosojoanna@gmail.com
Thomas Gravier thomas.gravier@gadz.org

October 28, 2024

1 Introduction

Objective. This assignment has three parts: questions about convolutional dictionary learning, 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 28$^{\text{th}}$ October 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: 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

Minimiser l'objectif :

$$L(\beta) = \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

où $\lambda > 0$ est un paramètre de régularisation.

Pour que $\beta = 0_p$ soit un minimiseur, il faut que :

$$0 \in -X^T y + \lambda[-1, 1]^p$$

(cf optimisation d'une sous différentielle) ce qui implique que $|X^T y_j| \leq \lambda$ pour chaque composante $j = 1, 2, \ldots, p$.

Le plus petit λ qui satisfait $|X^T y_j| \leq \lambda$ est :

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

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

Pour un dictionnaire fixé, on définit

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{pmatrix} \in \mathbb{R}^{K(N-L+1)}$$

comme la concaténation des activations z_k, et

$$D \in \mathbb{R}^{N \times K(N-L+1)}$$

comme une matrice bloc-diagonale contenant les matrices de convolution de Toeplitz $T(d_k)$ associées aux atomes d_k. Ainsi :

$$D = \begin{pmatrix} T(d_1) & 0 & \dots & 0 \\ 0 & T(d_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T(d_K) \end{pmatrix}$$

où chaque $T(d_k)$ est une matrice de Toeplitz qui applique la convolution du vecteur z_k avec l'atome d_k.

Un exemple de matrice de Toeplitz associée à l'atome

$$d_1 = \begin{pmatrix} d_1(1) \\ d_1(2) \\ \vdots \\ d_1(L) \end{pmatrix}$$

est :

$$T(d_1) = \begin{pmatrix} d_1(1) & 0 & \dots & 0 \\ d_1(2) & d_1(1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_1(L) & d_1(L-1) & \dots & d_1(1) \\ 0 & d_1(L) & \dots & d_1(2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_1(L) \end{pmatrix}$$

Cette matrice $T(d_1)$ réalise la convolution $d_1 * z_1$ lorsqu'elle est multipliée par le vecteur d'activation z_1.

En utilisant cette matrice de convolution, le problème peut être reformulé comme un problème de Lasso :

$$\min_Z \|x - DZ\|_2^2 + \lambda \|Z\|_1$$

où le paramètre de régularisation $\lambda_{\max}$ est donné par :

$$\lambda_{\max} = \|D^T x\|_\infty.$$

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 the sampling frequency by f_s, meaning that the index n corresponds to the time n/f_s. 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 $S(f)$ indicate 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 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 number of calculations.)

Question 3

In this question, let X_n $(n = 0, \dots, N-1)$ 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

Let X_n be a Gaussian white noise, with mean $\mu = 0$ and standard deviation $\sigma > 0$. Each X_i is a sample of this distribution, i.e., they are independent. Then, for τ:

$$\gamma(\tau) = \mathbb{E}[X_n X_{n+\tau}] = \operatorname{cov}(X_n, X_{n+\tau}) + \mathbb{E}[X_n]\mathbb{E}[X_{n+\tau}]$$

i.e.,

$$\gamma(\tau) = \begin{cases} 0 & \text{if } \tau \neq 0, \\ \sigma^2 & \text{otherwise.} \end{cases}$$

For the power spectrum:

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

$$= \gamma(0) e^0 + \sum_{\tau=-\infty}^{-1} \gamma(\tau) e^{-2\pi i f \tau / f_s} + \sum_{\tau=1}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau / f_s}$$

$$= \gamma(0)$$

$$S(f) = \sigma^2$$

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

Solution

We begin by calculating the expected value of the sample autocovariance:

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

$$= \frac{1}{N} \sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}]$$

$$= \frac{1}{N} \sum_{n=0}^{N-\tau-1} \gamma(\tau) \quad \text{because the process } X \text{ is weakly stationary}$$

$$= \frac{N-\tau}{N} \gamma(\tau)$$

Thus, we have:

$$\mathbb{E}[\hat{\gamma}_N(\tau)] \neq \gamma(\tau)$$

This shows that the estimator is biased. However, the expression above indicates that:

$$\lim_{N \to \infty} \mathbb{E}[\hat{\gamma}_N(\tau)] = \gamma(\tau)$$

To correct this bias, we can multiply the estimator by $\frac{N}{N-\tau}$. Therefore, the unbiased estimator is given by:

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

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 $|J(f_0)|^2, |J(f_1)|^2, \ldots, |J(f_{N/2})|^2$ 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

Let k be in $\{1, \ldots, N/2\}$.

$$\begin{aligned}
|J(f_k)|^2 &= J(f_k)\overline{J(f_k)} \\
&= \left(\frac{1}{\sqrt{N}}\right)\left(\frac{1}{\sqrt{N}}\right)\sum_{n=0}^{N-1} X_n \exp\left\{\frac{-2i\pi f_k n}{f_s}\right\}\sum_{n=0}^{N-1} X_n \exp\left\{\frac{2i\pi f_k n}{f_s}\right\} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{l=0}^{N-1} X_n X_l \exp\left\{\frac{-2i\pi f_k (n-l)}{f_s}\right\} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\left(\sum_{\tau=-n}^{N-1-n} X_n X_{n+\tau} \exp\left\{\frac{-2i\pi f_k \tau}{f_s}\right\}\right) \qquad (\tau = l - n) \\
&= \sum_{\tau=-(N-1)}^{N-1}\left(\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \exp\left\{\frac{-2i\pi f_k \tau}{f_s}\right\}\right),
\end{aligned}$$

where we observe that the change of index to τ with a fixed value corresponds to summing across all diagonals of the matrix. Recognizing the expression of the sample autocovariance, we then get:

$$\boxed{\forall k \in \{1, \ldots, N/2\},\ |J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) \exp\left\{\frac{-2i\pi f_k \tau}{f_s}\right\}}.$$

For the second part of the question, let $|J(f^{(N)})|^2$ be an estimator of $S(f)$, where $f^{(N)} \in \arg\min_{f_k} |f - f_k|$. Let's verify that it's asymptotically unbiased, with $f > 0$.

- $|f^{(N)} - f| \le |f^{(N)}| + |f| \le \left|\frac{f_s}{2N}\right| + \left|\frac{f_s}{2N}\right| = \frac{|f_s|}{N} \le \frac{|f_0|}{N}$, so $f^{(N)} \xrightarrow[N\to+\infty]{} f$.
- Thus, considering a fixed τ,

  $$\mathbb{E}\left[\hat{\gamma}(\tau)\exp\left\{\frac{-2i\pi f^{(N)}\tau}{f_s}\right\}\right] = \mathbb{E}\left[\hat{\gamma}(\tau)\right]\exp\left\{\frac{-2i\pi f^{(N)}\tau}{f_s}\right\} \xrightarrow[N\to+\infty]{} \gamma(\tau)\exp\left\{\frac{-2i\pi f\tau}{f_s}\right\},$$

  because $\hat{\gamma}(\tau)$ is asymptotically unbiased. Additionally, by definition, we know that $\gamma(\tau)$ is summable ($\sum_{\tau=-\infty}^{+\infty} |\gamma(\tau)| \ge +\infty$).
- Thus, by the dominated convergence theorem, we have

  $$\begin{aligned}
  \lim_{N\to+\infty}\sum_{\tau=1-N}^{N-1}\mathbb{E}\left[\hat{\gamma}(\tau)\right]\exp\left\{\frac{-2i\pi f^{(N)}\tau}{f_s}\right\} &= \sum_{\tau=-\infty}^{+\infty}\lim_{N\to+\infty}\mathbb{E}\left[\hat{\gamma}(\tau)\right]\exp\left\{\frac{-2i\pi f^{(N)}\tau}{f_s}\right\} \\
  &= \sum_{\tau=-\infty}^{+\infty}\gamma(\tau)\exp\left\{\frac{-2i\pi f\tau}{f_s}\right\} \\
  &= S(f).
  \end{aligned}$$

Thus, we have $\boxed{\lim_{N\to+\infty} |J(f^{(N)})|^2 = S(f)}$.

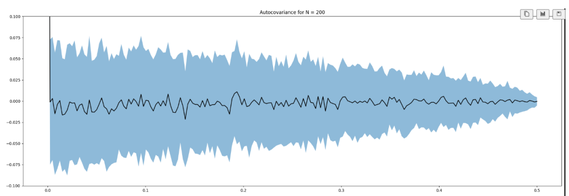
Question 6

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

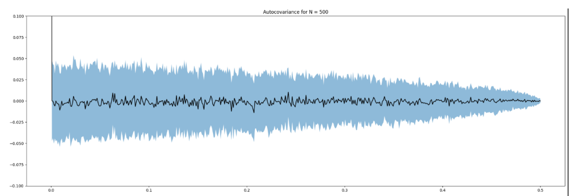
- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs τ) for 100 simulations of X. Plot the average value as well as the average $\pm$, the standard deviation. What do you observe?
- For $N \in \{200, 500, 1000\}$, compute the *periodogram* ($|J(f_k)|^2$ vs f_k)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

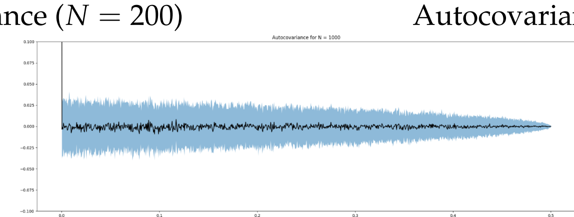
Answer 6



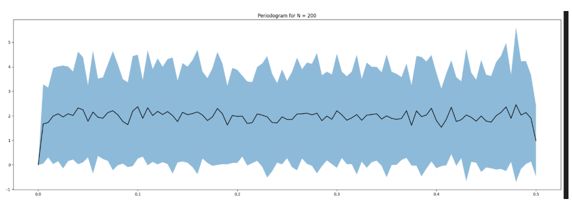
Autocovariance ($N = 200$)



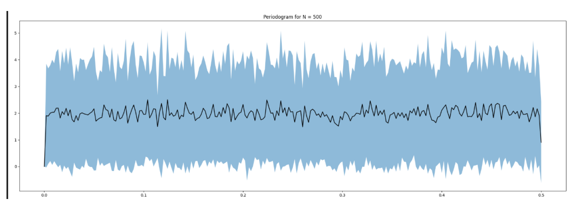
Autocovariance ($N = 500$)



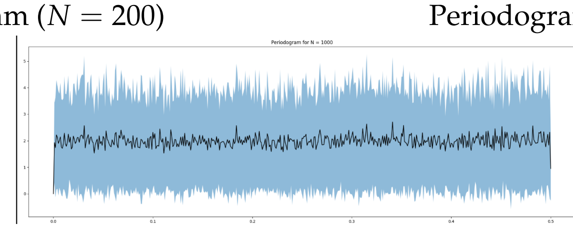
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Observations for autocovariance graphs : we observe that the standard variation is decreasing when the number N of samples grows. The mean is near to 0 for $\tau \neq 0$ and egal to 1 for $\tau = 0$.

Observations for periodogram graphs : the standard variation doesn't decreased when the number N of samples grows.

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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{6}$$

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

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

Answer 7

As we have proofed previously that $\hat{\gamma}(\tau)$ was asymptotically unbiaised, we just need to show $\mathbb{V}(\hat{\gamma}(\tau)) \xrightarrow[N \to +\infty]{} 0$ so that $\hat{\gamma}(\tau)$ can be consistent.
Let's derive the variance of this estimator:

$$\begin{aligned}
\mathbb{V}(\hat{\gamma}(\tau)) &= \mathbb{E}\left[\hat{\gamma}(\tau)^2\right] - \mathbb{E}\left[\hat{\gamma}(\tau)\right]^2 \\
&= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right)^2\right] - \left(\frac{1}{N}\sum_{n=0}^{N-\tau-1}\gamma(\tau)\right)^2 \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{l=0}^{N-\tau-1} \mathbb{E}\left[X_n X_{n+\tau} X_l X_{l+\tau}\right] - \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\gamma(\tau)\right)^2 \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{l=0}^{N-\tau-1} \left\{\mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_l X_{l+\tau}] + \mathbb{E}[X_n X_l]\mathbb{E}[X_{n+\tau} X_{l+\tau}] + \mathbb{E}[X_n X_{l+\tau}]\mathbb{E}[X_{n+\tau} X_l]\right\} \\
&\quad - \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\gamma(\tau)\right)^2 \text{ (Hint of Question 7)} \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{l=0}^{N-\tau-1} \gamma(\tau)^2 + \gamma(l-n)^2 + \gamma(l+\tau-n)\gamma(l-n-\tau) - \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\gamma(\tau)\right)^2 \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=-n}^{N-\tau-1-n} \gamma(\tau)^2 + \gamma(m)^2 + \gamma(m+\tau)\gamma(m-\tau) - \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\gamma(\tau)\right)^2 \quad (m = l-n) \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=-n}^{N-\tau-1-n} \gamma(m)^2 + \gamma(m+\tau)\gamma(m-\tau)
\end{aligned}$$

or, $\forall a \in \mathbb{R}$, $\frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=-n}^{N-\tau-1-n} a = \sum_{m=1+\tau-N}^{N-\tau-1}(N-\tau+|m|)a$ so we have

$$\mathbb{V}(\hat{\gamma}(\tau)) = \frac{1}{N^2}\sum_{m=1+\tau-N}^{N-\tau-1}(N-\tau+|m|)\gamma(m)^2 + \gamma(m+\tau)\gamma(m-\tau).$$

Thus we obtain $\boxed{\mathbb{V}(\hat{\gamma}(\tau)) = \frac{1}{N}\sum_{m=1+\tau-N}^{N-\tau-1}\left(1-\frac{\tau+|m|}{N}\right)\left[\gamma(m)^2+\gamma(m+\tau)\gamma(m-\tau)\right]}$.

For the second part of the question, let τ be fixed. We can first observe that $\forall m \in \{1+\tau-N,\dots,N-\tau-1\}$, $\left(1-\frac{\tau+|m|}{N}\right) \leq 1$. Thus:

$$\begin{aligned}\mathbb{V}(\hat{\gamma}(\tau)) &\leq \frac{1}{N}\sum_{m=1+\tau-N}^{N-\tau-1}\left[\gamma(m)^2+\gamma(m+\tau)\gamma(m-\tau)\right]\\ &= \frac{1}{N}\sum_{m=1+\tau-N}^{N-\tau-1}\gamma(m)^2 + \frac{1}{N}\sum_{m=1+\tau-N}^{N-\tau-1}\gamma(m+\tau)\gamma(m-\tau).\end{aligned}$$

Additionaly, $\gamma^2(\tau)$ is integrable so $\lim_{N\to+\infty}\sum_{m=1+\tau-N}^{N-\tau-1}\gamma^2(\tau) < +\infty$. We also have $\sum_{m=1+\tau-N}^{N-\tau-1}\gamma(m+\tau)\gamma(m-\tau) \leq \left(\sum_{m=-N}^{N}\gamma(m)\right)^2 < +\infty$.

We can thus major $N\mathbb{V}(\hat{\gamma}(\tau))$ by a constant C. Finally, we obtain $\boxed{\mathbb{V}(\hat{\gamma}(\tau)) \leq \frac{C}{N} \xrightarrow[N\to+\infty]{} 0}$.

The estimator $\hat{\gamma}(\tau)$ is consistent.

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 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 $f = f_0, f_1, \dots, f_{N/2}$ where $f_k = f_s k/N$.
- What is the distribution of the periodogram values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$.
- 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

First, we can observe easily that $J(f) = (1/N)(A(f)+iB(f))$.
Let's start to derive the mean of $A(f_k)$ and $B(f_k)$, $\forall k \in \{1,\dots,N/2\}$;

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

because $\forall n \in \{0,\dots,N-1\}, \mathbb{E}[X_n] = 0$. Same for $\mathbb{E}[B(f_k)]$.

Now we need to derive the variance. We compute the variance of $A(f_k)$, and the computing steps will be the same for $B(f_k)$.

$$\begin{aligned}
\mathbb{V}(A(f_k)) &= \mathbb{V}\left(\sum_{n=0}^{N-1} X_n \cos(2\pi f_k n/f_s)\right) \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} X_n \cos(2\pi f_k n/f_s)\right)^2\right] - \mathbb{E}\left[\sum_{n=0}^{N-1} X_n \cos(2\pi f_k n/f_s)\right]^2 \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} X_n \cos(2\pi kn/N)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos^2(2\pi kn/N)\right] \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left[X_n X_m\right] \cos(2\pi kn/N)\cos(2\pi km/N) \\
&= \sum_{n=0}^{N-1} \mathbb{E}\left[X_n^2\right] \cos^2(2\pi kn/N) \quad (if\ m \neq n,\ \mathbb{E}[X_n X_m] = 0) \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos^2(2\pi kn/N)
\end{aligned}$$

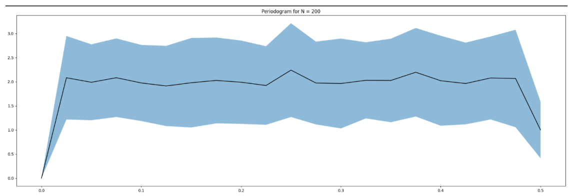
Question 9

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 into 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.

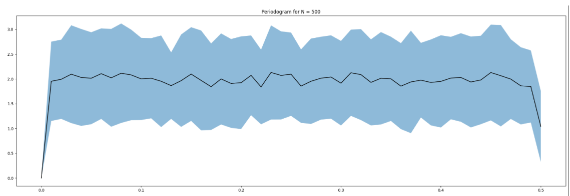
- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe?

Add your plots to Figure 2.

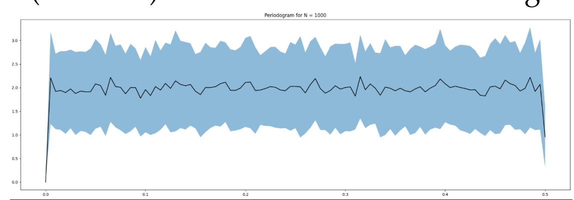
Answer 9



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Observation : The variance is indeed divided by 5 compared to Question 6 (around 3.85 to 0.74).

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 falls. 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 and 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.

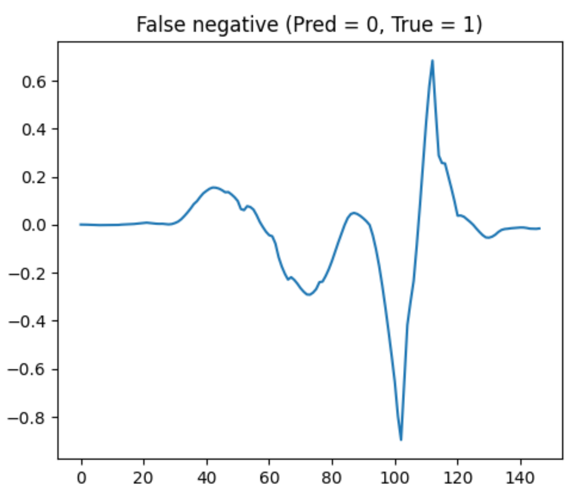
Answer 10

We tested k between 2 and 15 with 5 folds cross validation on the train set and we found the optimal $k = 3$ with $f1\_score = 0.88$ on the train set.
We evaluated the knn with $k = 3$ on the test set and we obtained the $f1\_score = 0.48$ on this set.
The is $f1\_score$ is close to the random experiment for the test set even if it performs well for the train set.

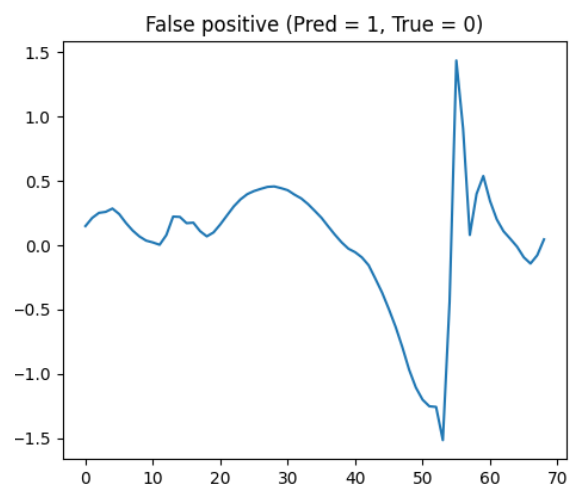
Question 11

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).