Assignment 1 (ML for TS) - MVA

Martin Jolif martinjolif@gmail.com
Enzo Abdallah-Herve enzo.abdallahherve@gmail.com

October 28, 2024

1 Introduction

Objective. This assignment has three parts: questions about convolutional dictionary learning, spectral features, and a data study using the DTW.

Warning and advice.

- Use code from the tutorials as well as from other sources. Do not code yourself well-known procedures (e.g., cross-validation or k-means); use an existing implementation.
- The associated notebook contains some hints and several helper functions.
- Be concise. Answers are not expected to be longer than a few sentences (omitting calculations).

Instructions.

- Fill in your names and emails at the top of the document.
- Hand in your report (one per pair of students) by Tuesday 28th October 23:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname2.pdf` and
  `FirstnameLastname1_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: LINK.

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 \quad + \quad \lambda \|\beta\|_1 \tag{1}$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ the design matrix, $\beta \in \mathbb{R}^p$ the vector of regressors and $\lambda > 0$ the smoothing parameter.

Show that there exists $\lambda_{\max}$ such that the minimizer of (1) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

As an intuition, if λ is big enough, minimizing (1) comes down to minimize $\lambda \|\beta\|_1$, which gives $\beta_{minimizer} = \mathbf{0}_p$.

If we note, for all $\beta \in \mathbb{R}^p$, $f(\beta) = \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$, then f is clearly convex as the sum of two convex functions.

The sub-differential of f at $\beta = \mathbf{0}_p$ is :

$$\partial f(\mathbf{0}_p) = \{-X^T y + \lambda v \mid v \in \mathbb{R}^p, \forall i \in [\![1, p]\!], v_i \in [-1, 1]\}$$

As f is convex, this set is non-empty, it means that :

$$\exists v \in [-1, 1]^p, \forall i \in [\![1, p]\!], |(X^T y)_i| = |\lambda v_i| \leq \lambda$$

So if we choose,

$$\lambda_{\max} = \max_{i \in [\![1,p]\!]} |(X^T y)_i| = \left\| X^T y \right\|_\infty$$

So for all $\lambda > \lambda_{\max}$, there exists by construction $v \in [-1, 1]^p$ such that $X^T y = \lambda v$ which implies that $\beta = \mathbf{0}_p$ is the minimizer of the lasso problem (because $\mathbf{0}_p \in \partial f(\mathbf{0}_p)$).

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

Solving problem (2) for a fixed dictionary is equivalent to solve the following problem :

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

The vector of regressors $\mathbf{Z}$ is a columns matrix per block where each block is a (z_k). $\mathbf{x}$ is the response vector and the design matrix is the product MB where $M \in \mathbb{R}^{n \times Kn}$ is a line matrix per block composed of K matrix I_n and $B_{i,j} = \mathbf{d}_{i-j+1}$ with $\mathbf{d}_l = 0$ if $l \leq 0$ or $l > L$.

The equivalent problem is :

$$\min_{\mathbf{Z}} \quad \|\mathbf{x} - MB\mathbf{Z}\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{Z}\|$$

As the solution of the sparse coding problem for a fixed dictionary is the minimizer of a lasso regression, we can use the answer to question 1. There exists λ_{max} such that the minimizer of the lasso regression is $\mathbf{0}_p$ for any $\lambda > \lambda_{max}$. However the minimizer of the regression is also the solution of the sparse coding problem, so for any $\lambda > \lambda_{max}$ the sparse codes are only 0.

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 the sampling frequency by f_s, meaning that the index n corresponds to the time n/f_s. For simplicity, let N be even.

The *power spectrum* S of the stationary random process X is defined as the Fourier transform of the autocovariance function:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{3}$$

The power spectrum describes the distribution of power in the frequency space. Intuitively, large values of $S(f)$ indicate that the signal contains a sine wave at the frequency f. There are many estimation procedures to determine this important quantity, which can then be used in a machine-learning pipeline. In the following, we discuss the large sample properties of simple estimation procedures and the relationship between the power spectrum and the autocorrelation.

(Hint: use the many results on quadratic forms of Gaussian random variables to limit the number of calculations.)

Question 3

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

Let $(X_n)_{n \in [\![0,N]\!]} \sim \mathcal{N}(0, \sigma^2)$ iid, then, by independence when $\tau \neq 0$:

$$\gamma(\tau) = \mathbb{E}(X_n X_{n+\tau}) = cov(X_n, X_{n+\tau}) + \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) = \begin{cases} 0 & \text{if } \tau \neq 0 \\ \sigma^2 & \text{otherwise.} \end{cases}$$

For the power spectrum, by definition and the last result:

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

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 $\tau = 0, 1, \ldots, N-1$ 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

First, as (X_n) is weakly stationary,

$$\mathbb{E}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau}) = \frac{N-\tau}{N} \gamma(\tau) \tag{5}$$

So, $\mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau) \neq 0$. So, the estimator $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$.

Moreover, from equation (5),

$$\mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau) \xrightarrow[N \to \infty]{} 0$$

So, the estimator is asymptotically unbiased.

A simple way to de-bias this estimator would be to multiply it by $\frac{N}{N-\tau}$, this gives:

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

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \tag{6}$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \ldots, |J(f_{N/2})|^2$ where $f_k = f_s k/N$. (They can be efficiently computed using the Fast Fourier Transform.)

- Write $|J(f_k)|^2$ as a function of the sample autocovariances.
- For a frequency f, define $f^{(N)}$ the closest Fourier frequency f_k to f. Show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Answer 5

$$
\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{N-1} X_n X_j e^{-2\pi i f_k (n-j)/f_s} \\
&= \frac{1}{N}\sum_{\tau=-N+1}^{N-1}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} e^{-2\pi i f_k \tau/f_s} \\
&= \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-2\pi i f_k \tau/f_s} \\
&= \hat{\gamma}(0) + 2\sum_{\tau=1}^{N-1} \hat{\gamma}(\tau) cos(2\pi f_k \tau/f_s)
\end{aligned}
$$

By the last result, and equation (5),

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

By definition $f^{(N)} = f + \mathbf{O}(\frac{1}{N})$, so $f^{(N)} \xrightarrow[N\to\infty]{} f$. So, for all τ,

$$
\frac{N-\tau}{N}\gamma(\tau) e^{-2\pi i f^{(N)}\tau/f_s} \xrightarrow[N\to\infty]{} \gamma(\tau) e^{-2\pi i f\tau/f_s}
$$

Moreover,

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

So, by the dominated convergence theorem, as $\sum_{\tau\in\mathbb{Z}} |\gamma(\tau)| < \infty$,

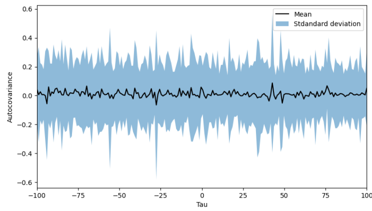
$$
\lim_{N\to+\infty} \mathbb{E}(|J(f^{(N)})|^2) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2i\pi f\tau/f_s} = S(f)
$$

Question 6

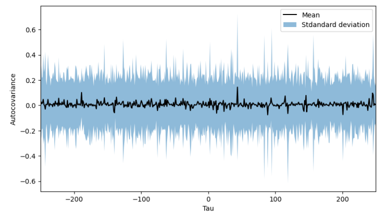
In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

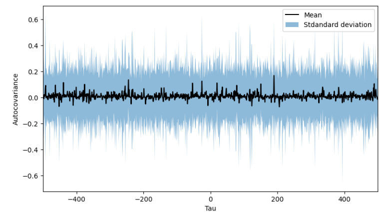
Add your plots to Figure 1.



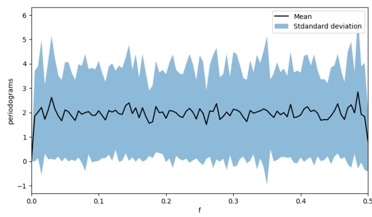
Autocovariance ($N = 200$)



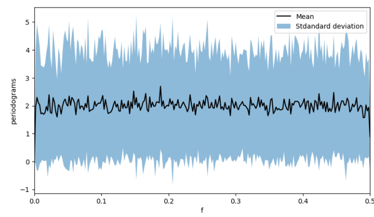
Autocovariance ($N = 500$)



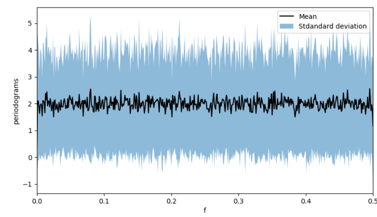
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

- For the *autocovariances*, as we do our simulations from (X_n) which can be considered independent, for $\tau \neq 0$ we obtain a mean around zero which is coherent with our results from question 3. However, we should find a mean of $\sigma^2 = 1$ for $\hat{\gamma}(0)$, which is not the case on our graphs.
- For the *periodograms*, we observe a mean around 2. However, the latest theoretical questions show that we should observe a mean around $\sigma^2 = 1$.

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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{7}$$

  (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

- Let be $\tau > 0$, by definition,

$$\mathrm{var}(\hat{\gamma}(\tau)) = \mathbb{E}(\hat{\gamma}(\tau)^2) - \mathbb{E}(\hat{\gamma}(\tau))^2$$

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

Thanks to the hint, we have:

$$= \frac{1}{N^2} \sum_{n=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} \left(\gamma(\tau)^2 + \gamma(m-n)^2 + \gamma(m+\tau-n)\gamma(m-n-\tau) \right) - \frac{1}{N^2} \sum_{n=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} \gamma(\tau)^2$$

$$= \frac{1}{N^2} \sum_{m=0}^{N-\tau-1} \sum_{n=0}^{N-\tau-1} \left(\gamma(m-n)^2 + \gamma((m-n)-\tau)\gamma((m-n)+\tau) \right)$$

$$= \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \sum_{m=0}^{N-\tau-|k|} \left(\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau) \right)$$

$$= \frac{1}{N} \sum_{n=-(N-\tau-1)}^{N-\tau-1} (\frac{N-\tau-|n|}{N}) \left(\gamma(n)^2 + \gamma(n-\tau)\gamma(n+\tau) \right).$$

We have equation (7).

- Let's show that $\hat{\gamma}(\tau)$ is consistent, let be $\epsilon > 0, a > 0$,

$$(|\hat{\gamma}(\tau) - \gamma(\tau)| > a) \iff (|\hat{\gamma}(\tau) - \mathbb{E}(\hat{\gamma}(\tau)) + \mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau)| > a)$$

So,

$$\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| > a) = \mathbb{P}(|\hat{\gamma}(\tau) - \mathbb{E}(\hat{\gamma}(\tau))| > a - |\mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau)|)$$

However, from question 4, we know that, $\mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau) \xrightarrow[N\to\infty]{} 0$. So, there exists $N_1 \in \mathbb{N}$ such that for $n \geq N_1$, $|\mathbb{E}(\hat{\gamma}(\tau)) - \gamma(\tau)| < \frac{a}{2}$.
So for $n \geq N_1$, by Bienaymé-Tchebychev,

$$\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| > a) = \mathbb{P}(|\hat{\gamma}(\tau) - \mathbb{E}(\hat{\gamma}(\tau))| > a/2) \leq \frac{4\mathrm{var}(\hat{\gamma}(\tau))}{a^2}$$

If we show that $\mathrm{var}(\hat{\gamma}(\tau)) \xrightarrow[N\to\infty]{} 0$, then there will exists $N_2 \in \mathbb{N}$ such that for $n \geq N_2$:

$$\mathrm{var}(\hat{\gamma}(\tau)) \leq \epsilon$$

Then for all $n \geq max(N_1, N_2)$,

$$\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| > a) \leq \frac{4\epsilon}{a^2}$$

ie $\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| > a) \xrightarrow[N\to\infty]{} 0$, $\hat{\gamma}$ is consistent!

Let's show that $\text{var}(\hat{\gamma}(\tau)) \xrightarrow[N\to\infty]{} 0$:

For $n \in [\![-(N-\tau-1), N-\tau-1]\!]$, $(1 - \frac{N-\tau+|n|}{N}) \leq 1$, so, from equation (7),

$$|\text{var}(\hat{\gamma}(\tau))| \leq \frac{1}{N}\left(\sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \gamma^2(n) + |\sum_{n=-(N-\tau-1)}^{n=N-\tau-1} (1 - \frac{N-\tau+|n|}{N})\gamma(n-\tau)\gamma(n+\tau)|\right)$$

So by Cauchy-Schwarz inequality in the second sum, and that for $n \in [\![-(N-\tau-1), N-\tau-1]\!]$, $(1 - \frac{N-\tau+|n|}{N}) \leq 1$, we obtain,

$$\leq \frac{1}{N}\left(\sum_{n\in\mathbb{Z}} \gamma(n)^2 + \left(\sum_{n\in\mathbb{Z}} \gamma(n-\tau)^2\right)^{1/2} \left(\sum_{n\in\mathbb{Z}} \gamma(n+\tau)^2\right)^{1/2}\right) \xrightarrow[N\to\infty]{} 0$$

The limits comes from the fact that autocovariances are square summable.

Finally, we proved that $\hat{\gamma}(\tau)$ is consistent.

Contrary to the correlogram, the periodogram is not consistent. It is one of the most well-known estimators that is asymptotically unbiased but not consistent. In the following question, this is proven for Gaussian white noise, but this holds for more general stationary processes.

Question 8

Assume that X is a Gaussian white noise (variance σ^2) and let $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s)$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n/f_s)$. Observe that $J(f) = (1/\sqrt{N})(A(f) + iB(f))$.

- 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 be $k \in [\![0, N/2]\!]$,

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

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

because X_n has a zero mean.

As the random variable are iid,

$$\mathrm{var}(B(f_k)) = \sum_{n=0}^{N-1} \mathrm{var}(X_n) \sin(-2\pi f n / f_s)^2$$

As for all $x \in \mathbb{R}, \sin(x)^2 = \frac{1}{2}(1 - \cos(2x))$,

$$= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} 1 - \cos(-4\pi k n / N)$$

$$= \frac{\sigma^2}{2} \left(N - \mathrm{Re} \left(\sum_{n=0}^{N-1} e^{-4i\pi k n / N} \right) \right)$$

$$= \frac{N\sigma^2}{2}$$

For A, we have as well,

$$\mathrm{var}(A(f_k)) = \sum_{n=0}^{N-1} \mathrm{var}(X_n) \cos(-2\pi f n / f_s)^2$$

$$= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} 1 + \cos(-4\pi k n / N)$$

$$= \frac{N\sigma^2}{2}$$

- As the $(X_n)_{n \in [\![0, N-1]\!]}$ are iid, $(X_0, \ldots, X_{N-1})$ is a gaussian vector, and therefore, $(A(f_k), B(f_k))$ is a gaussian vector too.

  Let's look at,

  $$\mathrm{cov}(A(f_k), B(f_k)) = \mathbb{E}(A(f_k)B(f_k)) - \mathbb{E}(A(f_k))\mathbb{E}(B(f_k))$$

  By the first point the expectations are equal to zero,

  $$\mathrm{cov}(A(f_k), B(f_k)) = \mathbb{E}(A(f_k)B(f_k))$$

  Then,

  $$\mathrm{cov}(A(f_k), B(f_k)) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}(X_n X_m) \cos(-2\pi k n / N) \sin(-2\pi k m / N)$$

  However, as the random variables are independant,

  $$\mathbb{E}(X_n X_m) = \begin{cases} \mathbb{E}(X_n)\mathbb{E}(X_m) & \text{if } n \neq m \\ \mathbb{E}(X_n^2) & \text{if } n = m \end{cases}$$

As X is a Gaussian white noise,

$$\mathbb{E}(X_n X_m) = \begin{cases} 0 & \text{if } n \neq m \\ \sigma^2 & \text{if } n = m \end{cases}$$

$$\text{cov}(A(f_k), B(f_k)) = \sum_{n=0}^{N-1} \sigma^2 \cos(-2\pi kn/N) \sin(-2\pi kn/N)$$

$$\text{cov}(A(f_k), B(f_k)) = \frac{\sigma^2}{2} \sum_{n=0}^{N-1} \sin(-4\pi kn/N) = \frac{\sigma^2}{2} \text{Im}(\sum_{n=0}^{N-1} (e^{-4\pi ik/N})^n) = 0$$

$(A(f_k), B(f_k))$ is a gaussian vector and the covariance is zero. Therefore $(A(f_k), B(f_k))$ are independent.

Moreover, $|J(f_k)|^2 = (1/N)(A(f_k)^2 + B(f_k)^2)$.

Therefore,

$$|J(f_k)|^2 \sim \frac{N\sigma^2/2}{N} \chi^2(2)$$

as $A(f_k)^2 \sim \frac{N\sigma^2}{2}\chi^2(1)$ and $B(f_k)^2 \sim \frac{N\sigma^2}{2}\chi^2(1)$. It follows up to a multiplicative constant a chi-squared distribution with 2 degrees of freedom.

- Then,

$$\text{var}(|J(f_k)|^2) = \frac{\sigma^4}{4} \text{var}(\chi^2(2)) = \sigma^4$$

This shows that the variance of the periodogram doesn't converge to 0, this is why the periodogram isn't consistent.

- We have,

$$\text{cov}(A(f_k), B(f_j)) = \sum_{n=0}^{N-1} \sigma^2 \cos(-2\pi kn/N) \sin(-2\pi jn/N)$$

$$\text{cov}(A(f_k), B(f_j)) = \sum_{n=0}^{N-1} \frac{\sigma^2}{2} (\sin(-2\pi(k+j)n/N) + \sin(-2\pi(k-j)n/N)) = 0$$

Moreover,

$$\text{cov}(A(f_k), A(f_j)) = \sum_{n=0}^{N-1} \sigma^2 \cos(-2\pi kn/N) \cos(-2\pi jn/N)$$

$$\text{cov}(A(f_k), A(f_j)) = \sum_{n=0}^{N-1} \frac{\sigma^2}{2} (\cos(-2\pi(k+j)n/N) + \cos(-2\pi(k-j)n/N)) = 0$$

And we also find that,

$$\text{cov}(B(f_k), B(f_j)) = 0$$

So $\text{cov}(|J(f_k)|^2, |J(f_j)|^2) = 0$

Therefore, there is no correlation between the $|J(f_k)|^2$, it explains the behaviour of the periodogram: it looks like a random process.

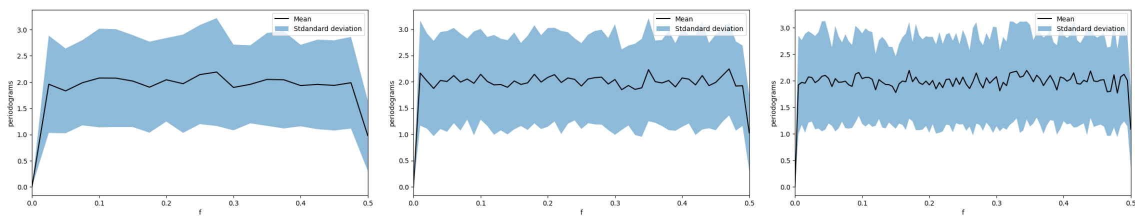
Question 9

As seen in the previous question, the problem with the periodogram is the fact that its variance does not decrease with the sample size. A simple procedure to obtain a consistent estimate is to divide the signal into K sections of equal durations, compute a periodogram on each section, and average them. Provided the sections are independent, this has the effect of dividing the variance by K. This procedure is known as Bartlett's procedure.

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

Add your plots to Figure 2.

Answer 9



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

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

We cannot take the value at the border but for the other value we find the same mean value but a lower standard deviation. It is around half the one found in the Question 6.

4 Data study

4.1 General information

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

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

Data. Data are described in the associated notebook.

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

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

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

Question 10

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

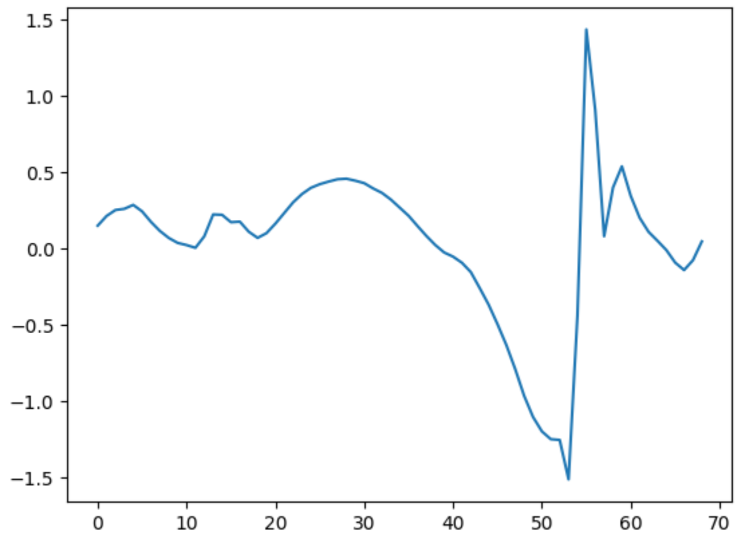
Answer 10

By combining the DTW and a k-neighbors classifier with 5-fold cross-validation, the optimal number of neighbors seems to be 5, with a F-score of 0.78 on the train dataset. The associated F-score on the test dataset is: 0.51. The classifier isn't really able to generalize on unseen data and the final F-score is not so good. Therefore, for a medical diagnosis we will not use the DTW/k-neighbors classifier method, because it will often classify the footsteps signal in the wrong class.

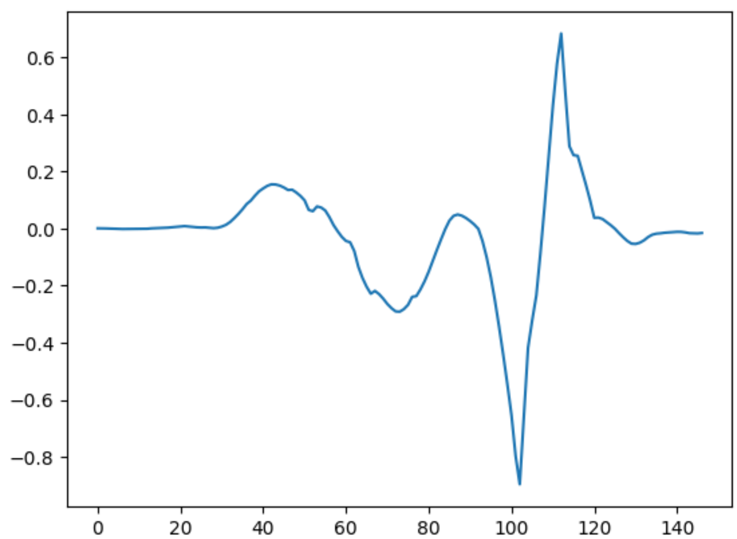
Question 11

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).