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

Baptiste CALLARD baptcallard@gmail.com
Adhemar Senneville adhemar.senneville@gmail.com

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 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 (1) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

$$\lambda_{\max} = \left\| X^T y \right\|_\infty \tag{2}$$

We note:

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

Let β^* be the minimizer of the Lasso problem (1), which is a convex problem. Since we are in a convex problem, KKT allows us to write the following first-order optimality condition : β^*est solution de (1) $\Leftrightarrow 0 \in \partial g(\beta^*) \Leftrightarrow 0 \in X^T(y - X\beta^*) + \partial \|\beta^*\|_1$. Thus we get :

$$\beta^* \text{is solution of (1)} \Leftrightarrow -X^T(y - X\beta^*) + \gamma\lambda = 0 \tag{3}$$

$$\text{with } \gamma \text{ the sub-gradient of } \|.\|_1 \ : \gamma_i \in \begin{cases} \text{sign}(\beta^*) & \text{if } \beta_i^* \neq 0 \\ [-1, 1] & \text{if } \beta_i^* = 0 \end{cases} \quad \forall i \in \{1, ..., d\}$$

We want to find the value of λ_{max} such that $\beta^* = 0_p$.

Let $\lambda > 0$:

$$\begin{aligned} \beta^* = 0 &\Longleftrightarrow X^T(y - X\beta^*) = X^T y = \lambda\gamma \text{ from (3)} \\ &\Longleftrightarrow \left| x_i^T y \right| \leq \lambda \quad \forall i = 1, ..., p \\ &\Longleftrightarrow \max_{i=1,...,p} \left| x_i^T y \right| \leq \lambda \\ &\Longleftrightarrow \left\| X^T y \right\|_\infty \leq \lambda \end{aligned}$$

Therfore, pour tout $\lambda \geq \lambda_{max} = \|X^T y\|_\infty$ is equivalent to $\beta^* = 0$. Thus, we have shown the existence of λ_{max}

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:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{4}$$

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 $d_k \in \mathbb{R}^L$ the K atoms of the fixed dictionary and $z_k \in \mathbb{R}^{N-L+1}$.

We have the following equation (5), which is very similar to the Lasso equation explained in the previous question. So, in the following optimization problem, we would like the term $\sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k$ to be of the form of Dz and then formulate the problem as (6).

For a fixed dictionary, we would like to rewrite the problem:

$$\min_{(\mathbf{z}_k)_k} \quad \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{5}$$

as the following

$$\min_{z \in \mathbb{R}^{KN}} \quad \|\mathbf{x} - D\mathbf{z}\|_2^2 \quad + \quad \lambda \|\mathbf{z}\|_1 \quad D \in \mathbb{R}^{N \times KN} \tag{6}$$

To write the convolution product $\mathbf{z}_k * \mathbf{d}_k$ as a matrix product, we define $\widetilde{\mathbf{z}}_k$ (resp. $\widetilde{\mathbf{d}}_k$) which correspond to the version of $\mathbf{z}_k$ (resp. $\mathbf{d}_k$) with padding of 0 such that the vectors are of size N.

$$\widetilde{\mathbf{z}}_k(j) = \begin{cases} \mathbf{z}_k(j) & \text{if} \quad j \in \{1, ..., N-L+1\} \\ 0 & otherwise \end{cases} \quad \widetilde{\mathbf{d}}_k(i-j) = \begin{cases} \mathbf{d}_k(i-j) & \text{if} \quad (i-j) \in \{1, ..., L\} \\ 0 & otherwise \end{cases}$$

We have

$\mathbf{z_k} * \mathbf{d_k} = \widetilde{\mathbf{z}}_k * \widetilde{\mathbf{d}}_k$ then for all i $\in \{1, ..., N\}$,

$$\left[\sum_{k=1}^{K} \widetilde{\mathbf{z}}_k * \widetilde{\mathbf{d}}_k\right](i) = \sum_{k=1}^{K} \sum_{j=1}^{N} \widetilde{\mathbf{z}}_k(j) \widetilde{\mathbf{d}}_k(i-j)$$

So we can rewrite the last term in matrix form.

Let Z be the vector that concatenates the vectors $\widetilde{\mathbf{z}}_k$:

$\widetilde{Z} = \begin{bmatrix} \widetilde{Z}(1) & ... & \widetilde{Z}(K) \end{bmatrix} \in \mathbb{R}^{KN}$

with

$$
\begin{aligned}
\widetilde{Z}_1 &= [\widetilde{z}_1(1), \quad ..., \quad \widetilde{z}_1(n)] \\
\vdots &= \qquad \vdots \\
\widetilde{Z}_K &= [\widetilde{z}_K(1), \quad ..., \quad \widetilde{z}_K(n)]
\end{aligned}
$$

Similarly, we define $\widetilde{D}$:

$\left[\widetilde{D}_1 \quad ... \quad \widetilde{D}_K\right] \in \mathbb{R}^{N\times KN}$ with $\widetilde{D}_k = (\widetilde{d}_k[i-j])_{1\leq i,j\leq N} \in \mathbb{R}^{N\times N}$

Let's show that

$\left[\sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k\right](i) = \widetilde{D}\widetilde{Z}(i)$

$$
\begin{aligned}
\widetilde{D}\widetilde{Z}(i) &= \sum_{k=1}^{K} \left[\widetilde{D}_k \widetilde{Z}_k\right](i) \\
&= \sum_{j=1}^{K} \sum_{k=1}^{N} \widetilde{D}_k(i,j)\widetilde{Z}_k(j) \\
&= \sum_{j=1}^{K} \sum_{k=1}^{N} \widetilde{d}_k[i-j]\widetilde{z}_k(j) \\
&= \left[\sum_{k=1}^{K} \widetilde{\mathbf{z}}_k * \widetilde{\mathbf{d}}_k\right](i) \\
&= \left[\sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k\right](i)
\end{aligned}
$$

Finally we also have :

$$
\left\|\widetilde{Z}\right\|_1 = \sum_{k=1}^{K} \|\widetilde{z}_k\|_1 = \sum_{k=1}^{K} \|z_k\|_1
$$

Therefore, we can rewrite (5) as

$$
\min_{\widetilde{z}\in\mathbb{R}^{KN}} \quad \left\|\mathbf{x} - \widetilde{D}\widetilde{Z}\right\|_2^2 \quad + \quad \lambda \left\|\widetilde{Z}\right\|_1 \quad \widetilde{D} \in \mathbb{R}^{N\times KN}
$$

We retreive the form of Lasso problem (1). Thus, we have already given the form of λ_{max}.

$$
\lambda_{\max} = \left\|\widetilde{D}^T x\right\|_\infty \tag{7}
$$

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{8}$$

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

We want to show that

$$\begin{gathered} X_n \sim \mathcal{N}(0, \sigma^2) \\ \gamma(\tau) = \sigma^2 \mathbb{1}_{\{\tau=0\}} \\ S(f) = \sigma^2 \end{gathered}$$

Recall that $\mathbb{E}(X_n) = 0$

$$\begin{aligned} \text{if } \tau = 0, \gamma(\tau) &= \mathbb{E}(X_n X_n) \\ &= Var(X_n) \text{ by independance of white noise} \\ &:= \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{if } \tau \neq 0, \gamma(\tau) &= \mathbb{E}(X_n X_{n+\tau}) \\ &= \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) \quad \text{by independance of white noise} \\ &= 0 \end{aligned}$$

$$
\gamma(\tau) = \mathbb{E}[(X(t))(X(t+\tau))] = \begin{cases} \sigma^2 & \text{if } \tau = 0 \\ 0 & \text{if } \tau \neq 0 \end{cases}
$$
$$
= \sigma^2 \mathbb{1}_{\{\tau=0\}}
$$

$$
\begin{aligned}
S(f) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s} \\
&= \sum_{\tau=-\infty}^{+\infty} (\sigma^2 \mathbb{1}_{\{\tau=0\}}) e^{-2\pi f \tau / f_s} \\
&= \sigma^2 e^{-2\pi f * 0 / f_s} \\
&= \sigma^2
\end{aligned}
$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$
\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{9}
$$

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

The bias is defined as $\mathbf{b}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$.

For $\tau \geq 0$

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

The same result is obtained for $\tau < 0$.

Hence, for all $\lambda \in \mathbb{R}$ the estimator is biaised $\mathbb{E}\left[\hat{\gamma}(\tau)\right] \neq \gamma(\tau)$, but this estimator is asymptotically unbiased because $\mathbb{E}\left[\hat{\gamma}(\tau)\right] \xrightarrow[N \to +\infty]{} \gamma(\tau)$.

It is suffisant to consider the estimator $\frac{N}{N-\tau} \hat{\gamma}(\tau)$ which is a version un-biaised of $\hat{\gamma}(\tau)$.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{10}$$

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

- First let's write $|J(f_k)|^2$ as a function of the sample autocovariances.

$$\begin{aligned}
|J(f_k)|^2 &= \left| \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}} \right|^2 \\
&= \frac{1}{N} \left| \sum_{n=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}} \right|^2 \\
&= \frac{1}{N} \left(\sum_{n=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}} \right) \overline{\left(\sum_{p=0}^{N-1} X_p e^{-\frac{2\pi i k p}{N}} \right)} \\
&= \frac{1}{N} \left(\sum_{n=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}} \right) \left(\sum_{p=0}^{N-1} X_p e^{\frac{2\pi i k p}{N}} \right) \\
&= \frac{1}{N} \left(\sum_{n=0}^{N-1} \sum_{p=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}} X_p e^{\frac{2\pi i k p}{N}} \right) \\
&= \frac{1}{N} \left(\sum_{n=0}^{N-1} \sum_{p=0}^{N-1} X_n X_p e^{-\frac{2\pi i k (n-p)}{N}} \right) \quad \text{let set } h = p - n \\
&= \frac{1}{N} \left(\sum_{n=0}^{N-1} \sum_{h=-n}^{N-1-n} X_n X_{h+n} e^{\frac{2\pi i k h}{N}} \right)
\end{aligned}$$

We can then permute the sum

$$\begin{cases} 0 \le n \le N-1 \\ -n \le h \le N-1-n \end{cases} \iff \begin{cases} 0 \le n \le N-1-h \\ -(N-1) \le h \le N-1 \end{cases}$$

$$
\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\left(\sum_{h=-N+1}^{N-1}\sum_{n=0}^{N-1-h} X_n X_{h+n} e^{\frac{2\pi ikh}{N}}\right) \\
&= \sum_{h=-N+1}^{N-1} \hat{\gamma}(h) e^{-\frac{2\pi ikh}{N}} \quad \text{as } \hat{\gamma}(\tau) := (1/N)\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \quad \text{and} \quad \hat{\gamma}(.) \text{ is even.}
\end{aligned}
$$

- Secondly we define $f^{(N)}$ the closest Fourier frequency f_k to f.

$$
\begin{aligned}
f^{(N)} = f_k &\Leftrightarrow \frac{k-0.5}{N} f_s \leq f < \frac{k+0.5}{N} f_s \\
&\Leftrightarrow k \leq \frac{f}{f_s} N + 0.5 < k+1 \\
&\Leftrightarrow k = \lfloor N\frac{f}{f_s} + 0.5 \rfloor \\
&\Leftrightarrow f^{(N)} = \lfloor N\frac{f}{f_s} + 0.5 \rfloor \frac{f_s}{N}
\end{aligned}
$$

To prove that $f^{(N)}$ tends to f as N tends to infinity, let's split the floor function into its integer and fractional parts:

$$
N\frac{f}{f_s} + 0.5 = \lfloor N\frac{f}{f_s} + 0.5 \rfloor + \{N\frac{f}{f_s} + 0.5\}
$$

Here, $\lfloor N\frac{f}{f_s} + 0.5 \rfloor$ is the integer part, and $\{N\frac{f}{f_s} + 0.5\}$ is the fractional part, $0 \leq \{N\frac{f}{f_s} + 0.5\} < 1$.

$$
\begin{aligned}
f^{(N)} &= \lfloor N\frac{f}{f_s} + 0.5 \rfloor \frac{f_s}{N} = (N\frac{f}{f_s} + 0.5 - \{N\frac{f}{f_s} + 0.5\})\frac{f_s}{N} \\
&= f + 0.5\frac{f_s}{N} - \{N\frac{f}{f_s} + 0.5\}\frac{f_s}{N} \\
\lim_{N\to\infty} f^{(N)} &= f + 0 - 0 \\
&= f
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[|J(f^{(N)})|^2\right] &= \sum_{\tau=-N+1}^{N-1} \mathbb{E}\left[\hat{\gamma}(\tau)\right] e^{-\frac{2\pi i f^{(N)}\tau}{f_s}} \\
&= \sum_{\tau=-N+1}^{N-1} \frac{N-\tau}{N}\gamma(\tau) e^{-\frac{2\pi i f^{(N)}\tau}{f_s}} \\
&= \sum_{\tau=-N+1}^{N-1} \left(1-\frac{\tau}{N}\right)\gamma(\tau) e^{-\frac{2\pi i f^{(N)}\tau}{f_s}} \\
&= \sum_{\tau=-\infty}^{\infty} \left(1-\frac{\tau}{N}\right)\gamma(\tau) e^{-\frac{2\pi i f^{(N)}\tau}{f_s}} \mathbb{1}_{\tau\in\{-N+1,\dots,N-1\}}
\end{aligned}
$$

As the term $\frac{\tau}{N}$ tends to 0 when $N \to \infty$, and $f^{(N)}$ tends to f. And we also have, for all τ, $\left|\left(1-\frac{\tau}{N}\right)\gamma(\tau)e^{-\frac{2\pi i f^{(N)}\tau}{f_s}}\mathbb{1}_{\tau\in\{-N+1,\dots,N-1\}}\right| < |\gamma(\tau)|$ and $\sum_{\tau=-\infty}^{\infty}|\gamma(\tau)| < +\infty$. Thus by dominated convergence, the limit is:

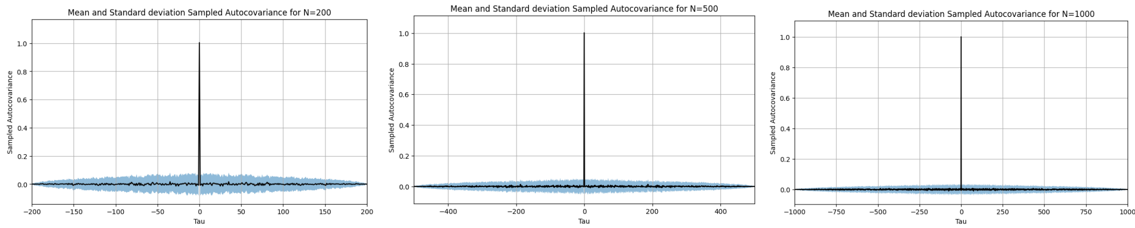
$$
\lim_{N\to\infty} \mathbb{E}\left[|J(f^{(N)})|^2\right] = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-\frac{2\pi i f\tau}{f_s}} = S(f)
$$

Question 6

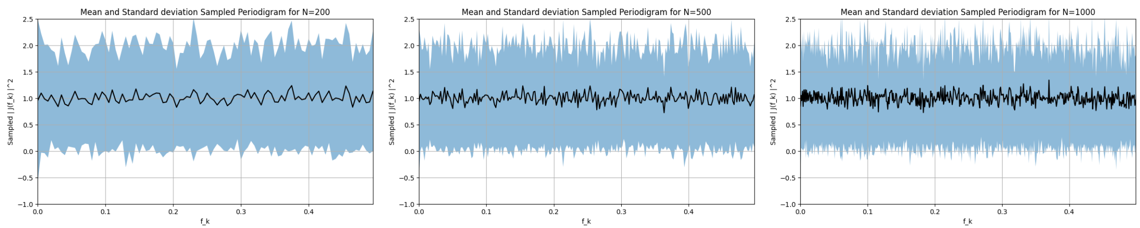
In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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.



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

In the sampled autocovariance plot for white noise, we observe that when the lag τ is zero, the average value of the sample covariance matches the standard deviation 1. For non-zero τ, the average sample covariance approaches zero. Additionally, as the number of generated samples increases, the standard deviation of the sample covariance decreases. This trend suggests that the average value of the sample covariance is converging to zero for $\tau \neq 0$. These observations are consistent with the theoretical expectation that $\hat{\gamma}(\tau)$ is an asymptotically unbiased estimator of $\gamma(\tau)$, where $\gamma(\tau) = \sigma^2 \mathbb{1}_{\{\tau=0\}}$.

In the sampled periodogram plot for white noise, we observe that the mean value is constant for all f_k and aproches the standard deviation 1. This matches theoretical results : $S(f) = \sigma^2$. However, the standard deviation doesn't decreases with the number of samples and is aproximatively 1.

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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{11}$$

  (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

- Let's derive the formula of $var(\hat{\gamma}(\tau))$

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

Indeed, according to question 4: $\mathbb{E}\left[\hat{\gamma}(\tau)\right] = \frac{N-\tau}{N}\gamma(\tau)$. Using the squared sum expansion formula.

$$
\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right)^2\right] = \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}\left[(X_n X_{n+\tau})^2\right] + \frac{2}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{p=n+1}^{N-\tau-1}\mathbb{E}\left[X_n X_{n+\tau} X_p X_{p+\tau}\right]
$$

Then you can use the hint:

$$
\begin{aligned}
\frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}\left[(X_n X_{n+\tau})^2\right] &= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}\left[X_n X_{n+\tau} X_n X_{n+\tau}\right] \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_n X_{n+\tau}] + \mathbb{E}[X_n^2]\mathbb{E}[X_{n+\tau}^2] + \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_n X_{n+\tau}] \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1} 2\left(\mathbb{E}[X_n X_{n+\tau}]\right)^2 + \mathbb{E}[X_n^2]\mathbb{E}[X_{n+\tau}^2] \\
&= \frac{N-\tau}{N^2}\left(2\gamma_X(\tau)^2 + \gamma_X(0)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{2}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{p=n+1}^{N-\tau-1}\mathbb{E}\left[X_nX_{n+\tau}X_pX_{p+\tau}\right] &= \frac{2}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{p=n+1}^{N-\tau-1}\mathbb{E}[X_nX_{n+\tau}]\mathbb{E}[X_pX_{p+\tau}]+\mathbb{E}[X_nX_p]\mathbb{E}[X_{n+\tau}X_{p+\tau}]+\\
&+\mathbb{E}[X_nX_{p+\tau}]\mathbb{E}[X_{n+\tau}X_p]\\
&= \frac{2}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{p=n+1}^{N-\tau-1}\gamma_X(\tau)^2+\gamma_X(n-p)^2+\gamma_X(n-p-\tau)\gamma_X(n+\tau-p)\\
(\textit{set } h=n-p \text{ and sum permutation}) &= \frac{2}{N^2}\sum_{h=1}^{N-\tau-1}\sum_{p=0}^{N-\tau-1-h}\gamma_X(\tau)^2+\gamma_X(h)^2+\gamma_X(h-\tau)\gamma_X(h+\tau)\\
(\textit{as } \gamma_X(.) \text{ is even}) &= \frac{1}{N^2}\sum_{\substack{h=-(N-\tau-1)\\ h\neq 0}}^{N-\tau-1}\sum_{p=0}^{N-\tau-1-|h|}\gamma_X(\tau)^2+\gamma_X(h)^2+\gamma_X(h-\tau)\gamma_X(h+\tau)\\
&= \frac{1}{N}\sum_{\substack{h=-(N-\tau-1)\\ h\neq 0}}^{N-\tau-1}(1-\frac{\tau+|h|}{N})[\gamma_X(\tau)^2+\gamma_X(h)^2+\gamma_X(h-\tau)\gamma_X(h+\tau)]
\end{aligned}
$$

But :

$$
\frac{1}{N}\sum_{\substack{h=-(N-\tau-1)\\ h\neq 0}}^{N-\tau-1}(1-\frac{\tau+|h|}{N})[\gamma_X(\tau)^2] = \gamma_X(\tau)^2\left[\frac{(N-\tau)(N-\tau-1)}{N}\right]
$$

Thus :

$$
\begin{aligned}
\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{n=0}^{N-\tau-1}X_nX_{n+\tau}\right)^2\right] &= \frac{2}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{p=n+1}^{N-\tau-1}\mathbb{E}\left[X_nX_{n+\tau}X_pX_{p+\tau}\right]+\frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}\left[(X_nX_{n+\tau})^2\right]\\
&= \frac{N-\tau}{N^2}\left(2\gamma_X(\tau)^2+\gamma_X(0)^2\right)+\gamma_X(\tau)^2\left[\frac{(N-\tau)(N-\tau-1)}{N^2}\right]\\
&+\frac{1}{N}\sum_{\substack{h=-(N+\tau-1)\\ h\neq 0}}^{N-\tau-1}(1-\frac{\tau+|h|}{N})[\gamma_X(h)^2+\gamma_X(h-\tau)\gamma_X(h+\tau)]\\
&= \frac{N-\tau}{N^2}\left(\gamma_X(\tau)^2+\gamma_X(0)^2\right)+\gamma_X(\tau)^2\left[\frac{(N-\tau)^2}{N^2}\right]\\
&+\frac{1}{N}\sum_{\substack{h=-(N+\tau-1)\\ h\neq 0}}^{N-\tau-1}(1-\frac{\tau+|h|}{N})[\gamma_X(h)^2+\gamma_X(h-\tau)\gamma_X(h+\tau)]
\end{aligned}
$$

Finally :

$$
\begin{aligned}
var(\hat{\gamma}(\tau)) &= -\left(\frac{N-\tau}{N}\gamma(\tau)\right)^2 + \left(\frac{N-\tau}{N}\gamma(\tau)\right)^2 + \frac{N-\tau}{N^2}\left(\gamma_X(\tau)^2 + \gamma_X(0)^2\right) \\
&\quad + \frac{1}{N}\sum_{\substack{h=-(N-\tau-1)\\ h\neq 0}}^{N-\tau-1}(1-\frac{\tau+|h|}{N})[\gamma_X(h)^2 + \gamma_X(h-\tau)\gamma_X(h+\tau)] \\
&= \frac{1}{N}\sum_{h=-(N+\tau-1)}^{N-\tau-1}\left(1-\frac{\tau+|h|}{N}\right)[\gamma_X(h)^2 + \gamma_X(h-\tau)\gamma_X(h+\tau)]
\end{aligned}
$$

- Let's show that $\hat{\gamma}(\tau)$ is consistent

Since the autocorrelation is square integrable we have : $\sum_{h\in\mathbb{Z}} |\gamma_X(h)|^2 < +\infty$ then we get :

$$
\lim_{N\to+\infty} var(\hat{\gamma}(\tau)) = \lim_{N\to+\infty}\frac{1}{N}\sum_{h=-(N-\tau-1)}^{N-\tau-1}\left(1-\frac{\tau+|h|}{N}\right)[\gamma_X(h)^2 + \gamma_X(h-\tau)\gamma_X(h+\tau)] = 0
$$

According to the Bienaymé-Chebyshev theorem :

$$
\mathbb{P}\left(\left|\hat{\gamma}(\tau) - \frac{N-\tau}{N}\gamma(\tau)\right| \geq 2\epsilon\right) \leq \frac{var(\hat{\gamma}(\tau))}{4\epsilon^2}
$$

Let $N_1 > 0$ such that for N $\geq N_1$ $\left|\frac{\tau\gamma(\tau)}{N}\right| < \epsilon$.

Then $\left|\hat{\gamma}(\tau) - \frac{N-\tau}{N}\gamma(\tau)\right| \leq |\hat{\gamma}(\tau) - \gamma(\tau)| + \left|\frac{\tau\gamma(\tau)}{N}\right| \leq |\hat{\gamma}(\tau) - \gamma(\tau)| + \epsilon$

Hence from (7): $\hat{\gamma}(\tau) \xrightarrow{\mathbb{P}} \gamma(\tau)$ as :

$$
\mathbb{P}\left(|\hat{\gamma}(\tau) - \gamma(\tau)| \geq \epsilon\right) \leq \frac{var(\hat{\gamma}(\tau))}{4\epsilon^2} \xrightarrow[N\to+\infty]{} 0
$$

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

- Let's derive the mean and variance or $A(f)$ and $B(f)$

X is a white noise Gaussian, i.e. it has zero mean and variance σ^2.

$$\mathbb{E}[A(f_k)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n] \cos(-2\pi f_k n / f_s) = 0$$
$$\mathbb{E}[B(f_k)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n] \sin(-2\pi f_k n / f_s) = 0$$

if k = 0 then var[A(f)] = Nσ^2. Let's $k > 0$:

$$\begin{aligned}
var[A(f)] &= var\left[\sum_{n=0}^{N-1} X_n \cos(-2\pi kn/N)\right] \\
&= \sum_{n=0}^{N-1} var[X_n] \left[\cos(-2\pi kn/N)\right]^2 \quad \text{(by independance)} \\
&= \sum_{n=0}^{N-1} \sigma^2 \left[\cos(-2\pi kn/N)\right]^2 \\
&= \sum_{n=0}^{N-1} \sigma^2 \left[\frac{cos(-4\pi kn/N)+1}{2}\right] \\
&= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} cos(-4\pi kn/N) + 1
\end{aligned}$$

if k = 0 then var[B(f)] = 0. Let's $k > 0$:

$$\begin{aligned}
var[B(f)] &= var\left[\sum_{n=0}^{N-1} X_n \sin(-2\pi kn/N)\right] \\
&= \sum_{n=0}^{N-1} var[X_n]\left[\sin(-2\pi kn/N)\right]^2 \quad \text{(by independance)} \\
&= \sum_{n=0}^{N-1} \sigma^2 \left[\sin(-2\pi kn/N)\right]^2 \\
&= \sum_{n=0}^{N-1} \sigma^2 \left[\frac{1 - cos(-4\pi kn/N)}{2}\right] \\
&= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} 1 - cos(-4\pi kn/N)
\end{aligned}$$

We know that the sum of the N/2-th roots of unity is zero (i.e. $\sum_{n=0}^{N-1} e^{\frac{2\pi ik}{N/2}} = 0$). In particular, its real part is zero.

$$var[A(f)] = var[B(f)] = \frac{\sigma^2 N}{2}$$

- Let's derive the distribution of the periodogram

First, X is a Gaussian white noise. Thus, $A(f)$ and $B(f)$ are linear combinations of independent Gaussian random variables, so $A(f)$ and $B(f)$ are Gaussian random variables.

$$\begin{aligned}
J(f_k) &= \frac{1}{\sqrt{N}}\left(A(f_k) + iB(f_k)\right) \\
|J(f_k)|^2 &= \frac{1}{\sqrt{N}}\left(A(f_k)^2 + B(f_k)^2\right)
\end{aligned}$$

Now we can show that $A(f)$ and $B(f)$ are independent. Since they are Gausian variables, we can show that they are independent if and only if they have zero covariance (we use the same argument for roots of unity):

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

Then :

$$
\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\left(A(f_k)^2 + B(f_k)^2\right) \\
&= \left(\frac{A(f_k)}{\sqrt{N}}\right)^2 + \left(\frac{B(f_k)}{\sqrt{N}}\right)^2 \\
\frac{2}{\sigma^2}|J(f_k)|^2 &= \frac{2}{N\sigma^2}\left(A(f_k)^2 + B(f_k)^2\right) \\
&= \left(\frac{A(f_k)}{\sqrt{\frac{\sigma^2 N}{2}}}\right)^2 + \left(\frac{B(f_k)}{\sqrt{\frac{\sigma^2 N}{2}}}\right)^2
\end{aligned}
$$

$\frac{2}{\sigma^2}|J(f_k)|^2$ is a sum of two independent centered and with 1 strandart deviation Gaussian random variables. It thus follows a χ^2 distribution.

$$
\begin{cases}
\frac{2}{\sigma^2}|J(f_k)|^2 \sim \chi^2_2 \quad if \quad k > 1 \\
\frac{2}{\sigma^2}|J(f_k)|^2 \sim \chi^2_1 \quad if \quad k = 0 \qquad (B_k = 0)
\end{cases}
$$

- Let's derive the variance of $|J(f_k)|^2$

The variance of a χ^2_p is 2p.

$$
\begin{cases}
var(|J(f_k)|^2) = \sigma^4 \quad if \quad k > 1 \\
var(|J(f_k)|^2) = \frac{\sigma^4}{2} \quad if \quad k = 0
\end{cases}
$$

The variance of the estimator is constant with respect to N, so it does not tend towards 0. Thus, the estimator is not consistent.

- Explanation of the erratic behavior of the periodogram

$$
\begin{aligned}
\text{if } k \neq p: \ & \operatorname{cov}(|J(f_k)|^2, |J(f_p)|^2) \\
& \propto \operatorname{cov}(A^2(f_k) + B^2(f_k), A^2(f_p) + B^2(f_p)) \\
& \propto \operatorname{cov}(A^2(f_k), A^2(f_p)) + \operatorname{cov}(A^2(f_k), B^2(f_p)) + \operatorname{cov}(B^2(f_k), A^2(f_p)) + \operatorname{cov}(B^2(f_k), B^2(f_p)) \\
& \propto 0 \qquad \text{As for} \quad k \neq p: \quad A(f_k) \perp\!\!\!\perp B(f_p) \text{ and } A(f_k) \perp\!\!\!\perp A(f_p) \text{ and } B(f_k) \perp\!\!\!\perp B(f_p)
\end{aligned}
$$

So there is no correlation between all the frequencies. This explains why the periodogram behaves erratically and is not at all smooth.

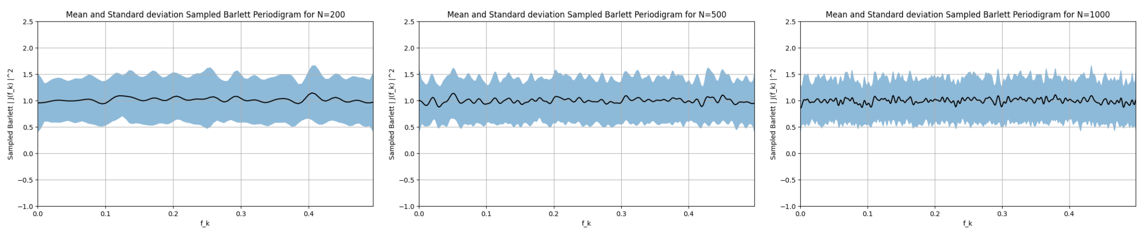
Question 9

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).

Using Bartlett's estimate, the standard deviation is approximately halved (so the variance by ≈ 5 as expected). However, again the standard deviation does not decrease with an increasing number of samples.

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

The KKN method only works with vectors of the same dimension. KNN can also be given a previously calculated distance matrix. Here, we have signals of varying size and so we can't use KNN directly. We will therefore use DTW as the distance to create the distance matrix and thus solve this problem, as DTW can handle vectors of different sizes.

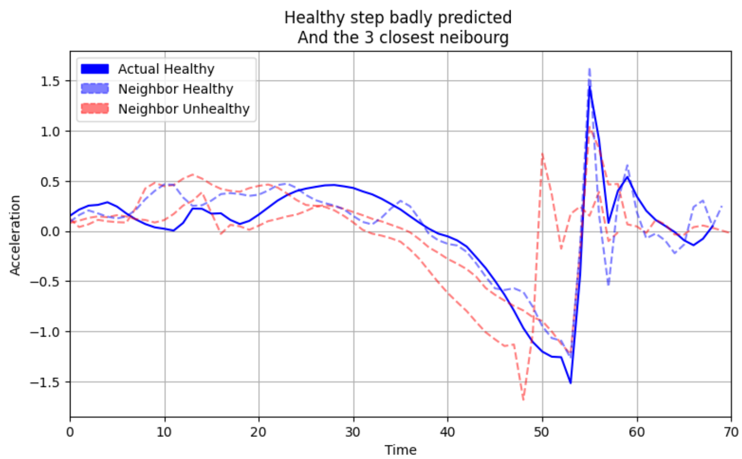
$n_{neighbors} = 3$
For the train set $FScore = 0.867$, $Accuracy = 0.875$, $Recall = 0.863$
For the test set $FScore = 0.482$, $Accuracy = 0.341$, $Recall = 0.370$
The F1-Score of the model for the test set is significantly lower compared to the training set. This discrepancy indicates that the model is not generalizing well to new, unseen data. After inspecting the data, it appears that certain signal shapes are present only in the test set but not in the training set, leading to this poor performance.

Question 11

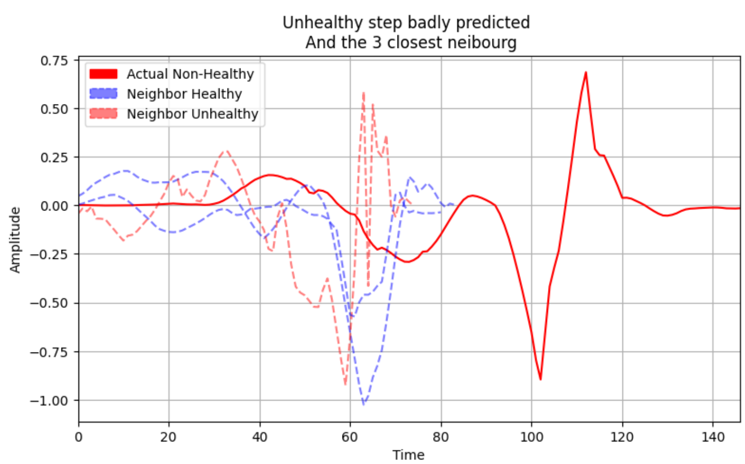
Display on Figure 3 a badly classified step from each class (healthy/non-healthy).

Answer 11

Here are two signals, healthy and non-healthy. To understand the decision, we have represented the misclassified signal and its K=3 nearest neighbours.



Badly classified healthy step



Badly classified non-healthy step

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

Note: for the second signal, its nearest neighbours are not the same size and this shows how DTW can be used to associate signals with perturbations (even if this is not conclusive here)..