

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

Yvann Le Fay yvann.lefay@ensae.fr
Antoine Schoonaert antoine.schoonaert@etu.emse.fr

November 5, 2023

1 Introduction

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

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/1FAIpQLSdTwJEyc6QloYTknjk12kJMtcKllFvPIWLk5LbyugW0YO7K6Q/viewform?usp=sf_link.

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

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

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

Answer 1

Let $\mathcal{L} : \beta \in \mathbb{R}^p \mapsto \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$, this is a non-differentiable (because of the l_1 -norm term) convex function. Its subdifferential is given by

$$\partial \mathcal{L}(\beta) = -X^\top (y - X\beta) + \lambda \partial \|\beta\|_1, \quad (2)$$

where the subdifferential of the l_1 -norm is given by

$$\partial \|\beta\|_1 = \{z \in \mathbb{R}^p : z_j = \text{sgn}(\beta_j) \text{ if } \beta_j \neq 0, z_j \in [-1, 1] \text{ if } \beta_j = 0\}. \quad (3)$$

The first-order optimality implies at the optimal $\beta(\lambda)$, there exists $z^* \in \partial \|\beta(\lambda)\|_1$ such that

$$X^\top (y - X\beta(\lambda)) + \lambda z^* = 0, \quad (4)$$

i.e.,

$$X^\top X\beta(\lambda) = X^\top y - \lambda z^*. \quad (5)$$

Thus, we have

$$\begin{aligned} 0 \leq \beta(\lambda) X^\top X\beta(\lambda) &= \langle \beta(\lambda), X^\top y - \lambda z^* \rangle \\ &= \sum_{j/\beta(\lambda)_j \neq 0} \beta(\lambda)_j (X_j^\top y - \lambda \text{sgn}(\beta(\lambda)_j)), \end{aligned} \quad (6)$$

where the X_j 's are the columns of X and the $\beta(\lambda)_j$'s are the components of $\beta(\lambda)$. The previous inequality implies that for $\lambda \geq \max_{j \in [1, p]} |X_j^\top y|$, $\beta(\lambda)_j = 0$ (however, it does not seem trivial to show that it is the minimal λ such that $\beta(\lambda)$ is 0). Thus, for any $\lambda \geq \lambda_{\max}$, where

$$\lambda_{\max} = \max_{j \in [1, p]} |X_j^\top y|, \quad (7)$$

we have $\beta(\lambda) = 0$.

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:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2 \leq 1} \left\| \mathbf{x} - \sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{z}_k\|_1 \quad (8)$$

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 $\lambda > \lambda_{\max}$.

Answer 2

We have $\sum_{k=1}^K (\mathbf{z}_k * \mathbf{d}_k)(i) = \sum_{k=1}^K \sum_{j=0}^{L-1} \mathbf{z}_k(j) \mathbf{d}_k(i-j)$ which can be rewritten $\sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k = D\mathbf{z}$ where

$$D = (D_1, \dots, D_K), \quad \mathbf{z} = [\mathbf{z}_1^\top, \dots, \mathbf{z}_K^\top]^\top, \quad (9)$$

and where the D_i 's are Toeplitz matrices for the convolution product by the \mathbf{d}_i 's, i.e.,

$$D_i = \begin{pmatrix} d_i(0) & 0 & \dots & 0 & 0 \\ d_i(1) & d_i(0) & & \dots & \dots \\ d_i(2) & d_i(1) & \dots & 0 & 0 \\ \dots & d_i(2) & \dots & d_i(0) & 0 \\ d_i(L-2) & \dots & \ddots & d_i(1) & d_i(0) \\ d_i(L-1) & d_i(L-2) & & \vdots & d_i(1) \\ 0 & d_i(L-1) & \ddots & d_i(L-3) & \vdots \\ 0 & 0 & \dots & d_i(L-2) & d_i(L-3) \\ \vdots & \vdots & & d_i(L-1) & d_i(L-2) \\ 0 & 0 & 0 & \dots & d_i(L-1) \end{pmatrix}. \quad (10)$$

In that case, using the result of the previous question, we have

$$\lambda_{\max} = \max_{i \in [1, K]} \|D_i^\top \mathbf{x}\|_\infty. \quad (11)$$

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2i\pi f \tau / f_s}. \quad (12)$$

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

Let $X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ ($n = 0, \dots, N-1$), then the autocovariance function is given, for any $\tau \in [0, N-1]$, by

$$\gamma(\tau) = \mathbb{E}(X_0 X_\tau) = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}, \quad (13)$$

and the power spectrum is given by

$$S = \sigma^2. \quad (14)$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \quad (15)$$

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

Let $\tau \in [0, N-1]$, we have

$$\begin{aligned} \mathbb{E}[\hat{\gamma}(\tau)] &= \frac{1}{N} \sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}] \\ &= \frac{N-\tau}{N} \gamma(\tau), \end{aligned} \quad (16)$$

which is biased up to a multiplicative constant, except for $\tau = 0$. As N goes to infinity, the multiplicative constant $\frac{N-\tau}{N}$ goes to 1, which shows that the estimator is asymptotically unbiased. Let $\hat{\gamma}^* = \frac{N}{N-\tau} \hat{\gamma}$, this is an de-biased version of the previous estimator.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \quad (17)$$

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 $k \in \mathbb{N}$, we have

$$\begin{aligned}
|J(f_k)|^2 &= J(f_k) \overline{J(f_k)} \\
&= \frac{1}{N} \sum_{n,m=0}^{N-1} X_n X_m e^{-2\pi i k(n-m)/N} \\
&= \frac{1}{N} \sum_{n=0, m=0}^{N-1} \sum_{\tau=-(N-1)}^{N-1} \mathbb{1}(n-m=\tau) X_n X_m e^{-2\pi i k(n-m)/N} \\
&= \frac{1}{N} \left\{ \sum_{\tau=1}^{N-1} \sum_{m=0}^{N-1-\tau} X_{m+\tau} X_m e^{-2i\pi k/N\tau} + \sum_{\tau=1}^{N-1} \sum_{n=0}^{N-1-\tau} X_n X_{n+\tau} e^{2i\pi k/N\tau} + \sum_{n=0}^{N-1} X_n^2 \right\} \quad (18) \\
&= \frac{1}{N} \left\{ 2 \sum_{\tau=1}^{N-1} \cos(2\pi k/N\tau) \sum_{n=0}^{N-1-\tau} X_n X_{n+\tau} + \sum_{n=0}^{N-1} X_n^2 \right\} \\
&= 2 \sum_{\tau=1}^{N-1} \cos(2\pi k/N\tau) \hat{\gamma}(\tau) + \hat{\gamma}(0) \\
&= \sum_{\tau=-(N-1)}^{N-1} e^{-2i\pi k/N\tau} \hat{\gamma}(\tau),
\end{aligned}$$

where we used $\gamma(\tau) = \gamma(-\tau)$ for $\tau \in [-(N-1), -1]$ for the last line. Using the previous result and formally computing the expectation, we have for $k(f)$ realising the $f_k(f) = f^{(N)}$

$$\mathbb{E}[|J(f^{(N)})|^2] = \sum_{\tau=-(N-1)}^{N-1} e^{-2i\pi k(f)/N\tau} (1 - \tau/N) \gamma(\tau). \quad (19)$$

Since the autocovariances are absolutely summable, and $k(f)/N$ converges to f/f_s , the sum converges uniformly to $S(f)$.

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 ??.

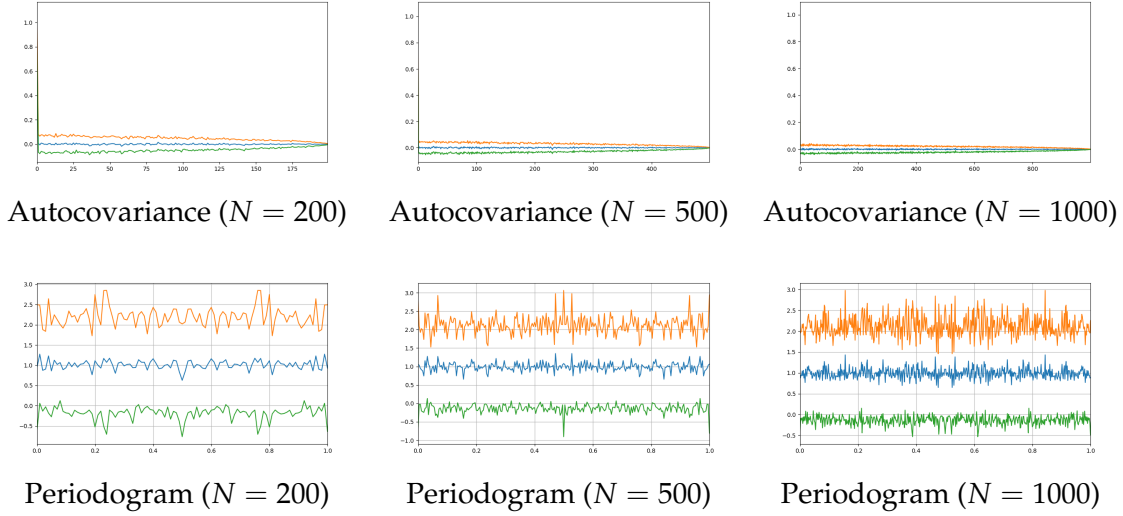


Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question ??).

Answer 6

We observe that the sample autocovariance functions goes to 0 as τ goes to N , which can be foreseen using Question 3. Similarly for the standard deviation, which can be foreseen since as τ goes to N , we have little available samples. Moreover, we observe that the autocovariance function goes to 0 for $\tau > 0$ as N goes to infinity, which is totally previsible because $\tau = \sigma^2 \mathbb{1}_{\{0\}}$.

Regarding the periodogram, the closed-formula given Question 3) for the power spectrum shows that $|J(f_k)|^2$ should be an estimator of the constant $\sigma^2 = 1$, which is the case. Moreover, the computed variance of $|J(f_k)|^2$ is the constant $\sigma^4 = 1$. This is confirmed by the plots. More importantly, we remark the standard deviation does not go to 0 as N goes to infinity.

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 $\tau > 0$

$$\text{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) [\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)]. \quad (20)$$

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

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

Answer 7

Let us compute the variance, using the above equality to go from the first to the second line, we have for $\tau > 0$

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}^2(\tau)] &= \frac{1}{N^2} \sum_{n,m=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] \\
&= \frac{1}{N^2} \sum_{n,m=0}^{N-\tau-1} \{ \mathbb{E}[X_n X_{n+\tau}] \mathbb{E}[X_m X_{m+\tau}] + \mathbb{E}[X_n X_m] \mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}] \mathbb{E}[X_m X_{n+\tau}] \} \\
&= \frac{1}{N^2} \sum_{n,m=0}^{N-\tau-1} \{ \gamma(\tau)^2 + \gamma(m-n)^2 + \gamma(n+\tau-m) \gamma(m+\tau-n) \}.
\end{aligned} \tag{21}$$

Thus, subtracting $\mathbb{E}[\hat{\gamma}(\tau)]^2 = (1 - \tau/N)^2 \gamma(\tau)^2 = \frac{1}{N^2} \sum_{n,m=0}^{N-\tau-1} \gamma(\tau)^2$, we have

$$\begin{aligned}
\text{var}(\hat{\gamma}(\tau)) &= \frac{1}{N^2} \sum_{n,m=0}^{N-\tau-1} \gamma(m-n)^2 + \gamma(n+\tau-m) \gamma(m+\tau-n) \\
&= \frac{1}{N^2} \sum_{n=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} \left\{ \sum_{k=-(N-\tau-1)}^{N-\tau-1} \mathbb{1}(m-n=k) \right\} (\gamma^2(k) + \gamma(k-\tau) \gamma(\tau+k)) \\
&= \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \text{card}((m,n) \in [0, N-\tau-1]^2 / m-n=k) (\gamma^2(k) + \gamma(k-\tau) \gamma(\tau+k)) \\
&= \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|) (\gamma^2(k) + \gamma(k-\tau) \gamma(\tau+k)) \\
&= \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|k|}{N}\right) (\gamma^2(k) + \gamma(k-\tau) \gamma(\tau+k)) \\
&\leq \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma^2(k) + |\gamma(k-\tau) \gamma(\tau+k)| \\
&\leq \frac{1}{N} \left(\sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma^2(k) + \sqrt{\sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma^2(k-\tau) \sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma^2(\tau+k)} \right) \\
&\leq \frac{2}{N} \sum_{k \in \mathbb{Z}} \gamma^2(k),
\end{aligned} \tag{22}$$

where to go from the first to the second line, we used $\gamma(k-\tau) = \gamma(\tau-k)$, computing the cardinal of pairs $(m,n) \in [0, N-\tau-1]^2$ such that $m-n=k$ can be done by a quick drawing, one obtains $N-\tau-|k|$. This is the desired result.

Now, let us prove that the estimator is indeed consistent by using Bienaymé-Tcheybychev inequality. For any $t > 0$, we have

$$\begin{aligned}
\mathbb{P}[|\hat{\gamma}(\tau) - \gamma(\tau)| > t] &\leq \frac{\text{var}(\hat{\gamma}(\tau))}{t^2} \\
&\leq \frac{c}{Nt^2},
\end{aligned} \tag{23}$$

where $c = 2 \sum_{n \in \mathbb{Z}} \gamma^2(n)$ is finite because the autocovariances are square-summable. Hence, for any $t > 0$, $\mathbb{P}[\hat{\gamma}(\tau) - \gamma(\tau) > t]$ goes to 0 as N goes to ∞ , which exactly means 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 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 ?? by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

Since the X_n 's have zero-mean, we obtain that $\mathbb{E}[A(f)] = \mathbb{E}[B(f)] = 0$. Let us compute the variance of $A(f_k)$ for $k \in [0, N/2]$, using the mutual independence of the X_n 's and trigonometric formulas,

$$\begin{aligned}
 \text{var}(A(f_k)) &= \sum_{n=0}^{N-1} \text{var}(X_n \cos(-2\pi k n / N)) \\
 &= \sigma^2 \sum_{n=0}^{N-1} \cos^2(-2\pi k n / N) \\
 &= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} (\cos(-4\pi k n / N) + 1) \\
 &= \frac{\sigma^2 N}{2}.
 \end{aligned} \tag{24}$$

Similarly, we have

$$\text{var}(B(f_k)) = \frac{\sigma^2 N}{2}. \tag{25}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}[A(f_k)B(f_k)] &= \mathbb{E}\left[\sum_{n,m=0}^{N-1} X_n X_m \cos(-2\pi kn/N) \sin(-2\pi km/N)\right] \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos(-2\pi kn/N) \sin(-2\pi kn/N) \\
&= -\frac{\sigma^2}{2} \sum_{n=0}^{N-1} \sin(4\pi kn/N) \\
&= -\frac{\sigma^2}{2} \operatorname{Im} \sum_{n=0}^{N-1} e^{4\pi i kn/N} \\
&= -\frac{\sigma^2}{2} (\operatorname{Im}(1 + e^{2\pi i k})) + \sum_{n=1}^{N/2-1} \operatorname{Im}(e^{4\pi i kn/N} + e^{4\pi i k(N-n)/N}) \\
&= 0,
\end{aligned} \tag{26}$$

where we remark that $e^{4\pi i kn/N}$ is the conjugate of $e^{4\pi i k(N-n)/N}$ and similar arguments were used for the two previous variances. Since, A and B are Gaussian variables and have null-covariance, they are independent. Hence, $|J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2)$ is the sum of the square of two independent Gaussian variables, i.e.,

$$|J(f_k)|^2 = \frac{\sigma^2}{2}(U_0^2 + U_1^2) \sim \frac{\sigma^2}{2}\chi^2(2), \tag{27}$$

where U_0 and U_1 are two independent standard Gaussian variables and the equality is to be understood as equality in law. Using the previous expression, we obtain that the variance of $|J(f_k)|^2$ is given by

$$\operatorname{var}(|J(f_k)|^2) = \sigma^4, \tag{28}$$

which is a constant. In particular, we've shown that the distribution of $|J(f_k)|^2$ does not depend on N , thus $\mathbb{P}[||J(f_k)|^2 - S(f)| > t]$ is a constant for any t and $|J(f_k)|^2$ is not a consistent estimator of $S(f)$.

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 ??, but replace the periodogram by Bartlett's estimate (set $K = 5$). What do you observe.

Add your plots to Figure ??.

Answer 9

We observe that the standard deviation is now close to 0.4, which can be predicted. Indeed, since we split the signals in 5 batches, it means that the standard deviation of our mean batch estimator has a standard deviation of $1/\sqrt{5} \approx 0.44$.

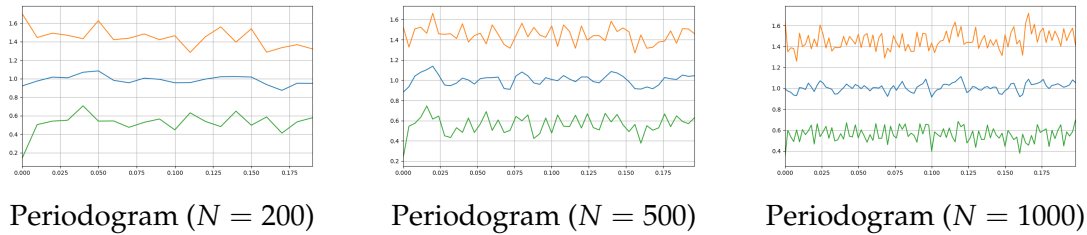


Figure 2: Barlett's periodograms of a Gaussian white noise (see Question ??).

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.

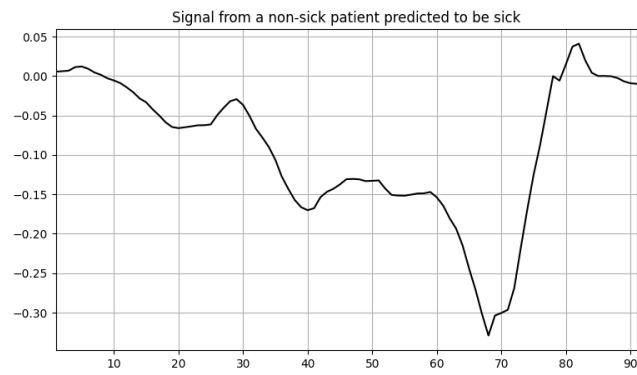
Answer 10

The reported optimal k is 7, however, this estimate is not consistent as the F-scores for the different k 's are very close while being sensitive to the random process of splitting the dataset into two parts. The associated F-score is 0.87.

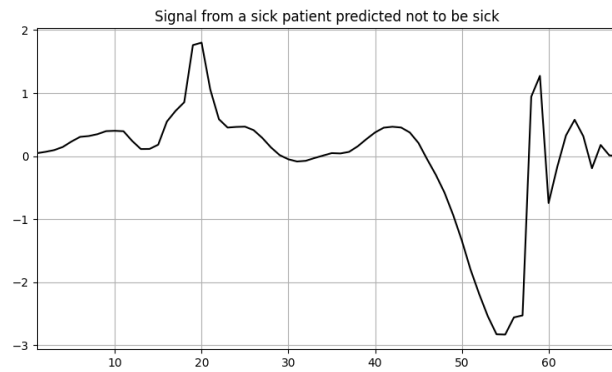
Question 11

Display on Figure ?? 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 ??).