Assignment 1 (ML for TS) - MVA 2022/2023

Matias Etcheverry matias.etcheverry9@gmail.com
Amric Trudel amric.trudel@ens-paris-saclay.fr

April 11, 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 Wednesday 1st February 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: dropbox.com/request/8uHP2WLfYTS1Js8LNkP6.

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

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

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

Intuitively, if λ is huge, the cost of not being sparse overcomes the cost of not fitting the data. Unfortunately, (1) is not derivable in 0. The sub-differential of the loss $\mathcal{L}$ is:

$$\partial(\mathcal{L}(\beta)) = \left\{ X^T(y - X\beta) + \lambda z \mid z \in [-1, 1] \right\}$$

Thus $\beta = 0$ is a solution of the problem if it belongs to the sub-differential, i.e. :

$$\beta \in \partial(\mathcal{L}(0)) \Rightarrow X^T y = \lambda(-z) \quad \text{with } z \in [-1, 1]$$

This condition can be achieved if $-\mathbb{1} \leq \frac{1}{\lambda} X^T y \leq \mathbb{1}$ element-wise, i.e. $\lambda \geq \boxed{\lambda_{\max} = \left\| X^T y \right\|_\infty}$

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

With such problem, without taking into account padding, it is quite clear that we can't reconstruct properly the sides of x. For a fixed dictionary, the response vector is then $y = x[(N - L + 1)//2 : -(N - L + 1)//2]$. We first construct $D \in \mathbb{R}^{L \times K}$ the matrix of dictionary and $Z \in \mathbb{R}^{N-L+1 \times K}$ the matrix of activations. With such matrices, the problem becomes a LASSO problem with:

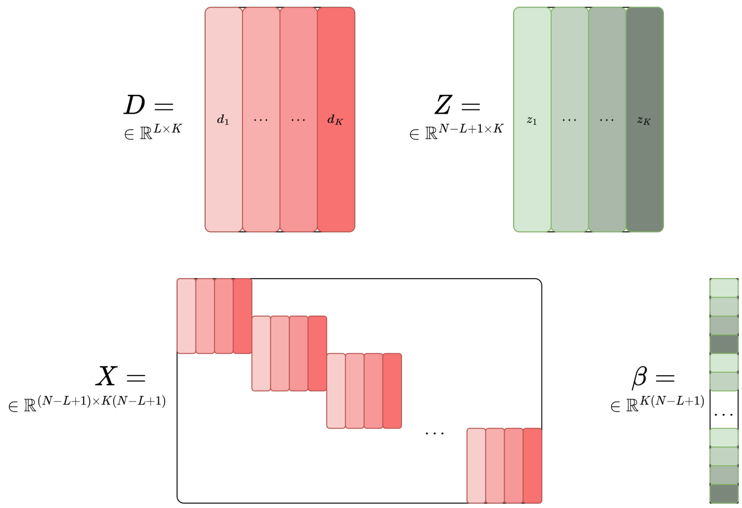


Figure 1: Matrices for Lass problem

with X a shifted version of D and β an interleaved version of Z. As in question 1,

$$\begin{aligned}\lambda_{\max} &= \left\| X^T y \right\|_\infty \\ &= \left\| \sum_{k=1}^{K} \mathbf{y} * \mathbf{d}_k \right\|_\infty\end{aligned}$$

where y is the response vector.

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 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:

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

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.

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 a Gaussian white noise, with mean $\mu = 0$ and std $\sigma > 0$. Each X_i is a sample of this distribution, i.e. there are independent. Then for τ:

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

For the power spectrum:

$$\begin{aligned} S(f) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s} \\ &= \gamma(0)e^0 + \sum_{\tau=-\infty}^{-1} \gamma(\tau) e^{-2\pi f \tau / f_s} + \sum_{\tau=1}^{+\infty} \gamma(\tau) e^{-2\tau / f_s} \\ &= \gamma(0) \\ &\boxed{S(f) = \sigma^2} \end{aligned}$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{4}$$

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

Let's compute the average sample autocovariance:

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}_N(\tau)] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right] \\
&= \frac{1}{N}\sum_{n=0}^{N-\tau-1} \mathbb{E}\left[X_n X_{n+\tau}\right] \\
&= \frac{1}{N}\sum_{n=0}^{N-\tau-1} \gamma(\tau) \quad \text{as } X \text{ is weakly stationary} \\
&= \frac{N-\tau}{N}\gamma(\tau) \\
\text{i.e.} \quad & \boxed{\mathbb{E}[\hat{\gamma}(\tau)] \neq \gamma(\tau)}
\end{aligned}$$

The estimator is biased. However, the above equation shows:

$$\boxed{\lim_{N\to\infty} \mathbb{E}[\hat{\gamma}_N(\tau)] = \gamma(\tau)}$$

A simple way to de-bias this estimator is to multiply is by $\frac{N}{N-\tau}$, i.e. we end up with:

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

Question 5

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

$$J(f) := (1/\sqrt{N})\sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{5}$$

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

First subquestion: Expressing $|J(f_k)|2$ as a function of the sample autocovariances:

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

This very last line aims at counting the values of a matrix made from the product between the vectors $[X_n e{-}\frac{-2i\pi kn}{N}]$ and $[X_m e{-}\frac{2i\pi km}{N}]$. Fig 2 shows how to count the values in this matrix. We see that if we count on the diagonals, we can make $\hat{\gamma}(\tau)$ appear.

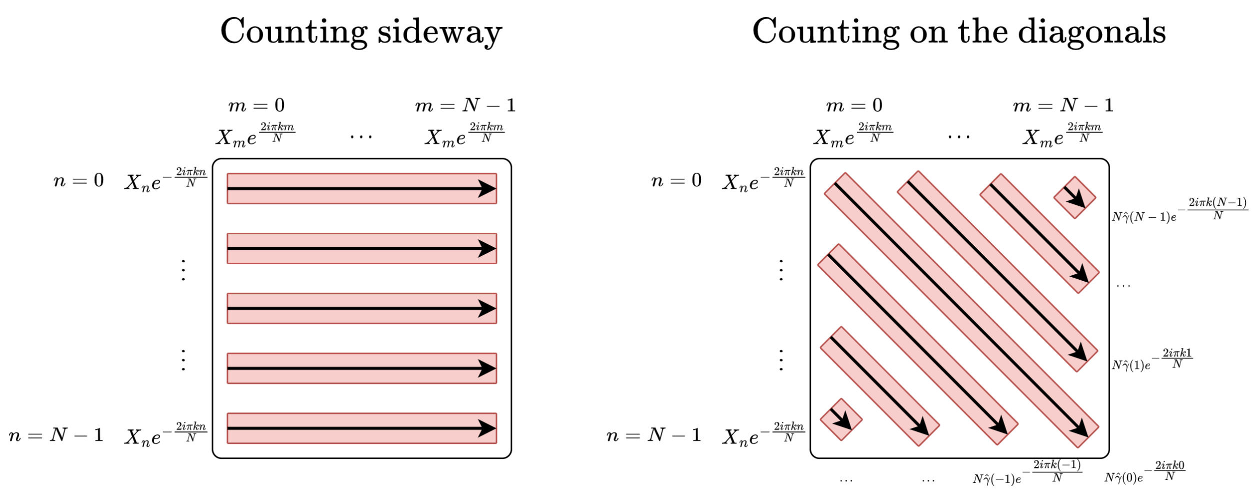


Figure 2: How to count values in matrix

Finally, we have:

$$
|J(f_k)|^2 = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} N\hat{\gamma}(\tau) e^{-\frac{2i\pi k\tau}{N}}
$$

$$
\text{ie} \quad \boxed{|J(f_k)|^2 = \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{\frac{-2\pi i k\tau}{N}}}
$$

Second subquestion: defining $f^{(N)}$ as the closest Fourier frequency f_k to f

$$f^{(N)} = \arg\min_{f_k} |f - f_k| \tag{6}$$

We denote $k^{(N)}$ the index so that $f^{(N)} = \frac{k^{(N)} f_s}{N}$. Third subquestion: Showing that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

We use the dominated convergence theorem. For any $f < f_s/2$:

- It is simple to prove that $|f^{(N)} - f| \leq \frac{f_0}{N}$. Then, we have $\lim_{N\to\infty} f^{(N)} = f$, and thus $\lim_{N\to\infty} \hat{\gamma}(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} = \gamma(\tau) e^{-\frac{2i\pi f\tau}{f_s}}$
- from a certain N, as $\lim_{N\to\infty} \hat{\gamma}(\tau) = \gamma(\tau)$, it is possible to have $\left|\hat{\gamma}(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}}\right| \leq 2|\gamma(\tau)|$ which is finite, and sommable.

Then

$$\begin{aligned}\lim_{N\to\infty} \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} &= \lim_{N\to\infty} \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-\frac{2i\pi k^{(N)}\tau}{N}} = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-\frac{2i\pi f\tau}{f_s}} \\ \text{i.e.} \quad \lim_{N\to\infty} |J(f^{(N)})|^2 &= S(f)\end{aligned}$$

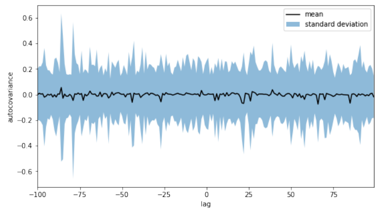
In the end, $\underline{|J(f^{(N)})|^2 \text{ is an asymptotically unbiased estimator of } S(f)}$.

Question 6

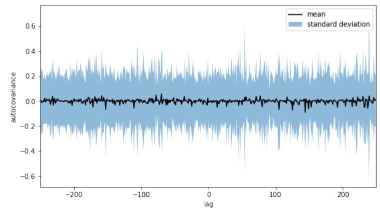
In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

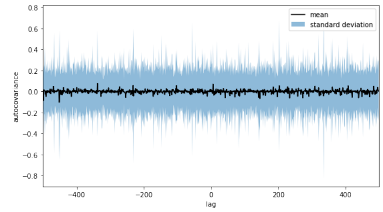
Add your plots to Figure 3.



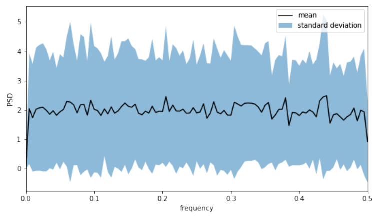
Autocovariance ($N = 200$)



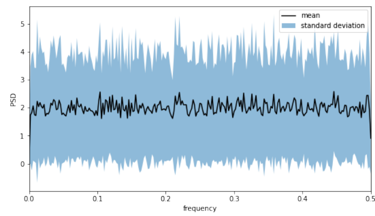
Autocovariance ($N = 500$)



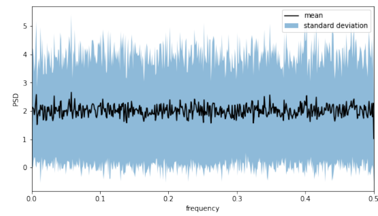
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

Figure 3: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

Answer 6

On the graphs, we observe the following:

- The average autocovariance and PSD of many white noise signals also look like white noise.
- The mean autocovariance and PSD is stationary.
- The standard deviation doesn't change across time.
- All properties discussed above are identical regardless of the length of the time series.

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 $\mathbb{E}[Y_1 Y_2 Y_3 Y_4] = \mathbb{E}[Y_1 Y_2]\mathbb{E}[Y_3 Y_4] + \mathbb{E}[Y_1 Y_3]\mathbb{E}[Y_2 Y_4] + \mathbb{E}[Y_1 Y_4]\mathbb{E}[Y_2 Y_3]$.)

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

Answer 7

We have:

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

To obtain the last equality, we count the number of times a certain m occurs in the sum.

We simply need $\lim_{N\to\infty}\mathrm{Var}(\hat{\gamma}(\tau)) = 0$ to prove that the sample autocovariance is consistent:

$$
\begin{aligned}
\mathrm{Var}(\hat{\gamma}(\tau)) &= \frac{1}{N^2}\sum_{m=-(N-\tau-1)}^{N-\tau-1}(N-\tau-|m|)(\gamma(m)^2 + \gamma(m+\tau)\gamma(m-\tau)) \\
&\leq \frac{1}{N^2}\sum_{m=-(N-\tau-1)}^{N-\tau-1} N(\gamma(m)^2 + \gamma(m+\tau)\gamma(m-\tau)) \\
&\leq \frac{1}{N}\left(\sum_{m=-N}^{N}\gamma(m)^2 + \left(\sum_{m=-N}^{N}\gamma(m)\right)^2\right) \\
&\xrightarrow[N\to\infty]{} 0
\end{aligned}
$$

Thus, the variance of $\hat{\gamma}(\tau)$ tends to 0. We apply Bienaymé-Tchebychev inequality to $\hat{\gamma}$. Let $\epsilon > 0$:

$$
\mathbb{P}[|\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| > \frac{\epsilon}{2}] \leq 2\frac{\mathrm{Var}(\hat{\gamma}(\tau))}{\epsilon}
$$

And $\mathbb{E}[\hat{\gamma}(\tau)]$ can be close to $\gamma(\tau)$ to $\frac{\epsilon}{2}$. This proves that $\hat{\gamma}(\tau)$ is consistent.

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

- The *sinus* and *cosinus* are just scale factors in $A(f)$ and $B(f)$. Thus it doesn't change the mean of the Gaussians, i.e. $\boxed{\mathbb{E}[A(f)] = \mathbb{E}[B(f)] = 0}$.

  In order to compute the variance, we use the previously mentioned hint. For any k

$$\begin{aligned}
\mathrm{Var}(A(f_k)) &= \mathbb{E}[A(f_k)^2] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\right] \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}[X_n X_m]\mathbb{E}\left[\cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\right] \quad \text{(hint)} \\
&= \sum_{n=0}^{N-1} \mathbb{E}[X_n^2]\mathbb{E}\left[\cos\left(\frac{2\pi kn}{N}\right)^2\right] \quad (\mathbb{E}[X_n X_m] \neq 0 \text{ for } n = m) \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right)^2 \\
&= \frac{\sigma^2}{2}\left(N + \mathrm{Re}\left(\sum_{n=0}^{N-1} e^{\frac{4i\pi kn}{N}}\right)\right)
\end{aligned}$$

$$\text{i.e.} \quad \boxed{\mathrm{Var}(A(f_k)) = \frac{\sigma^2 N}{2}}$$

  Similarly,

$$\begin{aligned}
\mathrm{Var}(B(f_k)) &= \mathbb{E}[B(f_k)^2] \\
&= \sigma^2 \sum_{n=0}^{N-1} \sin\left(\frac{2\pi kn}{N}\right)^2 \\
&= \sigma^2 \sum_{n=0}^{N-1} 1 - \cos\left(\frac{2\pi kn}{N}\right)^2
\end{aligned}$$

$$\text{i.e.} \quad \boxed{\mathrm{Var}(B(f_k)) = \frac{\sigma^2 N}{2}}$$

  A and B are identical Gaussians.

- we notice $|J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2)$. Let's compute the covariance between A and B:

$$
\begin{aligned}
\operatorname{cov}(A(f_k), B(f_k)) &= \mathbb{E}\left[A(f_k)B(f_k)\right] \\
&= \mathbb{E}\left[-\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\frac{2\pi kn}{N}\sin\frac{2\pi km}{N}\right] \\
&= -\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left[X_n X_m\right]\mathbb{E}\left[\cos\frac{2\pi kn}{N}\sin\frac{2\pi km}{N}\right] \quad \text{(hint)} \\
&= -\sum_{n=0}^{N-1} \mathbb{E}\left[X_n^2\right]\cos\frac{2\pi kn}{N}\sin\frac{2\pi km}{N} \quad (\mathbb{E}[X_n X_m] \neq 0 \text{ for } n = m) \\
&= -\frac{\sigma^2}{2}\sum_{n=0}^{N-1}\sin\frac{4\pi km}{N} \\
&= -\frac{\sigma^2}{2}\operatorname{Im}\left(\sum_{n=0}^{N-1} e^{\frac{4i\pi km}{N}}\right) \\
\text{i.e.} \quad \operatorname{cov}(A(f_k), B(f_k)) &= 0
\end{aligned}
$$

A and B are i.i.d. Gaussians. Thus <u>$|J(f_k)|^2$ is an unscaled chi-squared distribution with 2 degrees of freedom</u>.

- First, let's recall that if $X \hookrightarrow \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}[X^4] = 3\sigma^4$. We then have $\operatorname{Var}(X^2) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = 2\sigma^4$. We can now compute the variance of $|J(f_k)|^2$:

$$
\begin{aligned}
\operatorname{Var}(|J(f_k)|^2) &= \frac{1}{N^2}\left(\operatorname{Var}(A(f_k)^2) + \operatorname{Var}(B(f_k)^2)\right) \quad \text{as } A^2 \text{ and } B^2 \text{ are i.i.d.} \\
&= \frac{1}{N^2}\left(2\left(\sigma\sqrt{\frac{N}{2}}\right)^4 + 2\left(\sigma\sqrt{\frac{N}{2}}\right)^4\right) \quad \text{based on the previous remark} \\
\text{i.e.} \quad & \boxed{\operatorname{Var}(|J(f_k)|^2) = \sigma^4}
\end{aligned}
$$

Furthermore, we can compute the expectation of $|J(f_k)|^2$:

$$
\begin{aligned}
\mathbb{E}[|J(f_k)|^2] &= \frac{1}{N}\left(\mathbb{E}[A(f_k)^2] + \mathbb{E}[B(f_k)^2]\right) \\
&= \frac{1}{N}\left(\operatorname{Var}(A(f_k)) + \operatorname{Var}(B(f_k))\right) \\
\text{i.e.} \quad & \boxed{\mathbb{E}[|J(f_k)|^2] = \sigma^2}
\end{aligned}
$$

Thus, the distribution of $|J(f_k)|^2$ doesn't depend on N. This means that $|J(f_k)|^2$ can always be arbitrarily away from its expectation, i.e. <u>$|J(f_k)|^2$ is not consistent.</u>

- let's compute the covariance between the $|J(f_k)|^2$. Let k, l:
  - We have:

$$
\begin{aligned}
\mathbb{E}[A(f_k)^2 B(f_l)^2] &= \mathbb{E}[A(f_k)^2]\mathbb{E}[B(f_l)^2] \quad (A \text{ and } B \text{ are independent, so are } A^2 \text{ and } B^2) \\
&= \operatorname{Var}(A(f_k))\operatorname{Var}(B(f_l)) \\
\text{i.e.} \quad \mathbb{E}[A(f_k)^2 B(f_l)^2] &= \frac{\sigma^4 N^2}{4}
\end{aligned}
$$

– We have:

$$\mathbb{E}[A(f_k)^2 A(f_l)^2] = \mathbb{E}\left[\sum_{n,m,s,t}^{N-1} X_n X_m X_s X_t \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ls}{N}\cos\frac{2\pi lt}{N}\right]$$
$$= \sum_{n,m,s,t}^{N-1} \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ls}{N}\cos\frac{2\pi lt}{N}$$
$$(\mathbb{E}[X_n X_m]\mathbb{E}[X_s X_t] + \mathbb{E}[X_n X_s]\mathbb{E}[X_m X_t] + \mathbb{E}[X_n X_t]\mathbb{E}[X_m X_s])$$

The first term of the parenthesis leads to study cases where $n = m$ and $s = t$:

$$\sum_{n,s} \cos\left(\frac{2\pi kn}{N}\right)^2 \cos\left(\frac{2\pi ls}{N}\right)^2 \mathbb{E}[X_n^2]\mathbb{E}[X_s^2] = \sigma^4 \sum_{n,s} \cos\left(\frac{2\pi kn}{N}\right)^2 \cos\left(\frac{2\pi ls}{N}\right)^2$$
$$= \frac{\sigma^4 N^2}{4} \quad \text{see computation of Var}(A)$$

The second term of the parenthesis leads to study cases where $n = s$ and $m = t$.

$$\sum_{n,m} \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}$$
$$\cos\frac{2\pi ln}{N}\cos\frac{2\pi lm}{N}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] = \sigma^4 \sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}\sum_m \cos\frac{2\pi km}{N}\cos\frac{2\pi lm}{N}$$
$$= \frac{\sigma^4}{2}\sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}\sum_m \cos\frac{2\pi(l-k)m}{N} + \cos\frac{2\pi(l+k)m}{N}$$
$$= \frac{\sigma^4 N\delta_{l,k}}{2}\sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}$$
$$= \frac{\sigma^4 N^2\delta_{l,k}}{4} \quad \text{(see computation of Var}(A))$$

The third and last term of the parenthesis leads to study cases where $n = t$ and $m = s$.

$$\sum_{n,m} \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi lm}{N}\cos\frac{2\pi lt}{N}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] = \frac{\sigma^4 N^2\delta_{l,k}}{4} \quad \text{(same method as 2nd term)}$$

Finally, once we have suffered enough, we get:

$$\mathbb{E}[A(f_k)^2 A(f_l)^2] = \frac{\sigma^4 N^2}{4}(1 + 2\delta_{l,k})$$

– the same method gives us:

$$\mathbb{E}[B(f_k)^2 B(f_l)^2] = \frac{\sigma^4 N^2}{4}(1 + 2\delta_{l,k})$$

We can compute the covariance of $|J(f_k)|^2$ using what we did above:

$$\operatorname{cov}(|J(f_k)|^2, |J(f_l)|^2) = \mathbb{E}[|J(f_k)|^2|J(f_l)|^2] - \mathbb{E}[|J(f_k)|^2]\mathbb{E}[|J(f_l)|^2]$$
$$= \frac{1}{N^2}\mathbb{E}[A(f_k)^2 A(f_l)^2 + B(f_k)^2 B(f_l)^2 + A(f_k)^2 B(f_l)^2 + B(f_k)^2 A(f_l)^2] - \sigma^4$$
$$= \frac{1}{N^2}\left(\frac{\sigma^4 N^2}{4}(1 + 2\delta_{l,k}) + \frac{\sigma^4 N^2}{4}(1 + 2\delta_{l,k}) + \frac{\sigma^4 N^2}{4} + \frac{\sigma^4 N^2}{4}\right) - \sigma^4$$

i.e. $\boxed{\operatorname{cov}(|J(f_k)|^2, |J(f_l)|^2) = \sigma^4(1 + \delta_{l,k}) - \sigma^4}$

In particular, for $l \neq k$, $\mathrm{cov}(|J(f_k)|^2, |J(f_l)|^2) = 0$. AS there is no correlation between the $|J(f_k)|^2$, it explains the randomness in the periodogram. No pattern can't be extracted from white noise.

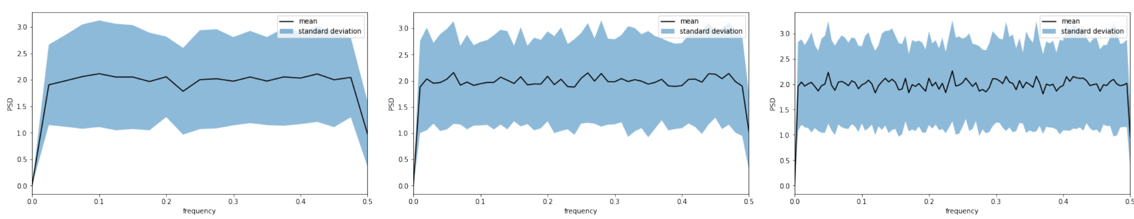
Question 9

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 4.

Answer 9



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

Figure 4: Barlett's periodograms of a Gaussian white noise (see Question 9).

The standard deviation of the PSD is around 0.75, while it was around 1.5 before, which is is consistent with the theoretic idea that the variance should be divided by 5.

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.

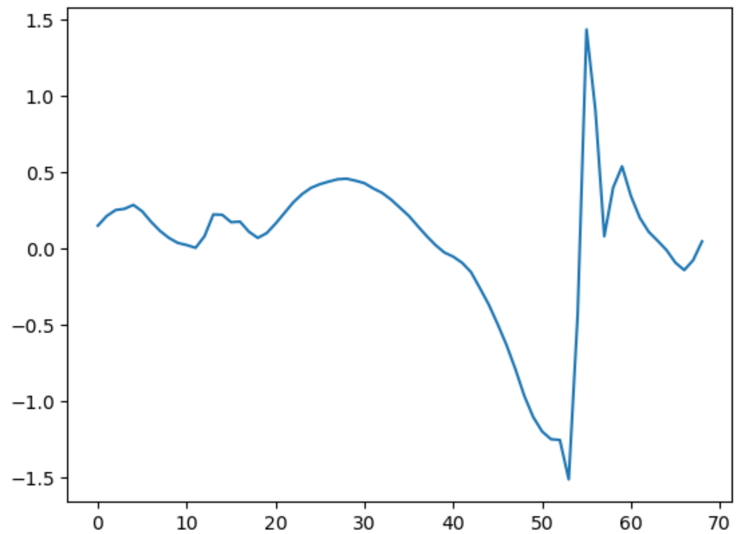
Answer 10

The optimal number of neighbors is **5**, with a validation f1 score of 0.741.
However, the f1-score on the test set is **0.513**. The classifier doesn't generalize very well on unseen data. We would not rely on the DTW+kNN approach for medical diagnosis.

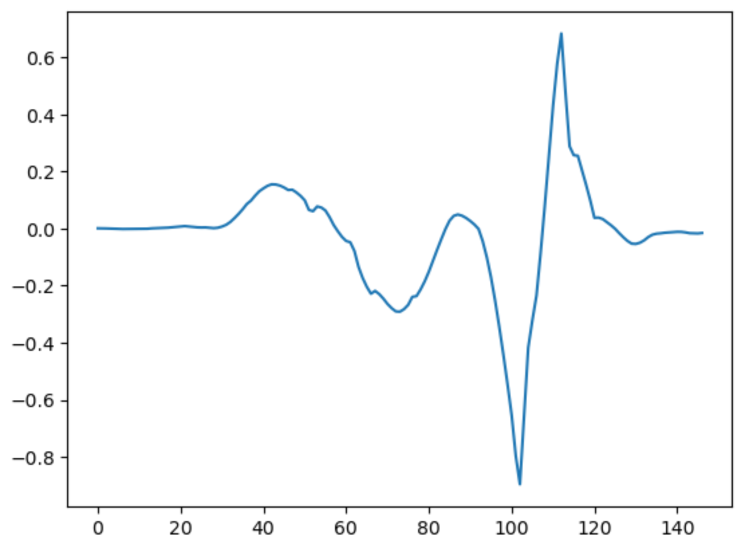
Question 11

Display on Figure 5 a badly classified step from each class (healthy/non-healthy).

Answer 11



Badly classified healthy step



Badly classified non-healthy step

Figure 5: Examples of badly classified steps (see Question 11).