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

Louis-Marie Lovichi lm.lovichi@me.com
Gaspard Berthelier gaspard.berthelier@student-cs.fr

December 12, 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 7th 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 (??) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

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

On cherche à trouver $\lambda_{\max}$ tel que $\forall \lambda > \lambda_{\max}$, $\mathrm{argmin}_\beta f(\beta, \lambda) = 0$.

Notons $\beta_\lambda^* = \mathrm{argmin}_\beta f(\beta, \lambda)$. On a alors : $f(\beta_\lambda^*) \leq f(0, \lambda) = \frac{1}{2}||y||_2^2$

C'est à dire :

$$\frac{1}{2}\sum_i (y_i - [X\beta_\lambda^*]_i)^2 + \lambda \sum_i |\beta_{\lambda,i}^*| \leq \frac{1}{2}\sum_i y_i^2$$

$$\Leftrightarrow \sum_{i,j}[X_{ij}^2 \beta_{\lambda,j}^{*2} - 2y_i X_{ij}\beta_{\lambda,j}^*] + 2\lambda \sum_i |\beta_{\lambda,i}^*| \leq 0$$

$$\Leftrightarrow \sum_{i,j}[X_{ij}^2|\beta_{\lambda,j}^*| + 2\lambda]|\beta_{\lambda,j}^*| \leq \sum_{i,j} 2y_i X_{ij}\beta_{\lambda,j}^*$$

Donc par majoration :

$$\sum_{i,j}[X_{ij}^2|\beta_{\lambda,j}^*| + 2\lambda]|\beta_{\lambda,j}^*| \leq \sum_{i,j} 2|y_i||X_{ij}||\beta_{\lambda,j}^*|$$

Si on fait l'hypothèse $\beta_\lambda^* \neq 0$:

$$\Leftrightarrow \forall(i,j), \ X_{ij}^2|\beta_{\lambda,j}^*| + 2\lambda \leq 2|y_i||X_{ij}|$$

$$\Leftrightarrow \forall(i,j), \ 0 \leq X_{ij}^2|\beta_{\lambda,j}^*| \leq 2(|y_i||X_{ij}| - \lambda)$$

On constate donc qu'il suffit de prendre $\lambda > \lambda_{\max} = ||y||_\infty ||X||_{\max}$ pour avoir une contradiction car le terme à droite serait négatif. Ainsi pour $\lambda > \lambda_{\max}$, on a bien $\beta_\gamma^* = 0$.

$$\lambda_{\max} = ||y||_\infty ||X||_\infty \tag{2}$$

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:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k ||\mathbf{d}_k||_2^2 \leq 1} \left\| \mathbf{x} - \sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^K ||\mathbf{z}_k||_1 \tag{3}$$

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

Let's consider a fixed dictionnary $(d_k)_k$ and the sparce coding problem :

$$(z_k)_k^{\lambda *} = \operatorname{argmin}_{(z_k)_k} \|x - \sum_k z_k * d_k\|_2^2 + \lambda \sum_k \|z_k\|_1$$

The quantity $z_k * d_k[n]$ can be defined for $n \in [\![1, N]\!]$ with :

$$z_k * d_k[n] = \sum_{m=1}^{N} \tilde{z}_k[m]\tilde{d}_k[n-m]$$

where $\tilde{z}_k$ and $\tilde{d}_k$ are padded versions of z_k and d_k, of size N and with zeros if z_k or d_k is not defined. So:

$$z_k * d_k[n] = (d_k^{\widetilde{(n)}})^T \tilde{z}_k$$

where $d_k^{\widetilde{(n)}}$ is equal to $\tilde{d}_k$ with an inversion followed by n-2 circular translations of the indexes.

We can concatenate the $\tilde{d}_k^{(n)}$ vectors as lines into a $\tilde{D}_k$ matrix of size $N * N$ such that $z_k * d_k$ is the vector $\tilde{D}_k \tilde{z}_k$ of size N. So we have :

$$\sum_{k=1}^{K} z_k * d_k = \sum_{k=1}^{K} \tilde{D}_k \tilde{z}_k$$

We can express this sum as one big matrix product by concatenating $\tilde{D}_k$ along the columns ($\tilde{D}$ matrix of size $N * KN$) and $\tilde{z}_k$ as one vector $\tilde{z}$ of size NK, which gives us :

$$\sum_{k=1}^{K} z_k * d_k = \tilde{D}\tilde{z}$$

As it turns out :

$$\|\tilde{z}\|_1 = \sum_{k,n} |z_k[n]| = \sum_{k=1}^{K} \sum_{n} |z_k[n]| = \sum_{k=1}^{K} \|z_k\|_1$$

So we have reformulated the problem as :

$$(z_k)_k^{\lambda *} = \text{argmin}_{(z_k)_k} \|x - \tilde{D}\tilde{z}\|_2^2 + \lambda \|\tilde{z}\|_1$$

By factorizing by $\frac{1}{2}$ and renaming 2λ as λ, $\tilde{z}$ as β, $\tilde{D}$ as X and taking $p = KN$, we have exactly the same problem as Question 1. So the sparse coding problem is indeed a lasso regression.

$$\lambda_{\max} = \|x\|_\infty \|\tilde{D}\|_\infty \tag{4}$$

3 Spectral feature

Let X_n ($n = 0, \ldots, N - 1$) 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:
$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{5}$$

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

By definition of a white noise, $X = (X_n)_n$ is wide-sense stationary, centered and $\gamma_X(\tau) = \sigma^2 \delta(\tau)$ where $\sigma^2 = V(X_0) = E(X_0^2)$.

So the autocovariance is equal to $\sigma^2 \delta(\tau)$.
Now we can compute the power spectrum :

$$S(f) = \sum_{\tau \in \mathbb{Z}} \gamma(\tau) e^{-2i\pi f \tau / f_s} = \sigma^2 \sum_{\tau \in \mathbb{Z}} \delta(\tau) e^{-2i\pi f \tau / f_s} = \sigma^2$$

The power spectrum is constant for all frequences and equal to the variance.

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{6}$$

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 compute the bias of the sample autocovariance :

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

$$\Rightarrow \text{bias} = E(\hat{\gamma}(\tau)) - \gamma(\tau) = \frac{-\tau}{N} \gamma(\tau)$$

We can see that for $\tau \neq 0$, we have bias $\neq 0$ so the estimator is biased. But the bias tends towards 0 as N increases, so it is asymptotically unbiased.

We could compute a de-biased estimator by setting $\tilde{\gamma}(\tau) = \frac{N}{N-\tau} \hat{\gamma}(\tau)$.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \tag{7}$$

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

We have :

$$|J(f)|^2 = J(f)J(f)^* = \frac{1}{N} \left(\sum_{n=0}^{N-1} X_n e^{-2i\pi f n / f_s} \right) \left(\sum_{m=0}^{N-1} X_n e^{+2i\pi f n / f_s} \right)$$

$$= \frac{1}{N} \left(\sum_{n=0}^{N-1} X_n^2 + \sum_{0 \leq n < m}^{m=N-1} X_n X_m \left(e^{-2i\pi f(n-m)/f_s} + e^{-2i\pi f(m-n)/f_s} \right) \right)$$

5

$$= \hat{\gamma}(0) + \frac{1}{N}\left(\sum_{n=0}^{N-1}\sum_{\tau=1}^{N-1-n} X_n X_{n+\tau}(e^{-2i\pi f\tau/f_s} + e^{2i\pi f\tau/f_s})\right)$$

Let's note $h(\tau) = e^{-2i\pi f\tau/f_s}$. We get :

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

$$= \hat{\gamma}(0) + \sum_{\tau=1}^{N-1}\hat{\gamma}(\tau)(h(\tau) + h(-\tau))$$

Since $\hat{\gamma}(-\tau) = \hat{\gamma}(\tau)$ and $h(0) = 1$, we obtain :

$$|J(f)|^2 = \sum_{\tau=-(N-1)}^{N-1}\hat{\gamma}(\tau)h(\tau)$$

So for $f = f_k = \frac{f_s k}{N}$, we get :

$$|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1}\hat{\gamma}(\tau)e^{-2i\pi\tau k/N}$$

For $f = f^{(N)}$, we get :

$$E(|J(f^{(N)})|^2) = \sum_{\tau=-(N-1)}^{N-1} E(\hat{\gamma}(\tau))e^{-2i\pi\tau f^{(N)}/f_s} = \sum_{\tau=-(N-1)}^{N-1}\frac{N-\tau}{N}\gamma(\tau)e^{-2i\pi\tau f^{(N)}/f_s}$$

We must show that this quantity tends towards $S(f)$ as N increases. We have :

$$|S(f) - E(|J(f^{(N)})|^2)| = |\sum_{\tau\in\mathbb{Z}}\gamma(\tau)e^{-2i\pi f\tau/f_s} - \sum_{\tau=-(N-1)}^{N-1}\frac{N-\tau}{N}\gamma(\tau)e^{-2i\pi\tau f^{(N)}/f_s}|$$

$$= |\sum_{|\tau|\geq N}\gamma(\tau)h(\tau) + \sum_{\tau=-(N-1)}^{N-1}\gamma(\tau)(e^{-2i\pi f\tau/f_s} - (1-\frac{\tau}{N})e^{-2i\pi\tau f^{(N)}/f_s})|$$

$$\leq \sum_{|\tau|\geq N}|\gamma(\tau)| + \sum_{\tau=-(N-1)}^{N-1}|\gamma(\tau)|(|e^{-2i\pi f\tau/f_s} - e^{-2i\pi\tau f^{(N)}/f_s}| + \frac{|\tau|}{N})$$

We have : $|e^{-2i\pi f\tau/f_s} + e^{-2i\pi\tau f^{(N)}/f_s}| = |2i\sin(i\pi\tau(f^{(N)} - f)\frac{1}{f_s})e^{-i\pi\tau(f+f^{(N)})/f_s}|$
$= 2|\sin(i\pi\tau(f^{(N)} - f)\frac{1}{f_s})| \leq 2\min(1, \pi\frac{|\tau|}{f_s}|f^{(N)} - f|)$

By definition of $f^{(N)}$, we have $\lim_{N\to+\infty}|f^{(N)} - f| = 0$.
Also, since $\sum|\gamma(\tau)|$ is summable, we have $\lim_{N\to+\infty}\sum_{|\tau|\geq N}|\gamma(\tau)| = 0$

For $\epsilon > 0$, let's use N_0 such that $\forall N \geq N_0$, $\max(\sum_{|\tau|\geq N}|\gamma(\tau)|, |f^{(N)} - f|) \leq \epsilon$. For $N \geq N_0$:

$$|S(f) - E(|J(f^{(N)})|^2)| \leq \epsilon + \sum_{\tau=-(N-1)}^{N-1}|\gamma(\tau)|(2\min(1, |\tau|\frac{\pi}{f_s}|f^{(N)} - f|) + \frac{|\tau|}{N})$$

Let's note $h(\tau, N) \to_{N \to \infty} = 0$ the term inside the sum.
We now have for $\tilde{N} \geq 0$:

$$|S(f) - E(|J(f^{(N)})|^2)| \leq \epsilon + \sum_{|\tau| < \tilde{N}} h(\tau, N) + \sum_{\substack{\tau = -(N-1) \\ |\tau| \geq \tilde{N}}}^{N-1} h(\tau, N)$$

The fist sum has a finite number of elements that tend towards 0. We can easily choose $\tilde{N} \geq N_0$ such that $\forall N \geq \tilde{N}$:

$$|S(f) - E(|J(f^{(N)})|^2)| \leq 2\epsilon + \sum_{\tilde{N} \leq |\tau| \leq N-1} h(\tau, N)$$

Where inside the sum :

$$h(\tau, N) \leq |\gamma(\tau)| \left[2 \min(1, \epsilon \frac{\pi}{f_s}) + 1 \right]$$

By taking ϵ small enough we get inside the sum : $h(\tau, N) \leq (1 + \epsilon 2\pi / f_s)|\gamma(\tau)|$ By summability of $|\gamma(\tau)|$ once again, the sum above tends to 0 hence we have proven that the estimator is asymptotically unbiased.
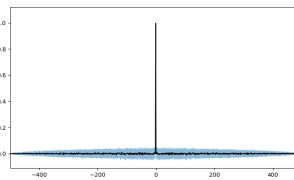
Question 6

In this question, let X_n ($n = 0, \ldots, N-1$) 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?
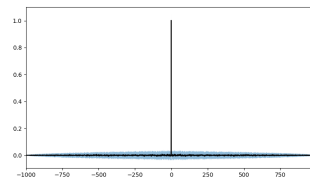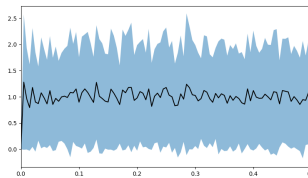
Add your plots to Figure **??**.
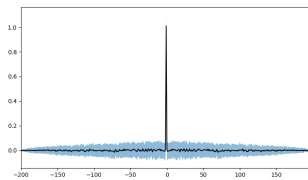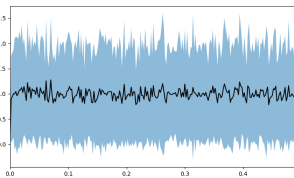

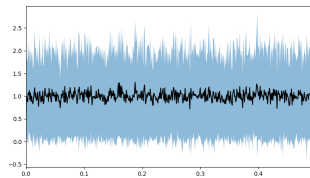Autocovariance ($N = 200$)   Autocovariance ($N = 500$)   Autocovariance ($N = 1000$)

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

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

Answer 6

- We can conclude that the variance of the estimator for the autocovariance decreases as $N \to +\infty$.

- We can conclude that the variance of the estimator for the periodogram does not decrease as $N \to +\infty$.

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 $\tau > 0$

$$\text{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N} \right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau) \right]. \tag{8}$$

(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

- For $\tau > 0$:

$$V(\hat{\gamma}(\tau)) = \frac{1}{N^2} V\Big(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \Big) = \frac{1}{N^2} \sum_{0 \le n \le m \le N-L-1} (2 - \delta_{n,m}) \text{cov}(X_n X_{n+\tau}, X_m X_{m+\tau})$$

With $2 - \delta_{n,m} = 1$ if $n = m$, otherwise 2. Also :

$$\text{cov}(X_n X_{n+\tau}, X_m X_{m+\tau}) = E(X_n X_{n+\tau} X_m X_{m+\tau}) - E(X_n X_{n+\tau})E(X_m X_{m+\tau})$$

It is not indicated that the X_n variables are centered but judging by the hint given in the subject, we will assume they are :

$$\text{cov}(X_n X_{n+\tau}, X_m X_{m+\tau}) = E(X_n X_m)E(X_{n+\tau} X_{m+\tau}) + E(X_n X_{m+\tau})E(X_{n+\tau} X_m)$$

$$= r_X(n, m)r_X(n+\tau, m+\tau) + r_X(n, m+\tau)r_X(n+\tau, m)$$

Also, we can rewrite the double sum as follows :

$$\sum_{0 \le n \le m \le N-L-1} (2 - \delta_{n,m}) \text{cov}(X_n X_{n+\tau}, X_m X_{m+\tau}) = \sum_{n=0}^{N-L-1} \sum_{k=0}^{N-L-1-n} (2 - \delta_k) \text{cov}(X_n X_{n+\tau}, X_{n+k} X_{n+k+\tau})$$

8

$$= \sum_{n=0}^{N-L-1} \sum_{k=0}^{N-L-1-n} (2 - \delta_k)[r_X(n, n+k)r_X(n+\tau, n+\tau+k) + r_X(n, n+\tau+k)r_X(n+\tau, n+k)]$$

$$= \sum_{n=0}^{N-L-1} \sum_{k=0}^{N-L-1-n} (2 - \delta_k)[\gamma(k)^2 + \gamma(\tau+k)\gamma(\tau-k)]$$

The switch from the correlation coefficients to the intercorrelation coefficients with no dependency on n comes from the stationary hypothesis. So we have :

$$V(\hat{\gamma}(\tau)) = \frac{1}{N^2} \sum_{k=0}^{N-L-1} \sum_{n=0}^{N-L-1-k} (2 - \delta_k)[\gamma(k)^2 + \gamma(\tau+k)\gamma(\tau-k)]$$

Let's call $h(k) = \gamma(k)^2 + \gamma(\tau+k)\gamma(\tau-k)$. It does not depend on n. So :

$$V(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{n=0}^{N-L-1} (1 - \frac{L+n}{N})(2 - \delta_n)h(n)$$

$$V(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{n=0}^{N-L-1} (1 - \frac{L+n}{N})(2 - \delta_n)h(n)$$

Since $h(-n) = h(n)$, we actually get :

$$V(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{n=-(N-L-1)}^{N-L-1} (1 - \frac{L+|n|}{N})h(n)$$

- So, the Byenaymé-Tcheybychev inequality let us write :

$$\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| > t) \leq \frac{\mathbb{V}(\hat{\gamma}(\tau))}{t^2} \underset{n \to +\infty}{\to} 0$$

Thus, $\hat{\gamma}(\tau)$ 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 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 $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi fn/f_s)$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi fn/f_s$. 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 **??** by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

- Since $\mathbb{E}(X_n) = 0$, on a $\mathbb{E}[A(f)] = \mathbb{E}[B(f)] = 0$. For $f = f_s\frac{k}{N}$, we compute :

$$
\begin{aligned}
\mathbb{V}(A(f)) &= \sum_{n=0}^{N-1} cos^2(-2\pi k\frac{n}{N})\mathbb{V}(X_n) \\
&= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} (cos(-4\pi k\frac{n}{N}) + 1) \\
&= N\frac{\sigma^2}{2}
\end{aligned}
$$

Similarly, we compute :

$$
\mathbb{V}(B(f)) = N\frac{\sigma^2}{2}
$$

- We obtain $|J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2)$. $|J(f_k)|^2$ is the sum of two independant Gaussian with mean equals 0 and variance $N\frac{\sigma^2}{2}$, then, $|J(f_k)|^2$ follows $\frac{\sigma^2}{2}\chi^2(2)$, where $\chi_2(2)$ is a chi-square distribution with 2 degrees of freedom.

- Then, $\mathbb{V}(|J(f_k)|^2) = \sigma^4$ and we conclude that the periodogram is not consistent.

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 **??**, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe.

Add your plots to Figure **??**.

Answer 9

- We can conclude that the variance indeed decreased by a $\sqrt{5}$-factor as it is expected.
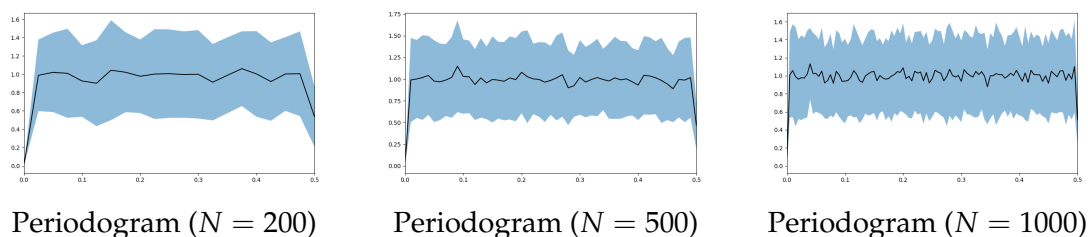
| Periodogram ($N = 200$) | Periodogram ($N = 500$) | Periodogram ($N = 1000$) |

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

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

With 5-fold cross-validation, we can conclude that the optimal number of neighbors is 3 whose associated F1-score is 0.786. You can find the code in the notebook. We have randomly shuffle the

train et test datasets to ensure good F1-score (the initial train et test datasets did not provide such a F1-score).
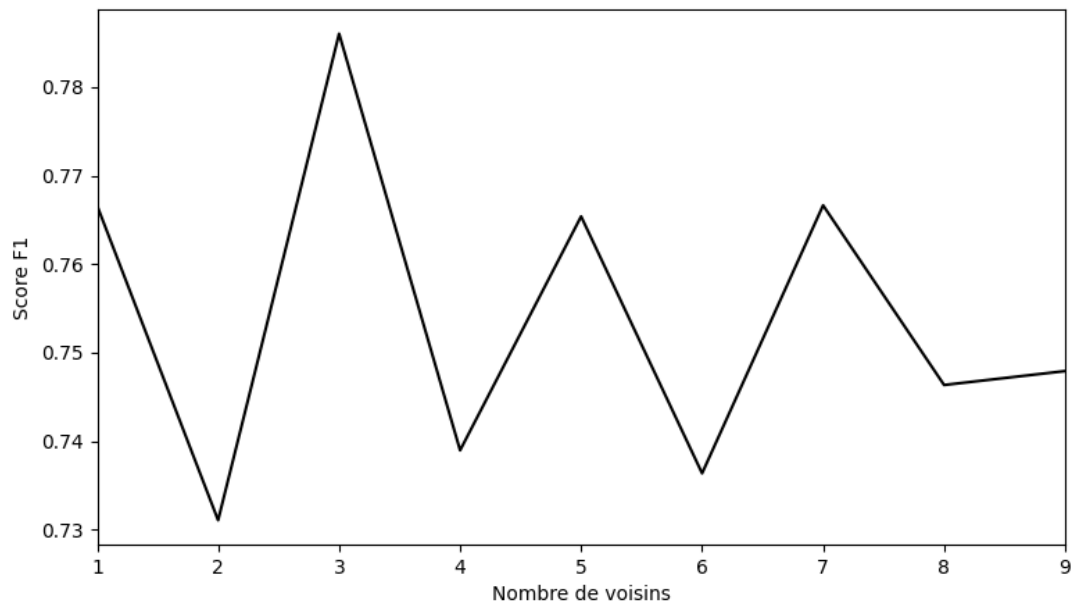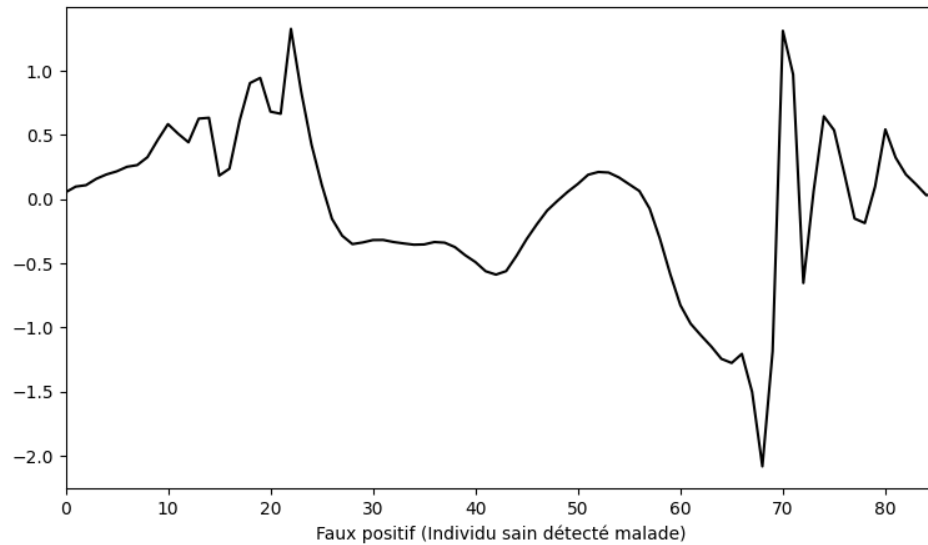

Figure 3: Selection of the optimal number of neighbors *k* via the associated F-score.
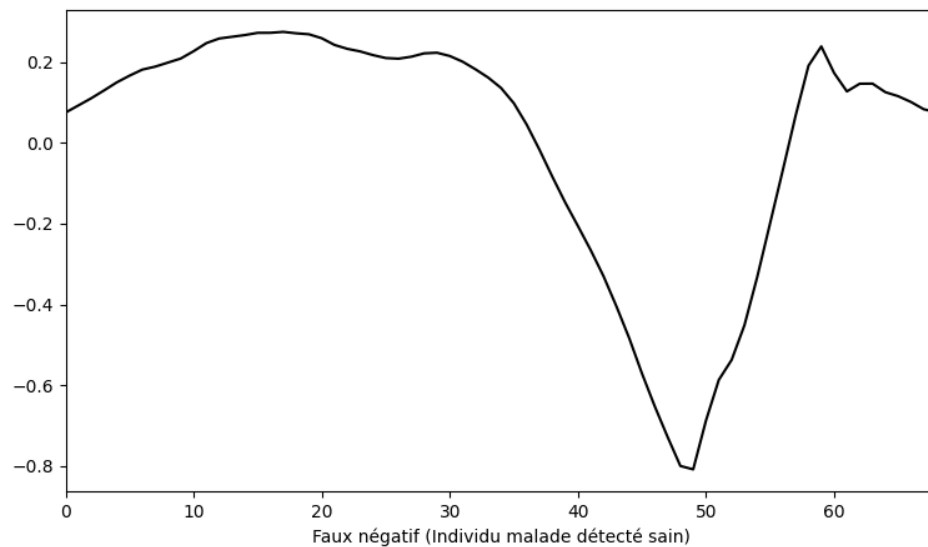
Question 11

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

Answer 11

Badly classified healthy step

Badly classified non-healthy step

Figure 4: Examples of badly classified steps (see Question **??**).