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

Rania Bennani rania.bennani@ens-paris-saclay.fr
Raphael Razafindralambo raphael.razafin@gmail.com

January 14, 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

Let's show that

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

Let $\mathcal{L} : \beta \mapsto \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$ and assume β is a solution of the minimization. Then, the first order condition gives

$$-X_j^T(y - X\beta) = \lambda g_j \tag{3}$$

where g is a subgradient of $\|.\|$ at β, defined for all j in $\{1, \ldots, p\}$ by:

$$g_j \in \begin{cases} \{1\} & \text{if } \beta_j > 0, \\ [-1, 1] & \text{if } \beta_j = 0, \\ \{-1\} & \text{if } \beta_j < 0. \end{cases} \tag{4}$$

It follows that $\beta = 0 \implies X_j^T y \in \lambda[-1, 1] \; \forall j \in \{1, \ldots, p\}$.

Now suppose that $\lambda > \left\| X^T y \right\|_\infty$. Then $|X_j^T y| < \lambda$ for all j, which means $-X_j^T y - \lambda < 0 < -X_j^T y + \lambda \iff X_j^T y \in \lambda] - 1, 1[$. The latter is only possible if $\beta_j = 0$ for all j.

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:

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

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

For all $z = (z_k)_k$ and $d = (d_k)_k$ such that $\|d_k\|_2^2 \leq 1$, let us define

$$f(\mathbf{z}, \mathbf{d}) = \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 . \tag{6}$$

By Parseval theorem and convolution theorem,

$$f(\mathbf{z},\mathbf{d}) = \left\| \hat{\mathbf{x}} - \sum_{k=1}^{K} \widehat{\mathbf{z}_k} \odot \widehat{\mathbf{d}_k} \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 . \tag{7}$$

where $\hat{}$ denotes the DFT and $\odot$ the component-wise product. If we define

$$\widehat{D} = \begin{bmatrix} \mathrm{diag}(\hat{\mathbf{d}}_1) & 0 & \cdots & 0 \\ 0 & \mathrm{diag}(\hat{\mathbf{d}}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{diag}(\hat{\mathbf{d}}_K) \end{bmatrix} \text{ and } \hat{z} = \begin{bmatrix} \hat{\mathbf{z}}_1 \\ \hat{\mathbf{z}}_2 \\ \vdots \\ \hat{\mathbf{z}}_K \end{bmatrix}$$

we have

$$\begin{aligned} f(\mathbf{z},\mathbf{d}) &= \left\| \hat{\mathbf{x}} - \sum_{k=1}^{K} \mathrm{diag}(\widehat{\mathbf{d}_k}) \widehat{\mathbf{z}_k} \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \\ &= \left\| \hat{\mathbf{x}} - \widehat{\mathbf{D}} \hat{\mathbf{z}} \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 . \end{aligned}$$

For a fixed dictionary the sparse coding problem is a lasso regression for response vector $\hat{\mathbf{x}}$ and design matrix $\widehat{\mathbf{D}}$. Therefore,

$$\lambda_{\max} = \left\| \widehat{D}^T \hat{x} \right\|_\infty . \tag{8}$$

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{9}$$

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

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

Question 3

In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise.

- Calculate the associated autocovariance function and power spectrum. (By analogy with the light, this process is called "white" because of the particular form of its power spectrum.)

Answer 3

By definition of a Gaussian white noise,

$$\gamma(\tau) = \mathbb{E}[X_n X_{n+\tau}] = \begin{cases} \sigma^2 & \text{if } \tau = 0, \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

Therefore,

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

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{11}$$

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

This estimator is biased.

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

However, it is asymptotically unbiased because

$$\frac{N-\tau}{N} \gamma(\tau) \xrightarrow[N \to +\infty]{} \gamma(\tau). \tag{12}$$

To de-bias this estimator, we consider $\tilde{\gamma}(\tau) = \frac{N}{N-\tau} \hat{\gamma}(\tau)$.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \tag{13}$$

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

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

Now, let's show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.
We have:

$$|J(f^N)|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{-2\pi i \frac{f^N \tau}{f_s}} \tag{14}$$

The mean is:

$$\mathbb{E}[|J(f^{(N)})|^2] = \sum_{\tau=-(N-1)}^{N-1} \frac{N-\tau}{N}\gamma(\tau)\mathbb{E}\left[e^{-2\pi i \frac{f^N \tau}{f_s}}\right]. \tag{15}$$

Also, we know that the f^N is the closest discrete frequency to f. Thus,

$$\mathbb{E}[e^{-2\pi i \frac{f^N \tau}{f_s}}] \xrightarrow[N\to+\infty]{} \mathbb{E}[e^{-2\pi i \frac{f \tau}{f_s}}] \tag{16}$$

Hence,

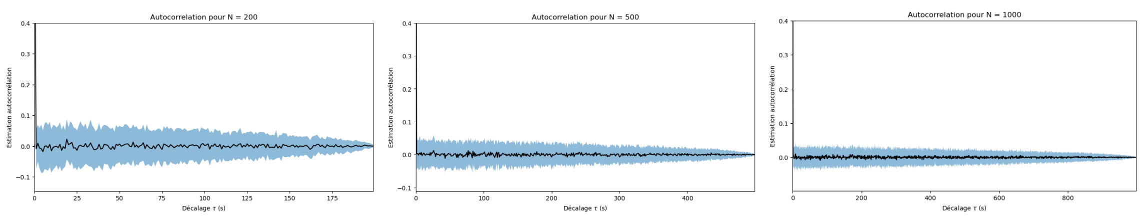
$$\lim_{N\to\infty} \mathbb{E}[|J(f^{(N)})|^2] = \lim_{N\to\infty} \sum_{\tau=-(N-1)}^{N-1} \frac{N-\tau}{N}\gamma(\tau)\mathbb{E}\left[e^{-2\pi i \frac{f\tau}{f_s}}\right]$$
$$= \lim_{N\to\infty} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-2\pi i \frac{f\tau}{f_s}}$$
$$= S(f).$$

Question 6

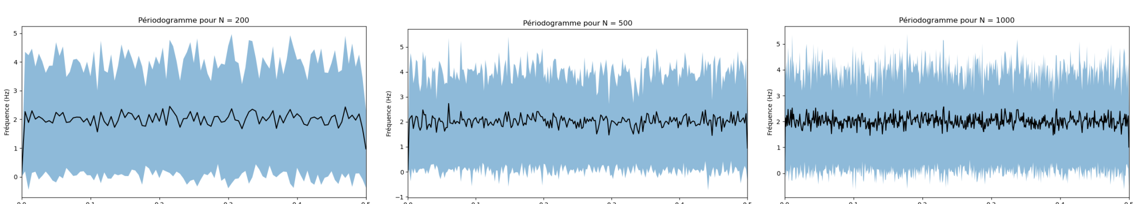
In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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 autocovariance's figures, we observe that: In three cases, the narrowing band decreases with increasing τ, which indicates that the estimates of the autocovariance become more precise for larger lags. However the band is less wide for higher values of N. The estimator's variance decreases, resulting in a tighter confidence band around the mean estimate.

- Regarding the periodogram, we have: In three cases, the confidence band is quite constant over frequency. Hence,even if N increases, the confidence band doesn't decrease (unlike the autocovariance).

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 $\tau > 0$

$$\mathrm{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{17}$$

  (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

$$Var(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)^2] - [\mathbb{E}[\hat{\gamma}(\tau)]]^2$$

We have:

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

By using the hint, we know that :

$$\begin{aligned}
\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] = &\ \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}] \\
&+ \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] \\
&+ \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_m X_{n+\tau}].
\end{aligned}$$

Also, we have : $\mathbb{E}[X_n X_{n+\tau}] = \gamma(\tau)$ and $\mathbb{E}[X_n X_m] = \gamma(|n-m|)$

Thus,

$$\mathbb{E}[\hat{\gamma}(\tau)^2] = \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}[\gamma(\tau)\gamma(\tau)+\gamma(|n-m|)\gamma(|n-m|)$$
$$+\gamma(|n-m-\tau|)\gamma(|n-m+\tau|)]$$
$$= \frac{(N-\tau)^2}{N^2}\gamma(\tau)^2 + \frac{1}{N^2}\sum_{n=-(N-\tau-1)}^{N-\tau-1}(N-\tau-|n|)[\gamma^2(n)+\gamma(n-\tau)\gamma(n+\tau)]\,.$$

Additionally,

$$(\mathbb{E}[\hat{\gamma}(\tau)])^2 = (\frac{N-\tau}{N}\gamma(\tau))^2$$
$$= \frac{(N-\tau)^2}{N^2}\gamma(\tau)^2$$

Finally,

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

Regarding the expression of the variance, we have:

- $$Var(\hat{\gamma}(\tau)) \xrightarrow[N\to+\infty]{} 0. \tag{18}$$
- $$\frac{N-\tau}{N}\gamma(\tau) - \gamma(\tau) \xrightarrow[N\to+\infty]{} 0. \tag{19}$$

We have shown that the variance goes to zero and the bias goes to zero then the estimator $\hat{\gamma}(\tau)$ 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 $J(f) = (1/N)(A(f) + iB(f))$.

- 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 of $A(f)$:

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

  Variance of $A(f)$:

$$\begin{aligned}Var(A(f)) &= \sum_{n=0}^{N-1} Var[X_n \cos(-2\pi f n / f_s)] \\ &= \sigma^2 \sum_{n=0}^{N-1} \cos^2(-2\pi f n / f_s)\end{aligned}$$

  Mean of $B(f)$:

$$\begin{aligned}\mathbb{E}[B(f)] &= \sum_{n=0}^{N-1} \mathbb{E}[X_n \sin(-2\pi f n / f_s)] \\ &= 0\end{aligned}$$

  Variance of $B(f)$:

$$\begin{aligned}Var(B(f)) &= \sum_{n=0}^{N-1} Var[X_n \sin(-2\pi f n / f_s)] \\ &= \sigma^2 \sum_{n=0}^{N-1} \sin^2(-2\pi f n / f_s)\end{aligned}$$

- Distribution of the periodogram values : The distribution of $|J(f_k)|^2$ for a white gaussian noise is a chi-squared (χ^2) distribution with 1 degree of freedom, denoted as $\chi^2(1)$, since X is a gaussian process and thus $J(f_k)$ follows a $\mathcal{N}(0, \sigma^2)$.

- Variance of $|J(f_k)|^2$:
  We have:

$$\begin{aligned}\mathrm{Var}(|J(f_k)|^2) &= \mathbb{E}[|J(f_k)|^4] - (\mathbb{E}[|J(f_k)|^2])^2 \\ &= \mathbb{E}\left[\left(\frac{1}{\sqrt{N}}\sqrt{A(f_k)^2 + B(f_k)^2}\right)^4\right] - \left(\mathbb{E}\left[\frac{1}{\sqrt{N}}\sqrt{A(f_k)^2 + B(f_k)^2}\right]^2\right) \\ &= \frac{1}{N^2}\mathbb{E}[(A(f_k)^2 + B(f_k)^2)^2] - \left(\frac{1}{N}\mathbb{E}[A(f_k)^2 + B(f_k)^2]\right)^2\end{aligned}$$

  The cross terms in the expression are null because we consider white gaussian noise, then we have:

$$
\begin{aligned}
&= \frac{1}{N^2}\left(2\sigma^4 \sum_{n=0}^{N-1} \cos^2\left(\frac{2\pi f_k n}{f_s}\right) + 2\sigma^4 \sum_{n=0}^{N-1} \sin^2\left(\frac{2\pi f_k n}{f_s}\right)\right) - \left(\frac{1}{N}\sigma^2 \sum_{n=0}^{N-1} 1\right)^2 \\
&= \frac{2\sigma^4}{N} \sum_{n=0}^{N-1} (\cos^2\left(\frac{2\pi f_k n}{f_s}\right) + \sin^2\left(\frac{2\pi f_k n}{f_s}\right)) - \left(\frac{\sigma^2}{N} \sum_{n=0}^{N-1} 1\right)^2 \\
&= \frac{2\sigma^4}{N} \sum_{n=0}^{N-1} 1 - \left(\frac{\sigma^2}{N} \sum_{n=0}^{N-1} 1\right)^2 \\
&= \frac{2\sigma^4}{N} N - \left(\frac{\sigma^2}{N} N\right)^2 \\
&= 2\sigma^4 - \left(\sigma^2\right)^2 \\
&= \sigma^4.
\end{aligned}
$$

Conclusion: By looking to the expression of the variance, we notice that: $\lim_{N\to\infty} Var(|J(f_k)|^2) \neq 0$ Hence, the periodogram is not consistent.

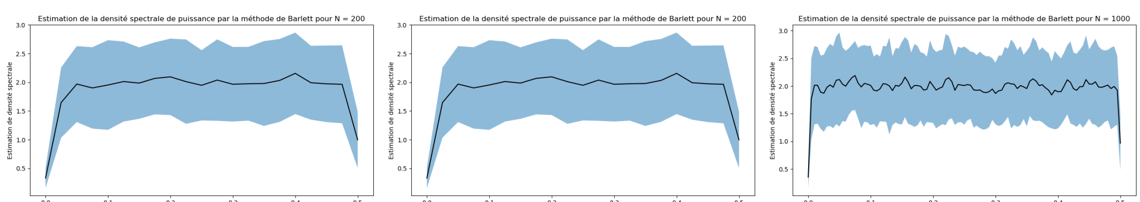
Question 9

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

- Rerun the experiment of Question 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).

The three graphs show that as the number of data points N increases, the estimates become smoother and the standard deviation bands around the average become narrower. This indicates that with more data (higher N), Bartlett's method becomes more stable and reliable, with reduced variance.

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

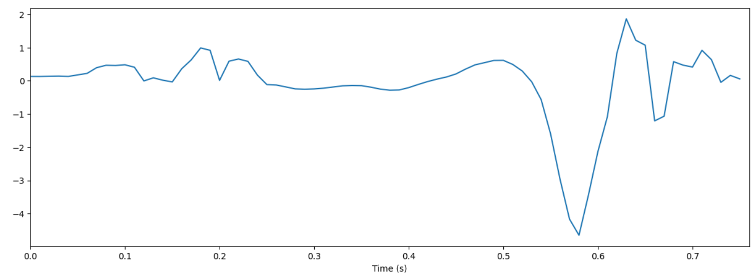
The optimal number of neighbors is $K = 1$ its associated F-score is `0.513`. The F-score is not really satisfying. It may be caused by the choice of only using the vertical acceleration of the left foot which may not provide enough discriminative information to accurately classify the activities, especially since activities like 'turning around' and 'stopping' may have subtle differences in acceleration patterns compared to 'walking' or 'standing still'.

Moreover, the dataset is high-dimensional, and it poses a challenge for the KNN algorithm which is known to suffer from the curse of dimensionality.

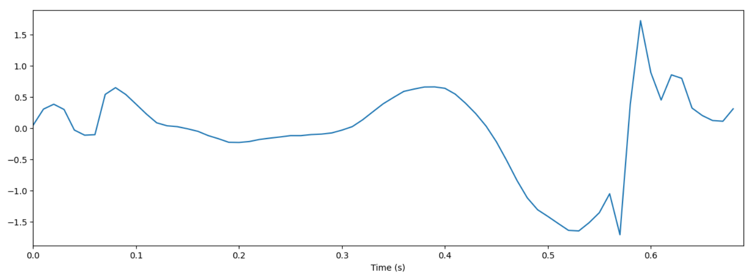
Question 11

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).