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

Bastien LE CHENADEC bastien.le-chenadec@eleves.enpc.fr
Sofiane EZZEHI sofiane.ezzehi@eleves.enpc.fr

January 10, 2024

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

$$\lambda_{\max} = \dots \tag{2}$$

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:

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

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 us consider a fixed dictionary. The variables to be optimized are the K activations signals $\mathbf{z}_1, \dots, \mathbf{z}_K$. Thus it is intuitive to define the vector of regressors β as the concatenation of the activation signals :

$$\beta = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_K \end{bmatrix} \in \mathbb{R}^{(n-L+1)K}$$

The response vector is the signal itself :

$$y = \mathbf{x}$$

And the design matrix is the concatenation of the shifted dictionary atoms such that an activation of an atom at time t leads to a contribution of this atom from time t to time $t + L - 1$ in the signal :

$$X = \begin{bmatrix} \mathbf{d}_1 & 0 & \dots & 0 & \mathbf{d}_2 & 0 & \dots & 0 & \dots & \mathbf{d}_K & 0 & \dots & 0 \\ 0 & \mathbf{d}_1 & \ddots & \vdots & 0 & \mathbf{d}_2 & \ddots & \vdots & \dots & 0 & \mathbf{d}_K & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & 0 & \dots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{d}_1 & 0 & \dots & 0 & \mathbf{d}_2 & \dots & 0 & \dots & 0 & \mathbf{d}_K \end{bmatrix} \in \mathbb{R}^{n \times (n-L+1)K}$$

Where each $\mathbf{d}_i = \begin{bmatrix} d_{i,1} \\ \vdots \\ d_{i,L} \end{bmatrix} \in \mathbb{R}^L$ is a dictionary atom, so that the matrix X is in fact written by blocks.

With these notations it is clear that the problem (3) rewrites as :

$$\min_{\beta \in \mathbb{R}^{(n-L+1)K}} \frac{1}{2} \|y - X\beta\|_2^2 \quad + \quad \lambda \|\beta\|_1$$

which we recognize as a Lasso regression (1).

It follows from the first question that for a fixed dictionary, there exists $\lambda_{\max}$ such that the sparse codes are only 0 for any $\lambda > \lambda_{\max}$ and we have :

$$\lambda_{\max} = \dots \tag{4}$$

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2i\pi f \tau / f_s}. \tag{5}$$

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 $(n = 0, \dots, N-1)$ 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

If X_n is a Gaussian white noise, then X_n is a Gaussian random variable with zero mean and variance σ^2. Furthermore X_n and X_m are independent for $n \neq m$.

Thus for $\tau \in \mathbb{Z}$:

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

Computing the power spectrum is then very straightforward:

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

The power spectrum is constant across all frequencies and equal to the variance of the white noise. That is why it is called *white* by analogy with light because white light is a mixture of all the colors of the spectrum.

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

$$\begin{aligned} b(\hat{\gamma}(\tau)) &= \mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau) \\ &= \frac{1}{N} \sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}] - \gamma(\tau) \\ &= \frac{1}{N} \sum_{n=0}^{N-\tau-1} \gamma(\tau) - \gamma(\tau) && \text{by definition of } \gamma(\tau). \\ &= \frac{-\tau}{N} \gamma(\tau) \end{aligned}$$

Thus $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$. However, we have :

$$\lim_{N \to +\infty} \frac{-\tau}{N} \gamma(\tau) = 0$$

So $\hat{\gamma}(\tau)$ is asymptotically unbiased. From our previous computations we see that we can de-bias like so :

$$\hat{\gamma}'(\tau) = \frac{\tau}{N} X_0 X_\tau + \hat{\gamma}(\tau)$$

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

Let $k \in [\![0, \frac{N}{2}]\!]$. Using the fact that for $z \in \mathbb{C}, |z|^2 = z\bar{z}$, we get :

$$\begin{aligned}
|J(f_k)|^2 &= \left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-\frac{2ik\pi}{N} n} \right) \left(\frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} X_m e^{\frac{2ik\pi}{N} m} \right) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m e^{-\frac{2ik\pi}{N}(n-m)} \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m \cos\left(\frac{2k\pi}{N}(n-m) \right) \qquad \text{(since } |J(f_k)|^2 \text{ is real)}
\end{aligned}$$

By summing over $\tau = n - m$, we get,

$$|J(f_k)|^2 = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \sum_{\substack{n,m=0 \\ n-m=\tau}}^{N-1} X_n X_m \cos\left(\frac{2k\pi}{N} \tau \right)$$

We now have to consider the different possibilites for n and m such that $n - m = \tau$.

1. If $\tau \geq 0$, then the set of possible couples is,

$$\left\{ (n, m) \in [\![0, N-1]\!]^2 \middle| n - m = \tau \right\} = \{(\tau, 0), (\tau + 1, 1), \dots, (N-1, N-1-\tau)\}.$$

2. If $\tau < 0$, then the set of possible couples is,

$$\left\{ (n, m) \in [\![0, N-1]\!]^2 \middle| n - m = \tau \right\} = \{(0, -\tau), (1, 1-\tau), \dots, (N-1+\tau, N-1)\}.$$

We can therefore rewrite $|J(f_k)|^2$ as,

$$|J(f_k)|^2 = \frac{1}{N} \sum_{\tau=0}^{N-1} \sum_{n=\tau}^{N-1} X_n X_{n+\tau} \cos\left(\frac{2k\pi}{N} \tau \right) + \frac{1}{N} \sum_{\tau=1}^{N-1} \sum_{n=0}^{N-1-\tau} X_n X_{n-\tau} \cos\left(\frac{2k\pi}{N} \tau \right).$$

Since, for $\tau = 0$, we have $\frac{1}{N}\sum_{n=0}^{N-1} X_n^2 = \hat{\gamma}(0)$,

we have,

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

Using the fact that $\hat{\gamma}(\tau) = \hat{\gamma}(-\tau)$, we can rewrite $|J(f_k)|^2$ as,

$$\boxed{|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1}\hat{\gamma}(\tau)\cos\left(\frac{2\pi k}{N}\tau\right)}$$

Since γ is symmetric it is clear that $S(f) \in \mathbb{R}$. We can write :

$$S(f) = \sum_{\tau=-\infty}^{+\infty}\gamma(\tau)\cos\left(\frac{2\pi f}{f_s}\tau\right)$$

We also have that $|f - f^{(N)}| < \frac{f_s}{2N}$. Let us consider the bias of $|J(f^{(N)})|^2$ as an estimator of $S(f)$:

$$\begin{aligned}
b(|J(f^{(N)})|^2) &= \mathbb{E}[|J(f^{(N)})|^2] - S(f) \\
&= \mathbb{E}[|J(f^{(N)})|^2] - \mathbb{E}[|J(f)|^2] + \mathbb{E}[|J(f)|^2] - S(f)
\end{aligned}$$

<u>First term :</u>

$$\begin{aligned}
\mathbb{E}[|J(f^{(N)})|^2] - \mathbb{E}[|J(f)|^2] &= \mathbb{E}\left[\sum_{\tau=-(N-1)}^{N-1}\hat{\gamma}(\tau)\cos\left(\frac{2\pi f^{(N)}}{f_s}\tau\right)\right] - \mathbb{E}\left[\sum_{\tau=-(N-1)}^{N-1}\hat{\gamma}(\tau)\cos\left(\frac{2\pi f}{f_s}\tau\right)\right] \\
&= \sum_{\tau=-(N-1)}^{N-1}\mathbb{E}[\hat{\gamma}(\tau)]\left[-2\sin\left(\frac{\pi}{f_s}(f^{(N)}+f)\tau\right)\sin\left(\frac{\pi}{f_s}(f^{(N)}-f)\tau\right)\right]
\end{aligned}$$

Since $|f - f^{(N)}| < \frac{f_s}{2N}$ we can prove that this term goes to 0 as $N \to +\infty$.

<u>Second term :</u>

$$
\begin{aligned}
\mathbb{E}[|J(f)|^2] - S(f) &= \mathbb{E}\left[\sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau)\cos\left(\frac{2\pi f}{f_s}\tau\right)\right] - \sum_{\tau=-\infty}^{+\infty} \gamma(\tau)\cos\left(\frac{2\pi f}{f_s}\tau\right) \\
&= R_N(S) + \sum_{\tau=-(N-1)}^{N-1} \cos\left(\frac{2\pi f}{f_s}\tau\right)\left(\mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau)\right) \\
&= R_N(S) + \sum_{\tau=-(N-1)}^{N-1} \cos\left(\frac{2\pi f}{f_s}\tau\right) b(\hat{\gamma}(\tau))
\end{aligned}
$$

Where $R_N(S)$ is the rest of a finite sum, therefore it goes to 0 as $N \to +\infty$. Furthermore we proved in question 4 that $\hat{\gamma}(\tau)$ is asymptotically unbiased, thus $b(\hat{\gamma}(\tau))$ goes to 0 as $N \to +\infty$. Therefore the whole term goes to 0 as $N \to +\infty$.

Conclusion :

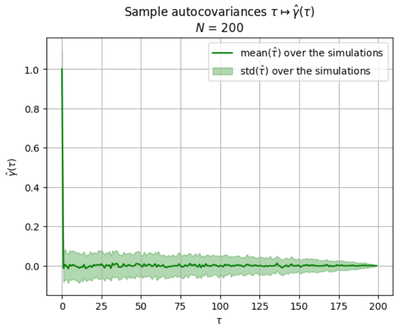
$|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Question 6

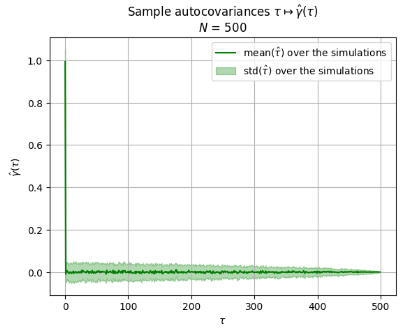
In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

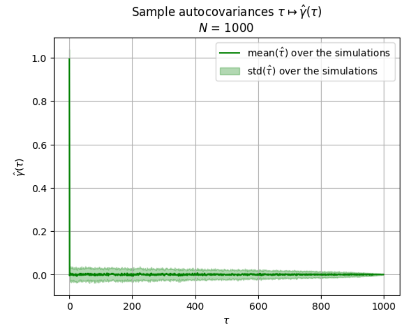
Add your plots to Figure 1.



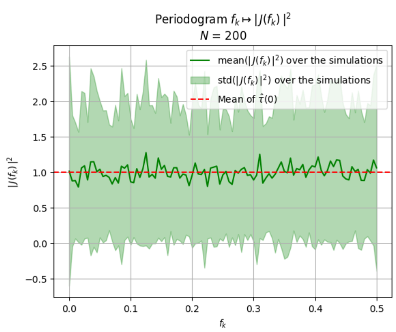
Autocovariance ($N = 200$)



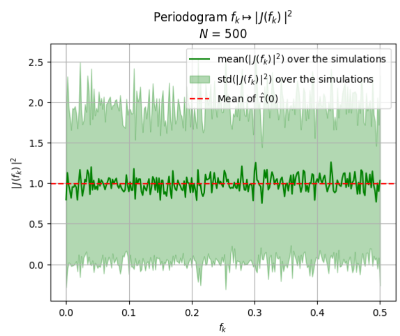
Autocovariance ($N = 500$)



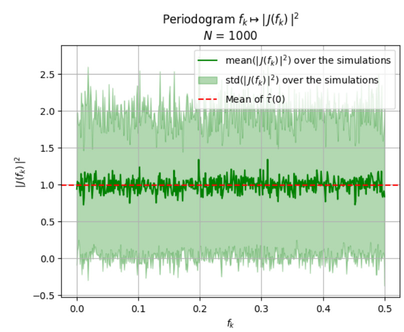
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

We observe that the sample autocovariances are very close to 0 for $\tau \neq 0$ and that the sample autocovariance for $\tau = 0$ is close to 1. This is consistent with the fact that X_n is a Gaussian white noise with variance $\sigma^2 = 1$ (as shown in Question 3).

We also observe that the periodogram is very close to 1 for all frequencies. This is consistent because, as shown in the previous question,

$$|J(f_k)|^2 = \gamma(0) + 2\sum_{\tau=1}^{N-1} \hat{\gamma}(\tau)\cos\left(\frac{2k\pi}{N}\tau\right),$$

and since $\hat{\gamma}(\tau)$ is close to 0 for $\tau \neq 0$, we have $|J(f_k)|^2 \approx \gamma(0) = 1$ for all $k \in [\![0, \frac{N}{2}]\!]$.

Question 7

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

- Show that for $\tau > 0$

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

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

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

Answer 7

We can write,

$$\begin{aligned}
\mathrm{var}(\hat{\gamma}(\tau)) &= \mathbb{E}[\hat{\gamma}(\tau)^2] - \mathbb{E}[\hat{\gamma}(\tau)]^2 \\
&= \mathbb{E}\left[\left(\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right)^2\right] - \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right]^2 \\
&= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] - \sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}]\right) \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \cancel{\mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}]} + \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_{n+\tau} X_m] \\
&\quad - \cancel{\mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}]} \qquad \text{using the hint.} \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_{n+\tau} X_m]
\end{aligned}$$

Let's deal with the first term inside the sum. We can write,

$$\mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] = \mathbb{E}\left[X_n X_{n+(m-n)}\right]\mathbb{E}\left[X_{n+\tau} X_{n+\tau+(m-n)}\right] = \gamma(n-m)^2$$

simply by definition of γ. Furthermore,

$$\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \gamma(n-m)^2 = \sum_{s=-(N-\tau-1)}^{N-\tau-1} Card\left(\{(n,m) \in [\![0, N-\tau-1]\!] | n-m=s\}\right)\gamma(s)^2$$

We can see that the cardinal of the set is equal to $N - \tau - |s|$. Thus,

$$\begin{aligned}
\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \gamma(n-m)^2 &= \sum_{s=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|s|)\gamma(s)^2 \\
&= N\sum_{s=-(N-\tau-1)}^{N-\tau-1}\left(1 - \frac{\tau+|s|}{N}\right)\gamma(s)^2
\end{aligned}$$

We deal with the second term inside the sum in the same way. We can write,

$$\mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_{n+\tau} X_m] = \mathbb{E}\left[X_n X_{n+(m-n+\tau)}\right]\mathbb{E}\left[X_{n+\tau} X_{n+\tau+(m-n-\tau)}\right] = \gamma(m-n+\tau)\gamma(m-n-\tau).$$

Thus,

$$\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}\gamma(m-n+\tau)\gamma(m-n-\tau) = \sum_{s=-(N-\tau-1)}^{N-\tau-1}(N-\tau-|s|)\gamma(s+\tau)\gamma(s-\tau)$$
$$= N\sum_{s=-(N-\tau-1)}^{N-\tau-1}\left(1-\frac{\tau+|s|}{N}\right)\gamma(s+\tau)\gamma(s-\tau)$$

Finally, by adding up the two terms, we have the desired result,

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

We write the mean square error of $\hat{\gamma}(\tau)$ as the sum of its variance and its bias squared (bias-variance decomposition),

$$\mathbb{E}\left[(\hat{\gamma}(\tau)-\gamma(\tau))^2\right] = \mathrm{var}(\hat{\gamma}(\tau)) + b(\hat{\gamma}(\tau))^2.$$

We have shown in Question 4 that $\hat{\gamma}(\tau)$ is asymptotically unbiased. Thus, $b(\hat{\gamma}(\tau))$ goes to 0 as N goes to infinity.

Let's now show that $\mathrm{var}(\hat{\gamma}(\tau))$ goes to 0 as N goes to infinity. We start by finding an upper bound for $\gamma(\tau)$. We use the Cauchy-Schwarz inequality,

$$\forall\tau,\quad \gamma(\tau) = \mathbb{E}[X_0X_\tau] \leq \sqrt{\mathbb{E}[X_0^2]\mathbb{E}[X_\tau^2]}$$

And since, X is a wide-sense stationary process, $\gamma(0) = \mathbb{E}[X_0^2] = \mathbb{E}[X_\tau^2], \forall\tau$.

Thus,

$$\forall\tau,\quad \gamma(\tau)\leq\gamma(0).$$

And since, $\tau+|s|\leq N-1, \forall s\in [\![-(N-\tau-1), N-\tau-1]\!]$, which gives us $1-\frac{\tau+|s|}{N}\leq 1-\frac{N-1}{N}$, we have,

$$\mathrm{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N}\sum_{s=-(N-\tau-1)}^{N-\tau-1}\left(1-\frac{N-1}{N}\right)\gamma(0)^2 = \frac{2(N-\tau)-1}{N}\left(1-\frac{N-1}{N}\right)\gamma(0)^2.$$

Thus,

$$0\leq\lim_{N\to\infty}\mathrm{Var}(\hat{\gamma}(\tau)) \leq \lim_{N\to\infty}\frac{2(N-\tau)-1}{N}\left(1-\frac{N-1}{N}\right)\gamma(0)^2 = 0.$$

Therefore, $\lim_{N\to\infty}\mathrm{Var}(\hat{\gamma}(\tau)) = 0$.

We thus have,

$$\lim_{N\to\infty}\mathbb{E}\left[(\hat{\gamma}(\tau)-\gamma(\tau))^2\right] = \lim_{N\to\infty}\mathrm{var}(\hat{\gamma}(\tau)) + b(\hat{\gamma}(\tau))^2 = 0.$$

Thus, $\hat{\gamma}(\tau)$ converges in L^2 to $\gamma(\tau)$, which implies that $\hat{\gamma}(\tau)$ converges in probability. 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 $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n/f_s$. Observe that $J(f) = (1/\sqrt{N})(A(f) + iB(f))$.

- 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

Let's start by computing the mean and variance of $A(f)$ and $B(f)$.

We have,

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

Similarly, $\mathbb{E}[B(f)] = 0$.

We now compute the variance of $A(f_k)$, for $k \in [\![0, N/2]\!]$. We have,

$$
\begin{aligned}
\operatorname{var}(A(f_k)) &= \mathbb{E}\left[A(f_k)^2\right] - \mathbb{E}\left[A(f_k)\right]^2 \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} X_n \cos\left(\frac{2\pi k n}{N}\right)\right)^2\right] - 0 \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(\frac{2\pi k n}{N}\right)\cos\left(\frac{2\pi k m}{N}\right)\right] \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left[X_n X_m\right] \cos\left(\frac{2\pi k n}{N}\right)\cos\left(\frac{2\pi k m}{N}\right)
\end{aligned}
$$

Since X is a Gaussian white noise,

$$
\mathbb{E}\left[X_n X_m\right] = \begin{cases} \sigma^2 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}.
$$

Thus,

$$
\begin{aligned}
\mathrm{var}(A(f_k)) &= \sum_{n=0}^{N-1} \mathbb{E}\left[X_n^2\right] \cos\left(\frac{2\pi k n}{N}\right)^2 \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos\left(\frac{2\pi k n}{N}\right)^2 \\
&= \sigma^2 \sum_{n=0}^{N-1} \frac{1+\cos\left(\frac{4\pi k n}{N}\right)}{2} && \text{(linearization of } \cos^2\text{)} \\
&= \frac{N\sigma^2}{2} + \frac{\sigma^2}{2} \sum_{n=0}^{N-1} \cos\left(\frac{4\pi k n}{N}\right) \\
&= \frac{N\sigma^2}{2} + \mathrm{Re}\left(\frac{\sigma^2}{2} \sum_{n=0}^{N-1} \exp\left(\frac{4i\pi k}{N}\right)^n\right) \\
&= \frac{N\sigma^2}{2} + \mathrm{Re}\left(\frac{\sigma^2}{2} \left(1-\exp\left(\frac{4i\pi k}{N}\right)^N\right)\left(1-\exp\left(\frac{4i\pi k}{N}\right)\right)^{-1}\right) && \text{(geometric sum)}
\end{aligned}
$$

Taking the real part,

$$
\begin{aligned}
\mathrm{Re}\left(\frac{1-\exp\left(\frac{4i\pi k}{N}\right)^N}{1-\exp\left(\frac{4i\pi k}{N}\right)}\right) &= \frac{\mathrm{Re}\Big((1-\exp(4i\pi k))(1-\exp(-\frac{4i\pi k}{N}))\Big)}{2|\sin(\frac{2\pi k}{N})|} \\
&= \frac{\mathrm{Re}\Big(1-\exp(4i\pi k)-\exp(-\frac{4i\pi k}{N})+\exp(4i\pi k(1-\frac{1}{N}))\Big)}{2|\sin(\frac{2\pi k}{N})|} \\
&= \frac{-\cos(\frac{4\pi k}{N})+\cos(4\pi k(1-\frac{1}{N}))}{2|\sin(\frac{2\pi k}{N})|} \\
&= \frac{\cos(4\pi k\frac{N-1}{N})-\cos(\frac{4\pi k}{N})}{2|\sin(\frac{2\pi k}{N})|} \\
&= 0
\end{aligned}
$$

Finally :

$$
\mathrm{var}(A(f_k)) = \frac{N\sigma^2}{2}
$$

Similarly, we have,

$$
\mathrm{var}(B(f_k)) = \frac{N\sigma^2}{2}
$$

Let's now compute the distribution of $|J(f_k)|^2$.

$$\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\left(A(f_k)^2 + B(f_k)^2\right) \\
&= \frac{1}{N}\left(\sum_{n=0}^{N-1} X_n \cos\left(-\frac{2\pi kn}{N}\right)\right)^2 + \frac{1}{N}\left(\sum_{n=0}^{N-1} X_n \sin\left(-\frac{2\pi kn}{N}\right)\right)^2 \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(-\frac{2\pi kn}{N}\right)\cos\left(-\frac{2\pi km}{N}\right) + \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \sin\left(-\frac{2\pi kn}{N}\right)\sin\left(-\frac{2\pi km}{N}\right) \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(-\frac{2\pi k(n-m)}{N}\right)
\end{aligned}$$

Like we did in the previous question, using the fact that X is a gaussian white noise, this expression leads to $\mathbb{E}[|J(f_k)|^2] = \sigma^2$. We're going to use the same argument for this next part : let's compute the variance of $|J(f_k)|^2$ as its covariance with itself.

$$\begin{aligned}
\mathrm{Var}(|J(f_k)|^2) &= \mathrm{cov}(|J(f_k)|^2, |J(f_k)|^2) \\
&= \mathbb{E}\Bigg[\left(\frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(-\frac{2\pi k(n-m)}{N}\right) - \sigma^2\right) \\
&\times \left(\frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(-\frac{2\pi k(n-m)}{N}\right) - \sigma^2\right)\Bigg] \\
&= \mathbb{E}\left[\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_i X_j X_n X_m \cos\left(-\frac{2\pi k(n-m)}{N}\right)\cos\left(-\frac{2\pi k(i-j)}{N}\right)\right] \\
&- 2\sigma^2\mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(-\frac{2\pi k(n-m)}{N}\right)\right] + \sigma^4
\end{aligned}$$

Using the hint given in the previous question, we can write,

$$\begin{aligned}
\mathrm{Var}(|J(f_k)|^2) &= \frac{1}{N^2} \times 3N^2\sigma^4 - 2\sigma^2 \times \sigma^2 + \sigma^4 \\
&= 2\sigma^4
\end{aligned}$$

This variance does not depend on N so the periodogram cannot be consistent. Furthermore this variance is consistent with the fact that $|J(f_k)|^2$ follows a χ^2 distribution with 2 degrees of freedom (we can show that A and B are independent with similar computations as above, but we will not do it here as our computations are long enough as it is).

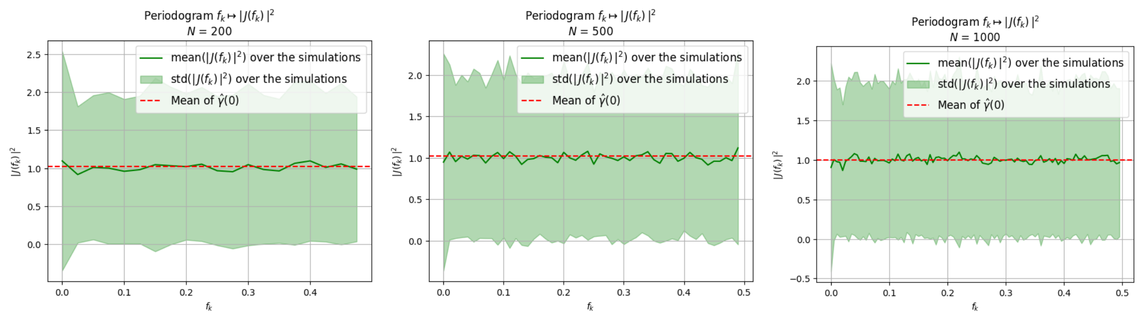
Question 9

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).

We observe, as expected, that Barlett's periodograms have $K = 5$ times less variance. While the variance of the periodograms still doesn't decrease with N, it is still greatly reduced.

4 Data study

4.1 General information

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

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

Data. Data are described in the associated notebook.

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

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

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

Question 10

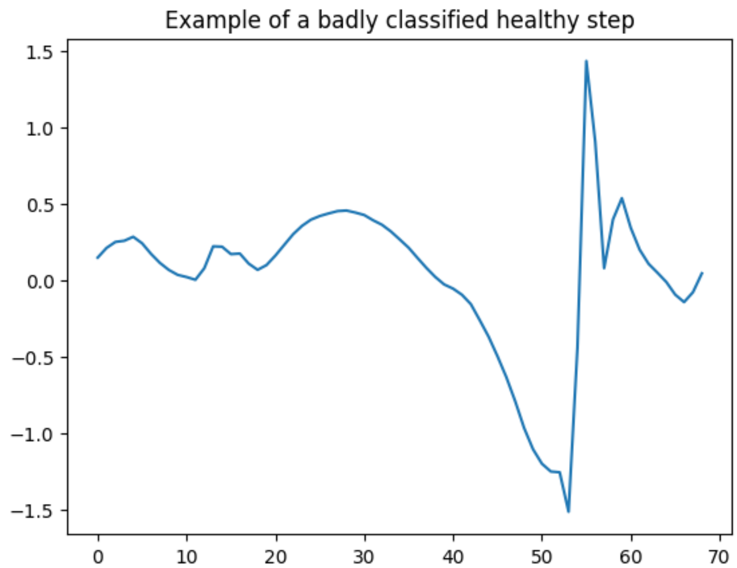
Combine the DTW and a k-neighbors classifier to classify each step. Find the optimal number of neighbors with 5-fold cross-validation and report the optimal number of neighbors and the associated F-score. Comment briefly.

Answer 10

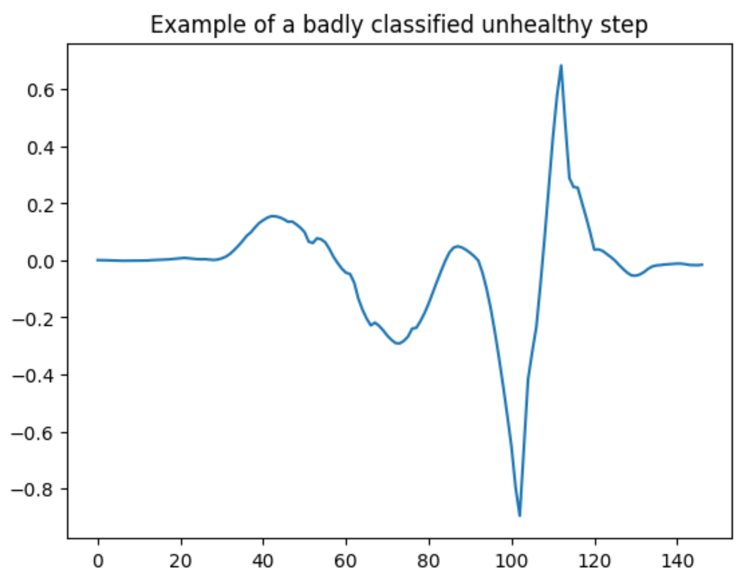
Question 11

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

Answer 11



Badly classified healthy step



Badly classified non-healthy step

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