Assignment 1 (ML for TS) - MVA

Pierre Aguié pierre.aguie@gmail.com
Paco Goze paco.goze@polytechnique.edu

October 26, 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 28$^{\text{th}}$ 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

Let $\lambda > 0$, and let

$$f : \begin{cases} \mathbb{R}^p & \longrightarrow \mathbb{R} \\ \beta & \longmapsto \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \end{cases}.$$

f is a convex function over $\mathbb{R}^p$. Thus, we have that $\beta^* \in \operatorname{argmin}_\beta f(\beta) \iff \mathbf{0}_p \in \partial f(\beta^*)$ for all $\beta^* \in \mathbb{R}^p$. The sub-differential of f at $\beta \in \mathbb{R}^p$ is:

$$\partial f(\beta) = \left\{ -X^\top y + X^\top X\beta + \lambda s \;\middle|\; s_j \in \begin{cases} \{\operatorname{sign}(\beta_j)\} & \text{if } \beta_j \neq 0 \\ [-1,1] & \text{otherwise} \end{cases}, \quad j = 1, \dots, p \right\}.$$

We then have the following equivalences:

$$\begin{aligned} \mathbf{0}_p \in \partial f(\mathbf{0}_p) &\iff \mathbf{0}_p \in \left\{ -X^\top y + \lambda s \;\middle|\; s_j \in [-1,1],\ j = 1, \dots, p \right\} \\ &\iff \exists s \in [-1,1]^p,\ \lambda s = X^\top y \\ &\iff \forall j \in \{1, \dots, p\},\ \left| X_j^\top y \right| \leqslant \lambda \\ &\iff \left\| X^\top y \right\|_\infty \leqslant \lambda, \end{aligned}$$

where for all $j \in \{1, \dots, p\}$, X_j is the j-th column of X.

We then deduce that $\mathbf{0}_p$ is a minimizer of f if and only if $\lambda \geqslant \lambda_{\text{max}}$, where

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

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 λ_{max} (which depends on the dictionary) such that the sparse codes are only 0 for any $\lambda > \lambda_{\text{max}}$.

Answer 2

Let us fix $(\mathbf{d}_k) \in (\mathbb{R}^L)^K$ a dictionary of atoms. For all activation signals $(\mathbf{z}_k) \in (\mathbb{R}^{N-L+1})^K$, we have

$$\begin{aligned}\forall k \in \{1,\dots,K\}, \forall n \in \{1,\dots,N\}, (\mathbf{z}_k * \mathbf{d}_k)[n] &= \sum_{\substack{i=1\\ i \geqslant L-N+n\\ i \leqslant n}}^{L} \mathbf{d}_k[i]\mathbf{z}_k[n-i+1] \\ &= \mathbf{D}_k \mathbf{z}_k,\end{aligned}$$

where for all $k \in \{1,\dots,K\}$,

$$\mathbf{D}_k := \begin{bmatrix} \mathbf{d}_k[1] & 0 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ \mathbf{d}_k[2] & \mathbf{d}_k[1] & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & \dots & \dots & \dots & \dots & \vdots \\ \mathbf{d}_k[L] & \ddots & \ddots & \ddots & \dots & \dots & \dots & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \dots & \dots & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \dots & \dots & 0 \\ 0 & \dots & 0 & \mathbf{d}_k[L] & \mathbf{d}_k[L-1] & \dots & \mathbf{d}_k[1] & \dots & 0 \end{bmatrix} \in \mathbb{R}^{N \times (N-L+1)}$$

We then introduce $\mathbf{D} \in \mathbb{R}^{N \times K(N-L+1)}$ and $\mathbf{z} \in \mathbb{R}^{K(N-L+1)}$, defined by

$$\begin{aligned}\mathbf{D} &:= \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \dots & \mathbf{D}_K \end{bmatrix}, \\ \mathbf{z} &:= \begin{bmatrix} \mathbf{z}_1^\top & \mathbf{z}_2^\top & \dots & \mathbf{z}_K^\top \end{bmatrix}^\top.\end{aligned}$$

The sparse coding problem with fixed dictionary $(\mathbf{d}_k)$ can then be rewritten as the following Lasso regression problem, with design matrix $\mathbf{D}$ and response vector $\mathbf{x}$:

$$\min_{\mathbf{z} \in \mathbb{R}^{K(N-L+1)}} \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1.$$

Using Question 1, we have that for a fixed dictionary $(\mathbf{d}_k)$, the sparse codes are only 0 if $> \lambda_{\text{max}}$, where

$$\lambda_{\max} = \left\| \mathbf{D}^\top \mathbf{x} \right\|_\infty. \tag{4}$$

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 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:

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

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

We have $\gamma(\tau) = \mathbb{E}(X_n X_{n+\tau})$

If $\tau \neq 0, \gamma(\tau) = \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) = 0$ (independence of the samples).

If $\tau = 0, \gamma(\tau) = Var(X_n) = \sigma^2$.

Therefore :

$$S(f) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s} = \gamma(0) = \sigma^2. \tag{6}$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{7}$$

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

$$\mathbb{E}(\hat{\gamma}(\tau)) = \mathbb{E}\left((1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right) = (1/N) \sum_{n=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau}) = \frac{N-\tau}{N}\gamma(\tau) \tag{8}$$

The estimator is asymptotically unbiased :

$$\lim_{N \to +\infty} \mathbb{E}(\hat{\gamma}(\tau)) = \gamma(\tau)$$

.

A simple way to de-bias this estimator would be to divide the sum by $N - \tau$ rather than N.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{9}$$

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

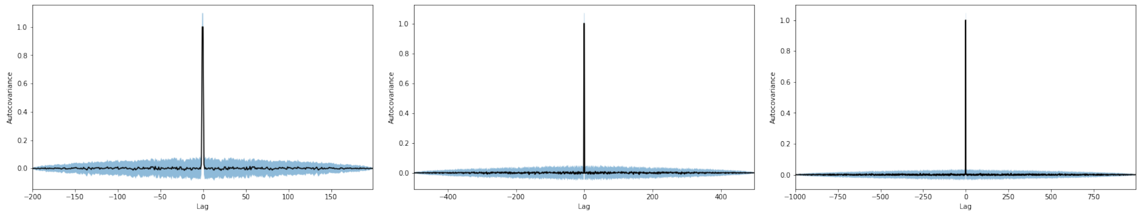
$$\begin{aligned}
|J(f_k)|^2 &= (1/N)\left(\sum_{n=0}^{N-1} X_n e^{-2\pi i n k/N}\right)\left(\sum_{j=0}^{N-1} X_j e^{2\pi i j k/N}\right) \\
&= (1/N)\left(\sum_{n=0}^{N-1} {X_n}^2\right) + (1/N)\left(\sum_{n=0}^{N-1}\sum_{\substack{j=0\\ j\neq n}}^{N-1} X_n X_j e^{\frac{2\pi i k(j-n)}{N}}\right) \\
&= (1/N)\left(\sum_{n=0}^{N-1} {X_n}^2\right) + (1/N)\left(\sum_{n=0}^{N-1}\sum_{\tau=1}^{N-1-n} 2X_n X_j \cos\left(\frac{2\pi k\tau}{N}\right)\right) \\
&= \hat{\gamma}(0) + \sum_{\tau=1}^{N-1} 2\cos\left(\frac{2\pi k\tau}{N}\right)\hat{\gamma}(\tau)
\end{aligned} \tag{10}$$

Question 6

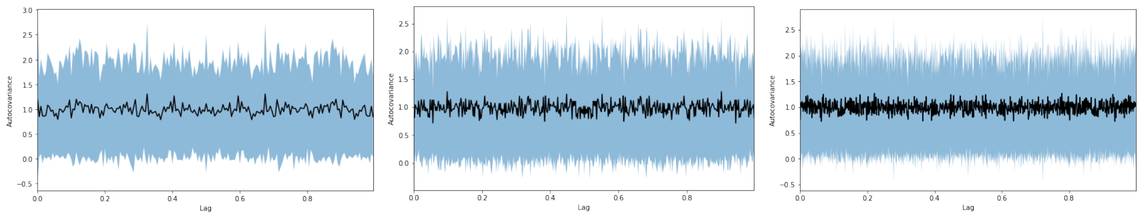
In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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

Regarding the sample autocovariances:

- The estimations show an autocovariance of 1 for $\tau = 0$, and an autocovariance of 0 for all other lags. This is the result we expect for a Gaussian white noise of variance $\sigma^2 = 1$. This is coherent with the fact that $\hat{\gamma}(\tau)$ is asymptotically unbiased, as shown in Question 4.
- For a fixed N, the standard deviation of $\hat{\gamma}(\tau)$ decreases as the lag τ increases. Thus, it seems that $\hat{\gamma}(\tau)$ is a consistent estimator.
- For a fixed τ, the standard deviation of $\hat{\gamma}(\tau)$ decreases as N increases.

Regarding the periodograms:

- The estimations of $S(f)$ are close to 1 for all values of f, which is coherent with the results from Question 3 for a Gaussian white noise of variance $\sigma^2 = 1$. It seems that $\left|J(f^{(N)})\right|^2$ is an asymptotically unbiased estimator of $S(f)$.
- The standard deviation of $|J(f_k)|^2$ does not vary with f_k or N, and is quite high. It seems that $\left|J(f^{(N)})\right|^2$ is not a consistent estimator.

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 $\tau > 0$

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

(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

$$\begin{aligned}
var(\hat{\gamma}(\tau)) &= \frac{1}{N^2} var\left(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right) \\
&= \frac{1}{N^2}\left(\mathbb{E}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} X_n X_{n+\tau} X_m X_{m+\tau}\right) - \left(\sum_{n=0}^{N-\tau-1}\mathbb{E}\left(X_n X_{n+\tau}\right)\right)^2\right) \\
&= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}\mathbb{E}\left(X_n X_{n+\tau} X_m X_{m+\tau}\right) - (N-\tau)^2\gamma(\tau)^2\right) \\
&= \frac{1}{N^2}\Bigg[\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}\mathbb{E}(X_n X_{n+\tau})\mathbb{E}(X_m X_{m+\tau}) + \mathbb{E}(X_n X_{m+\tau})\mathbb{E}(X_m X_{n+\tau}) + \\
&\mathbb{E}(X_n X_m)\mathbb{E}(X_{n+\tau} X_{m+\tau}) - (N-\tau)^2\gamma(\tau)^2\Bigg] \\
&= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}\gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)\right) \\
&= \frac{1}{N^2}\left(\sum_{n-m=-(N-\tau-1)}^{N-\tau-1}(N-\tau-|n-m|)\left(\gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)\right)\right) \\
&= \frac{1}{N^2}\left(\sum_{n=-(N-\tau-1)}^{N-\tau-1}(N-\tau-|n|)\left(\gamma(n)^2 + \gamma(n-\tau)\gamma(n+\tau)\right)\right) \\
&= \frac{1}{N}\left(\sum_{n=-(N-\tau-1)}^{N-\tau-1}\left(1-\frac{\tau+|n|}{N}\right)\left(\gamma(n)^2 + \gamma(n-\tau)\gamma(n+\tau)\right)\right)
\end{aligned} \tag{12}$$

We have previously shown that $\hat{\gamma}(\tau)$ is asymptotically unbiased. For the consistency of the estimator, it is therefore sufficient to prove that :

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

$$\begin{aligned}
0 \le var(\hat{\gamma}(\tau)) &\le \frac{1}{N}\left(\sum_{n=-(N-\tau-1)}^{N-\tau-1}\left(\gamma(n)^2 + \gamma(n-\tau)\gamma(n+\tau)\right)\right) \\
&\le \frac{1}{N}\left(\sum_{n=-(N-\tau-1)}^{N-\tau-1}\left(\gamma(n)^2\right)\right) + \frac{1}{N}\left(\sum_{n=-(N-\tau-1)}^{N-\tau-1}\gamma(n-\tau)\gamma(n+\tau)\right)
\end{aligned} \tag{13}$$

Since the autocovariances are square summable, the left term tends towards 0. Given the Cauchy Schwarz inequality, and the fact that the autocovariances are absolutely summable, we easily show that the right term tends towards 0 as well.

Therefore :

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

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 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

We have the following :

$$\mathbb{E}\left(A(f_k)\right) = \mathbb{E}\left(B(f_k)\right) = 0. \tag{14}$$

$$\begin{aligned} var(A(f_k)) &= var\left(\sum_{n=0}^{N-1} X_n cos(-2\pi nk/N)\right) \\ &= \sum_{n=0}^{N-1} \sigma^2 cos(-2\pi nk/N)^2 \\ &= \sigma^2 \frac{N}{2}. \end{aligned} \tag{15}$$

$$\begin{aligned} var(B(f_k)) &= var\left(\sum_{n=0}^{N-1} X_n sin(-2\pi nk/N)\right) \\ &= \sum_{n=0}^{N-1} \sigma^2 sin(-2\pi nk/N)^2 \\ &= \sigma^2 \frac{N}{2}. \end{aligned} \tag{16}$$

Since X is a Gaussian white noise, $\tilde{X} = (X_0, X_1, ..., X_{N-1})$ is a Gaussian vector. As linear transformations of $\tilde{X}$ coordinates, $A(f)$ and $B(f)$ therefore follow a normal distribution $\mathcal{N}(0, \frac{N}{2}\sigma^2)$.

$A(f_k)$ and $B(f_k)$ are jointly normal. Let us show that they are independent by calculating their covariance :

$$\begin{aligned}
Cov(A(f_k), B(f_k)) &= \mathbb{E}(A(f_k)B(f_k)) \\
&= -\mathbb{E}\left(\sum_{n=}^{N-1}\sum_{m=0}^{N-1} X_n X_m cos(2\pi nk/N) sin(2\pi mk/N)\right) \\
&= -\mathbb{E}\left(\sum_{n=}^{N-1} {X_n}^2 cos(2\pi nk/N) sin(2\pi nk/N)\right) \\
&= -\sum_{n=}^{N-1} \frac{1}{2}^2 sin(4\pi nk/N) \\
&= 0
\end{aligned} \tag{17}$$

$A(f_k)$ and $B(f_k)$ are independent ; $A(f_k)^2$ and $B(f_k)^2$ are therefore independent.

Finally :

$|J(f_k)|^2 = \frac{1}{N}((A(f_k)^2 + B(f_k)^2) \sim \frac{\sigma^2}{2}\chi^2(2)$.

Using the previous expression, we directly have :

$var(|J(f_k)|^2) = \sigma^4$

We here proved (for a Gaussian white noise process) what we observed in **Question 6** : $var(|J(f_k)|^2)$ is fixed and does not tends towards 0 when N tends towards $+\infty$. The periodogram is not consistent.

Similarly to what we did before, we can show that $Cov(|J(f_k)|^2, |J(f_l)|^2) = O$ when kl. The periodogram values are uncorrelated, hence the erratic behavior of the periodogram in **Question 6**.

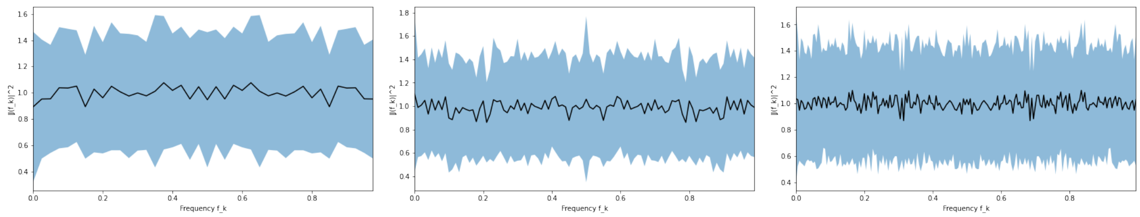
Question 9

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 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).

As expected, the variance of Barlett's estimate seems to have been divided by K. Indeed, in the plots in Figure 1, the periodograms had a standard deviation close to 1, and in the plots in Figure 2, the periodograms have a standard deviation close to approximately $0.4 \approx \frac{1}{\sqrt{K}}$ for $K = 5$. Just like before, it does not seem that the standard deviations are affected by f_k or by N. This is because Barlett's estimate is not a consistent estimator.

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.

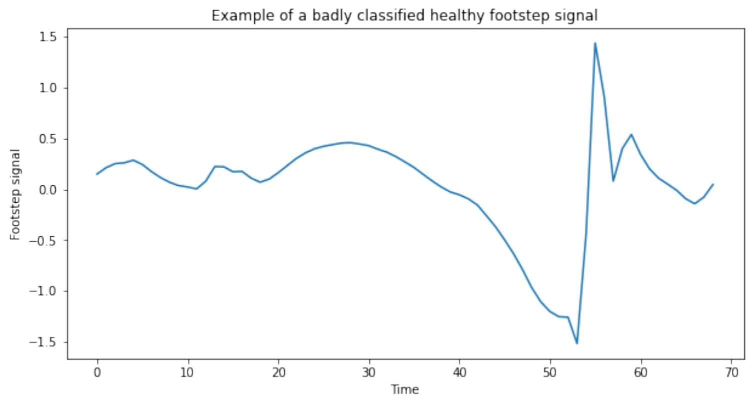
Answer 10

Using 5-fold cross-validation, we find an optimal number of neighbors of 5, with an associated average F-score over the 5 folds of $F1 \approx 0.77$. Training over the entirety of the training set a 5-neighbors classifier and testing it over the test set yields a test F-score of $F1 \approx 0.45$. This performacne can be explained by the fact that nearest-neighbor methods generalize poorly when the training set is small and in high-dimensional settings (The training set had 168 time series, which contained on average 74 samples).

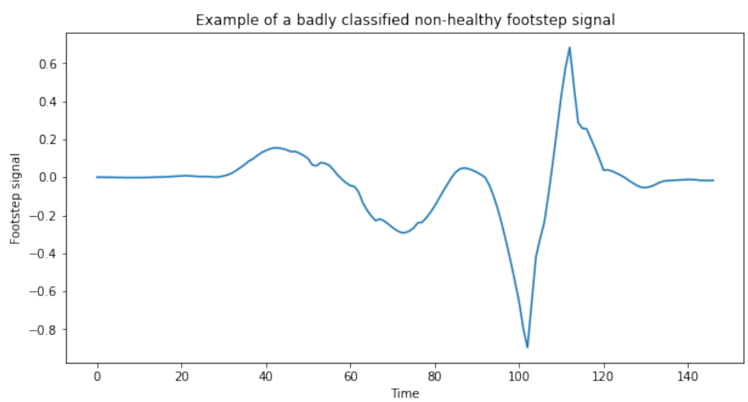
Question 11

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).