Assignment 1 (ML for TS) - MVA 2022/2023

Théo Di Piazza theo.dipiazza@gmail.com
Stanislas du Ché stanislasduche@gmail.com

February 1, 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.

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

To start, let's use the first order condition on (1). It comes that:

$$-X^T(y - X\beta) + \lambda sign(\beta) = 0$$

Where $\beta \in \mathbb{R}^p$ such that $\forall\ i \in \{1, .., p\} : sign(\beta)_i \in \begin{cases} [-1, 1] & \text{if } \beta_i = 0 \\ \{1\} & \text{if } \beta_i > 0 \\ \{-1\} & \text{if } \beta_i < 0 \end{cases}$

We're looking to show that there exists λ_{max} such that the minimizer of (1) defined as β_{min} is 0_p for any $\lambda > \lambda_{max}$. Hence for $\beta_{min} = 0_p$, it comes that :

$$\lambda sign(\beta_{min}) = X^T y$$

Moreover, it can be written that ($\lambda > 0$):

$$\lambda ||sign(\beta_{min})||_\infty = ||X^T y||_\infty$$

Then if $\forall\ i\ \in \{1, .., p\} : sign(\beta_{min})_i \notin \{-1, 1\} \iff ||sign(\beta_{min})||_\infty < 1$, we can write that :

$$||sign(\beta_{min})||_\infty < 1 \iff \lambda ||sign(\beta_{min})||_\infty < \lambda \iff ||X^T y||_\infty < \lambda$$

Reciprocity : We set $\lambda_{max} = ||X^T y||_\infty$. Suppose $\lambda > \lambda_{max}$ and β minimiser of (1) different to 0_p. We have:

$$\begin{aligned}
&\lambda sign(\beta) = X^T(y - X\beta) \\
\Rightarrow &\lambda \beta^T sign(\beta) = \beta^T X^T (y - X\beta) \\
\Rightarrow &\lambda = \frac{\beta^T}{||\beta||_1} X^T Y - \beta^T X^T X \frac{\beta}{||\beta||_1} \text{ and we have } \frac{\beta^T}{||\beta||_1} \leq 1_p \text{ and } X^T X \in S_n^+ \\
\Rightarrow &\lambda < \lambda_{max} \text{ Contradiction !}
\end{aligned}$$

Finally : $\boxed{\forall \lambda > \lambda_{max} = ||X^T y||_\infty \text{, we have that } \beta_{min} = 0_p}$

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

To show that the sparse coding problem is a lasso regression, it's first needed to recall that :

- Dictionary patern : $d_1, ..., d_K$ such that $d_k\ \in \mathbb{R}^L$ where L is the size of dictionary.
- Activation signals : $z_1, ..., z_K$ such that $z_k\ \in \mathbb{R}^{N-L+1}$.

The main issue is to be able to write the convolution problem as a matrix product such that

$$DZ = \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k$$

To do this, let's define Z and D such that :

$$Z = \begin{bmatrix} z_1 \\ \dots \\ z_K \end{bmatrix} \text{ where } Z \in \mathbb{R}^{K.N} \text{ and } z_k(i) = \begin{cases} (z_k)_i & \text{if } 0 \leq i \leq N-L+1 \\ 0 & \text{otherwise} \end{cases}$$

$$D = \begin{bmatrix} \tilde{d}_1 & \dots & \tilde{d}_K \end{bmatrix} \text{ where } D \in \mathbb{R}^{N.N.K} \text{ such that } (\tilde{d}_k)_{i,j} = (d_k)_{i-j}$$

$$\text{and } (\tilde{d}_k)_{i,j} = \begin{cases} d_k(i-j) & \text{if } 0 \leq i-j \leq L \\ 0 & \text{otherwise} \end{cases}$$

Then, it comes that :

$$(DZ)_i = (\sum_{k=1}^{K} D_k Z_k)_i = (\sum_{k=1}^{K} \tilde{d}_k z_k)_i = \sum_{k=1}^{K} (\tilde{d}_k z_k)_i = \sum_{k=1}^{K} \sum_{j=1}^{N} (\tilde{d}_k)_{i,j} (z_k)_j = \sum_{k=1}^{K} \sum_{j=1}^{N} d_k(i-j) z_k(j)$$

Hence,

$$(DZ)_i = (\sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k)_i \text{ so } \boxed{\mathrm{DZ} = \textstyle\sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k}$$

About the second part of the objective function, wich have that $\sum_{k=1}^{K} ||z||_1 = ||Z||_1$.

Hence, the problem can be written such that :

$$\min_{Z} \frac{1}{2} ||x - DZ||_2^2 + \lambda ||Z||_1$$

Finally with Queqtion 1, for a fixed dictionary there exists λ_{max} such that the sparse codes are only 0 for any $\lambda > \lambda_{max}$ such that :

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

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 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.

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

In the case of a Gaussian white noise, samples are supposed independant such that $X_n \sim \mathcal{N}(0, \sigma^2)$. About the autocovariance function, it comes that :

$$\mathbb{E}\left[X_n X_{n+\tau}\right] = \begin{cases} \mathbb{E}\left[X_n\right]\mathbb{E}\left[X_{n+\tau}\right] = 0 & \text{if } \tau \neq 0 \\ \mathbb{E}\left[X_n\right]\mathbb{E}\left[X_n\right] = Var(X_n) = \sigma^2 & \text{if } \tau = 0 \end{cases}$$

Then for the power spectrum, it comes that :

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

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

- 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

3 cases must be distinguished : case with $\tau < 0$, $\tau > 0$ and $\tau = 0$.

Case with $\tau > 0$ It comes that :

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

Case with $\tau < 0$ It comes that :

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

Hence, $\boxed{\hat{\gamma}(\tau) \text{ is a biased estimator of } \gamma(\tau).}$

Also, it comes that $\lim_{N\to\infty} \frac{N-\tau}{N}\gamma(\tau) = \gamma(\tau)$

Hence, $\boxed{\hat{\gamma}(\tau) \text{ is an asymptotically unbiased estimator of } \gamma(\tau).}$

Remark : We notice that for $\tau = 0$, case of independency, we recover the results of Q.3

<u>Case with $\tau = 0$</u> It comes that :

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

Finally, a way to de-bias this estimator would be to use a new estimator noted $\hat{\gamma}_2(\tau)$ such that :

$$\boxed{\hat{\gamma}_2(\tau) = \frac{1}{N-\tau}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}}$$

Since $\forall\tau$, we would have that :

$$\mathbb{E}\left[\hat{\gamma}_2(\tau)\right] = \mathbb{E}\left[\frac{1}{N-\tau}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right] = \frac{1}{N-\tau}\sum_{n=0}^{N-\tau-1} \mathbb{E}\left[X_n X_{n+\tau}\right] = \frac{N-\tau}{N-\tau}\gamma(\tau) = \gamma(\tau)$$

Question 5

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

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

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

First solution with conjugate of complex

Suppose $k \in [\![1; N]\!]$.

$$
\begin{aligned}
|J(f_k)|^2 &= J(f_k) \times J(\bar{f_k}) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{-2i\pi \frac{kn}{N}} \sum_{m=0}^{N-1} X_m e^{2i\pi \frac{km}{N}} \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_m X_n e^{-2i\pi \frac{k(n-m)}{N}}. \text{ On pose : } h = n - m \text{Cela nous fait } n \longleftarrow h + m \\
&= \frac{1}{N} \sum_{h=-N+1}^{N-1} \sum_{m=0}^{N-1-h} X_m X_{h+m} e^{-2i\pi \frac{kh}{N}} \\
&= \sum_{h=-N+1}^{N-1} e^{-2i\pi \frac{kh}{N}} \gamma(\hat{h})
\end{aligned}
$$

With Euler, we finally have the result :

$$
\boxed{|J(f_k)|^2 = \sum_{h=-N+1}^{N-1} e^{-2i\pi \frac{kh}{N}} \hat{\gamma}(h)} = \hat{\gamma}(0) + 2 \sum_{h=1}^{N-1} \hat{\gamma}(h) cos(\frac{2\pi kh}{N})
$$

We note k the index of $f^{(N)}$.
$\mathbb{E}\left[|J(f_k)|^2\right] = \sum_{h=-N+1}^{N-1} e^{-2i\pi \frac{hf_k}{f_s}} \mathbb{E}[\hat{\gamma}(h)]$

- We have $f_k = \frac{kf_s}{N}$; so $||f - f^{(N)}||^2 \underset{N \longrightarrow \infty}{\longrightarrow} 0. \Rightarrow e^{-2i\pi \frac{hf_k}{f_s}} \underset{N \longrightarrow \infty}{\longrightarrow} 1$
- $\mathbb{E}[\hat{\gamma}(h)] \underset{N \longrightarrow \infty}{\longrightarrow} \gamma(h)$
- $\sum_{h=-N+1}^{N-1} |\mathbb{E}[\hat{\gamma}(h)]| <= \sum_{h=-N+1}^{N-1} |\gamma(h)|$ by majoring term by term. And $\sum_{h=-N+1}^{N-1} |\gamma(h)|$ converges by hypothesis.

So by the theorem of dominated convergence (applied in inversion Series and Limits), we finally have that $\boxed{|J(f^{(N)})|^2 \text{ is an asymptotically unbiased estimator of } S(f) \text{ for } f > 0}$.

Second solution with norm of complex (same result)

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

$$
\Longrightarrow |J(f_k)|^2 = \frac{1}{N} \left[(\sum_{n=0}^{N-1} X_n cos(\frac{2\pi kn}{N}))^2 + (\sum_{n=0}^{N-1} X_n sin(\frac{2\pi kn}{N}))^2 \right]
$$

$$= \frac{1}{N}[\sum_{n=0}^{N-1} X_n^2 cos(\frac{2\pi kn}{N})^2 + \sum_{n=0}^{N-1} X_n^2 sin(\frac{2\pi kn}{N})^2 + 2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n cos(\frac{2\pi kn}{N}) X_m cos(\frac{2\pi km}{N})$$

$$+2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n sin(\frac{2\pi kn}{N}) X_m sin(\frac{2\pi km}{N})]$$

$$= \frac{1}{N}\left[\sum_{n=0}^{N-1} X_n^2 + 2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n X_m (cos(\frac{2\pi kn}{N})cos(\frac{2\pi km}{N}) + sin(\frac{2\pi kn}{N})sin(\frac{2\pi km}{N}))\right]$$

$$= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n X_m cos(\frac{2\pi k(n-m)}{N})$$

$$= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{\tau=1}^{N-n-1} X_n X_{n+\tau} cos(\frac{2\pi k\tau}{N}) = \boxed{\hat{\gamma}(0) + 2\sum_{\tau=1}^{N-1} \hat{\gamma}(\tau) cos(\frac{2\pi k\tau}{N})}$$

Finally with what we showed, it comes that :

$$\lim_{N\to\infty} |J(f^{(N)}|^2 - S(f) = 0$$

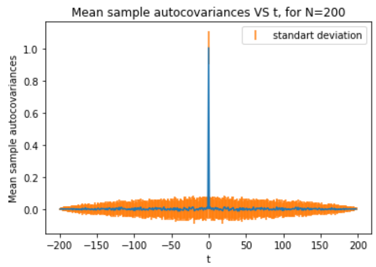
Hence $\boxed{|J(f^{(N)}|^2 \text{ is an asymptotically unbiased estimator of } S(f) \text{ for } f > 0.}$

Question 6

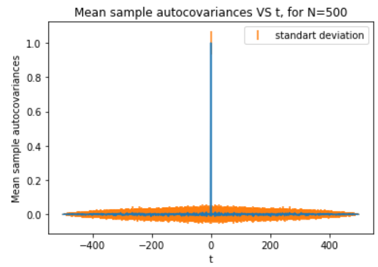
In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

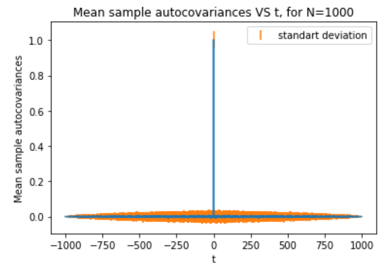
Add your plots to Figure 1.



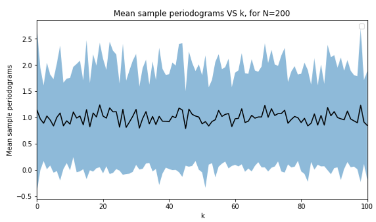
Autocovariance ($N = 200$)



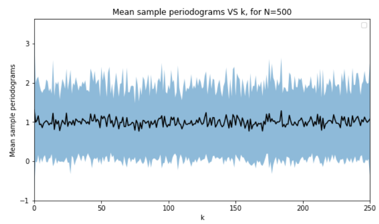
Autocovariance ($N = 500$)



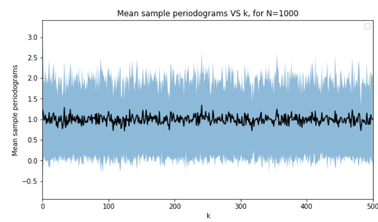
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

Observations on the Autocovariance : We notice first of all that $\tau = 0$, the average value of the sample covariance is equal to the standart deviation and that for $\tau \neq 0$, the average value of the sample covariance is very close to 0. Moreover by observing the standart-deviations, we notice that the more the number of generations is big, the weaker the standart deviation is and that the average value of the sample covariance seems to be tending to 0 for $\tau \neq 0$. This result is coherent because we have shown that $\hat{\gamma}(\tau)$ is an asymptotically unbiased estimator of $\gamma(\tau)$.

Observations on the Periodogram : We observe an unstable or even random behavior that does not show any coherent trend. In summary, the behavior of the periodogram is erratic. Moreover, the variance does not seem to decrease when the number of samples increases.

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 $\mathbb{E}[Y_1Y_2Y_3Y_4] = \mathbb{E}[Y_1Y_2]\mathbb{E}[Y_3Y_4] + \mathbb{E}[Y_1Y_3]\mathbb{E}[Y_2Y_4] + \mathbb{E}[Y_1Y_4]\mathbb{E}[Y_2Y_3]$.)

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

Answer 7

To begin with, remember that :

$$\begin{aligned} Var(\hat{\gamma}(\tau)) &= \mathbb{E}(\hat{\gamma}(\tau) - \mathbb{E}\hat{\gamma}(\tau))^2 \\ &= \mathbb{E}(\hat{\gamma}(\tau)^2) - \mathbb{E}(\hat{\gamma}(\tau))^2 \text{ and from Q4, } \mathbb{E}(\hat{\gamma}(\tau)) = \frac{N-\tau}{N}\gamma(\tau) \end{aligned}$$

It comes that :

$$\mathbb{E}(\hat{\gamma}(\tau)^2) = \frac{1}{N^2}\mathbb{E}\left(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right)^2 = \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\mathbb{E}[(X_n X_{n+\tau})^2] + \frac{1}{N^2}\sum_{0\le i<j\le N-\tau-1}\mathbb{E}(X_i X_{i+\tau} X_j X_{j+\tau})$$

Thanks to the indication given, we have:

- $\mathbb{E}[(X_n X_{n+\tau})^2] = \mathbb{E}[X_n X_{n+\tau}]^2 + \mathbb{E}[X_n^2]\mathbb{E}[X_{n+\tau}^2] + \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_n X_{n+\tau}]$
  $\Rightarrow \mathbb{E}[(X_n X_{n+\tau})^2] = \gamma(\tau)^2 + \gamma(0)^2 + \gamma(\tau)^2$
- $\mathbb{E}(X_i X_{i+\tau} X_j X_{j+\tau}) = \mathbb{E}[X_i X_{i+\tau}]\mathbb{E}[X_j X_{j+\tau}] + \mathbb{E}[X_i X_j]\mathbb{E}[X_{i+\tau} X_{j+\tau}] + \mathbb{E}[X_i X_{j+\tau}]\mathbb{E}[X_j X_{i+\tau}]$
  $\Rightarrow \mathbb{E}(X_i X_{i+\tau} X_j X_{j+\tau}) = \gamma(\tau)^2 + \gamma(j-i)^2 + \gamma(j-i+\tau)\gamma(j-i-\tau)$

We realise the variable change $n = j - i$ in the second sum and we finally have :

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

and

$$\begin{aligned} \frac{1}{N^2}\sum_{0\le i<j\le N-\tau-1}\mathbb{E}(X_i X_{i+\tau} X_j X_{j+\tau}) &= \frac{2}{N^2}\sum_{n=1}^{N-\tau-1}(N-\tau-n)\left(\gamma(\tau)^2 + \gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau)\right) \\ &= \frac{1}{N}\sum_{n=-(N-\tau-1),n\neq 0}^{N-\tau-1}(1 - \frac{\tau+|n|}{N})\left(\gamma(\tau)^2 + \gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau)\right) \end{aligned}$$

It comes that:

$$\mathbb{E}(\hat{\gamma}(\tau)^2) = \frac{1}{N}\sum_{n=-(N-\tau-1)}^{N-\tau-1}\left(1 - \frac{\tau+|n|}{N}\right)[\gamma(n)^2 + \gamma(n-\tau)\gamma(n+\tau)] + \frac{(N-\tau)^2}{N^2}\gamma(\tau)^2$$

Hence we get that : $\boxed{\text{var}(\hat{\gamma}(\tau)) = (1/N)\sum_{n=-(N-\tau-1)}^{n=N-\tau-1}\left(1 - \frac{\tau+|n|}{N}\right)\left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right].}$

Finally, as $\sum_{\tau\in\mathbb{Z}}|\gamma(\tau)|^2 < \infty$ and $(1 - \frac{\tau+|n|}{N})$ bounded for every n, with Tchebychev it comes that it is consistent since :

$$var(\hat{\gamma}(\tau)) \leq \frac{2}{N} \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \gamma(n)^2 \xrightarrow[N\to\infty]{} 0$$

$$\implies \lim_{N\to\infty} var(\hat{\gamma}(\tau)) = 0$$

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

- Mean and Variance of $A(f)$

$$\mathbb{E}[A(f)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n]\, cos(\frac{-2\pi kn}{N}) = 0$$

$$var(A(f)) = \sum_{n=0}^{N-1} var(X_n) cos(\frac{-2\pi kn}{N})^2 = \sum_{n=0}^{N-1} \sigma^2 cos(\frac{-2\pi kn}{N})^2$$

  Mean and Variance of $B(f)$

$$\mathbb{E}[B(f)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n]\, sin(\frac{-2\pi kn}{N}) = 0$$

$$var(B(f)) = \sum_{n=0}^{N-1} var(X_n) sin(\frac{-2\pi kn}{N})^2 = \sum_{n=0}^{N-1} \sigma^2 sin(\frac{-2\pi kn}{N})^2$$

  And we have :

$$\sum_{n=0}^{N-1} sin(\frac{-2\pi kn}{N})^2 + \sum_{n=0}^{N-1} cos(\frac{-2\pi kn}{N})^2 = N$$

$$\sum_{n=0}^{N-1} sin(\frac{-2\pi kn}{N})^2 - \sum_{n=0}^{N-1} cos(\frac{-2\pi kn}{N})^2 = \sum_{n=0}^{N-1} cos(2\frac{-2\pi kn}{N})$$

$$\sum_{n=0}^{N-1} cos(2\frac{-2\pi kn}{N}) = \mathcal{R}\left(\sum_{n=0}^{N-1} exp(\frac{-4i\pi kn}{N})\right) = 0$$

Finally, we have $Var(A(f)) = Var(B(f)) = \sigma^2 \frac{N}{2}$

- Distribution of the periodogram values

$$J(f) = \frac{1}{\sqrt{N}}(A(f) + iB(f)) \Longrightarrow |J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2) = (\frac{A(f_k)}{\sqrt{N}})^2 + (\frac{B(f_k)}{\sqrt{N}})^2$$

As linear combination of gaussians, $A(f)$ and $B(f)$ are also gaussian, and $J(f)$ is the sum of their projection in two orthogonal spaces : $\mathbb{R}$ and $i\mathbb{R}$.

So by Cochran Theorem, $\frac{|A(f_k)|^2}{Var(A(f_k))} \sim \chi^2(1)$ and $\frac{|B(f_k)|^2}{Var(B(f_k))} \sim \chi^2(1)$ and are independent.

Thus, $|J(f_k)|^2 \sim \frac{\sigma^2}{2}\left(\chi^2(1) + \chi^2(1)\right)$ and $\chi^2(1) + \chi^2(1) \sim \chi^2(2)$.

So finally we have, $|J(f_k)|^2 \sim \frac{\sigma^2}{2}\chi^2(2)$.

- Variance of the $|J(f_k)|^2$

From $|J(f_k)|^2 \sim \frac{\sigma^2}{2}\chi^2(2)$, the variance of $|J(f_k)|^2$ is σ^4.
The variance is independant of N, hence $\boxed{\text{the periodogram is not consistent}}$.

- Covariance between the $|J(f_k)|^2$

Recall that $|J(f_k)|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m e^{\frac{-2i\pi k(n-m)}{N}}$. It comes that :

Let's assume that $k \neq l$.

$$\begin{aligned} cov(|J(f_k)|^2, |J(f_l)|^2) &= \mathbb{E}\left[(|J(f_k)|^2 - \mathbb{E}[|J(f_k)|^2])(|J(f_l)|^2 - \mathbb{E}[|J(f_l)|^2])\right] \\ &= \mathbb{E}\left[|J(f_k)|^2 |J(f_l)|^2\right] - \sigma^4 \end{aligned}$$

Let's detail for $\mathbb{E}\left[|J(f_k)|^2|J(f_l)|^2\right]$:

$$\mathbb{E}\left[|J(f_k)|^2|J(f_l)|^2\right] = \mathbb{E}\left[(\frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m e^{\frac{-2i\pi k(n-m)}{N}})(\frac{1}{N}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} X_p X_q e^{\frac{-2i\pi l(p-q)}{N}})\right]$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\mathbb{E}[X_n X_m X_p X_q] e^{\frac{-2i\pi(k(n-m)+l(p-q))}{N}}$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}(\mathbb{E}[X_n X_m]\mathbb{E}[X_p X_q] + \mathbb{E}[X_n X_p]\mathbb{E}[X_m X_q] + \mathbb{E}[X_n X_q]\mathbb{E}[X_m X_p]) e^{\frac{-2i\pi(k(n-m)+l(p-q))}{N}}$$

$$= \frac{1}{N^2}(\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] e^{\frac{-2i\pi(k(n-n)+l(m-m))}{N}} + \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] e^{\frac{-2i\pi(k(n-m)+l(n-m))}{N}}$$

$$+\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2]e^{\frac{-2i\pi(k(n-m)+l(m-n))}{N}})$$

$$=\frac{1}{N^2}(N^2\sigma^4+\sigma^4\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}e^{\frac{-2i\pi}{N}(n(k+l)+m(-k-l))}+\sigma^4\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}e^{\frac{-2i\pi}{N}(n(k-l)+m(l-k))})$$

$$=\sigma^4+\frac{1}{N^2}(\sigma^4\sum_{n=0}^{N-1}e^{\frac{-2i\pi}{N}n(k+l)}\sum_{m=0}^{N-1}e^{\frac{-2i\pi}{N}m(-k-l)}+\sigma^4\sum_{n=0}^{N-1}e^{\frac{-2i\pi}{N}n(k-l)}\sum_{m=0}^{N-1}e^{\frac{-2i\pi}{N}m(l-k)})$$

$$=\sigma^4+\frac{1}{N^2}(\sigma^4\frac{1-e^{-2i\pi(k+l)}}{1-e^{\frac{-2i\pi}{N}(k+l)}}\frac{1-e^{-2i\pi(-k-l)}}{1-e^{\frac{-2i\pi}{N}(-k-l)}}+\sigma^4\frac{1-e^{-2i\pi(k-l)}}{1-e^{\frac{-2i\pi}{N}(k-l)}}\frac{1-e^{-2i\pi(l-k)}}{1-e^{\frac{-2i\pi}{N}(l-k)}})=\sigma^4$$

Finally it comes that :

$$\text{For } k\neq l:\ \boxed{\mathrm{cov}(|\mathrm{J}(\mathrm{f}_k)|^2,|J(f_l)|^2)=\sigma^4-\sigma^4=0}$$

$$\text{For } k=l:\ \boxed{\mathrm{cov}(|\mathrm{J}(\mathrm{f}_k)|^2,|J(f_l)|^2)=\sigma^4}$$

For $k\neq l$, the covariance is zero. This explains the erratic behavior observed in question 6.

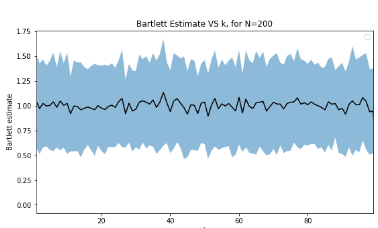
Question 9

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.

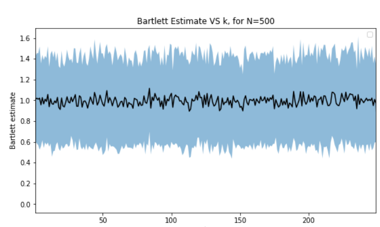
- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K=5$). What do you observe.

Add your plots to Figure 2.

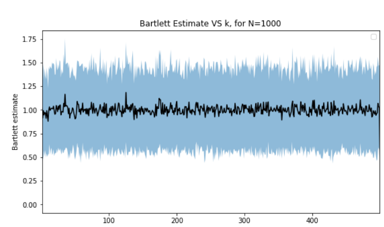
Answer 9



Periodogram ($N=200$)



Periodogram ($N=500$)



Periodogram ($N=1000$)

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

Answer 9

We can see that the standard deviation obtained is much smaller than the one obtained in question 6. For question 6 we obtain a standard deviation about equal to 1. Here with Barlett's estimate, we obtain a standard deviation about equal to 0.44, that is divided by a factor $\frac{1}{0.44} \approx 2.2 \approx \sqrt{5}$. This result is consistent with the statement, this procedure has the effect of dividing the variance by $K = 5$, thus dividing the standard deviation by $\sqrt{K} = \sqrt{5}$.

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

Training procedure The obtained signals have different lengths. However, the KNeighborsClassifier method of sklearn only takes into account inputs of the same size to compute the distance, or square distance matrices (**KNeighborsClassifier documentation.**). Thus, we propose to modify the size of the signals by adjusting them all to the same length: the length of the longest signal. This is done by adding the value -10 to the signals until the desired value is obtained. As we are working with the DTW, this distance is able to take into account the change in variation and compute the "true DTW" despite the change in length of the signals.

Training results The KNN is trained with a fine-tuning of the hyper-parameter k which corresponds to the number of neighbors taken into account when computing the distances. The training is performed with 5-fold cross-validation and the optimal number of neighbors found is 5.

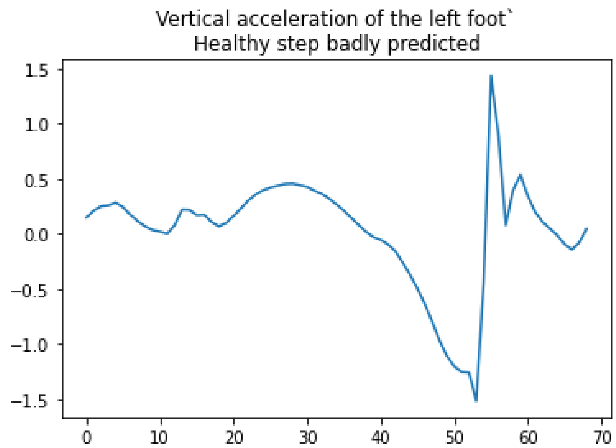
For label 1, the F-score obtained is equal to 0.87 for the training sample and 0.45 for the test sample.

We notice an over-fitting on the training sample because the performances are much better on the train sample than on the test sample. This may be related to the fact that the kind of steps of a part of the test sample have never been seen in the train sample by the model.

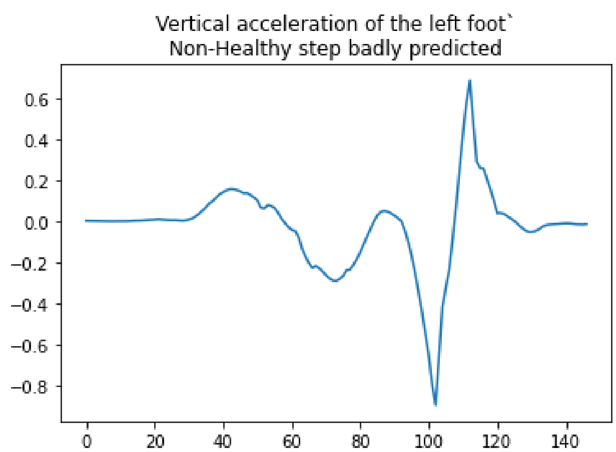
Question 11

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).