Assignment 1 (ML for TS) - MVA

Fotios Kapotos fotiskapotos@gmail.com
Adonis Jamal adonis.jamal@student-cs.fr

October 30, 2025

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 Sunday 9th November 23:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname2.pdf` and
  `FirstnameLastname1_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_ValerioGuerrini.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

We denote by $F_\lambda(\beta)$ the Lasso objective. For every β,

$$F_\lambda(\beta) - F_\lambda(0) = \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1 - \frac{1}{2}\|y\|_2^2$$
$$= -y^\top X\beta + \frac{1}{2}\|X\beta\|_2^2 + \lambda\|\beta\|_1.$$

Or, using Hölder's inequality,

$$|y^\top X\beta| \;\leq\; \|X^\top y\|_\infty \,\|\beta\|_1.$$

Hence,

$$F_\lambda(\beta) - F_\lambda(0) \;\geq\; \left(\lambda - \|X^\top y\|_\infty\right)\|\beta\|_1 + \frac{1}{2}\|X\beta\|_2^2.$$

By noting that $\lambda_{\max} = \|X^\top y\|_\infty$, we can conclude that for all $\lambda \geq \lambda_{\max}$, the unique minimizer is $\beta = 0$.

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:

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

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

For a fixed dictionnary the problem becomes

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

Since convolution with a fixed atom $\mathbf{d}_k$ is linear in $\mathbf{z}_k$, there exists a matrix D_k such that $D_k\mathbf{z}_k = \mathbf{z}_k * \mathbf{d}_k$. Stacking all activations $\mathbf{z}_k$ into $\boldsymbol{\beta} = [\mathbf{z}_1^\top, \ldots, \mathbf{z}_K^\top]^\top$ and defining $X = [D_1\ D_2\ \cdots\ D_K]$, the problem becomes

$$\min_{\boldsymbol{\beta}} \|\mathbf{x} - X\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

which is exactly a Lasso regression with response $\mathbf{x}$ and design matrix X.

To answer the following question we just apply the result proved in the first question.

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:

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

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

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

Question 3

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

By definition of (zero-mean) Gaussian white noise, the samples (X_n) are i.i.d. with $\mathbb{E}[X_n] = 0$ and $\mathrm{Var}(X_n) = \sigma^2$.

The autocovariance function is

$$\gamma_X[k] \;=\; \mathbb{E}[X_n X_{n+k}] \;=\; \begin{cases} \sigma^2, & k = 0, \\ 0, & k \neq 0\,, \end{cases}$$

i.e. $\gamma_X[k] = \sigma^2\, \delta[k]$.

The power spectrum (power spectral density), defined as the discrete-time Fourier transform of $\gamma_X[k]$, is then

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} \gamma_X[k] e^{-j\omega k} = \sigma^2.$$

Therefore, the spectrum is flat (constant in ω), which motivates the term "white".

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

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

Thus $\hat{\gamma}(\tau)$ is biased for any finite N unless $\tau = 0$.

However,

$$\lim_{N\to\infty} \mathbb{E}[\hat{\gamma}(\tau)] = \lim_{N\to\infty} \frac{N-\tau}{N} \gamma(\tau) = \gamma(\tau),$$

so the estimator is asymptotically unbiased.

We can debias this estimator by replacing the division by N by $N - \tau$

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

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

Answer 5

We have

$$|J(f_k)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m \, e^{-2\pi i (k/N)(n-m)}.$$

Grouping terms with the same lag $\tau = n - m$ yields

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

where

$$\hat{\gamma}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} X_n X_{n+|\tau|}, \qquad \hat{\gamma}(-\tau) = \hat{\gamma}(\tau).$$

Hence, the periodogram is the discrete Fourier transform of the sample autocovariance sequence.

Recall that the power spectral density is

$$S(f) \;=\; \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, e^{-2\pi i f \tau / f_s},$$

From the first part,

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

Taking expectation,

$$\mathbb{E}\left[|J(f_k)|^2\right] = \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}[\hat{\gamma}(\tau)]\, e^{-2\pi i (k/N)\tau}.$$

From the previous question, $\mathbb{E}[\hat{\gamma}(\tau)] = \frac{N-|\tau|}{N}\gamma(\tau)$. (The absolute value is used here to take negative τ values) Thus,

$$\mathbb{E}\left[|J(f_k)|^2\right] = \sum_{\tau=-(N-1)}^{N-1} \frac{N-|\tau|}{N}\, \gamma(\tau)\, e^{-2\pi i (k/N)\tau}.$$

As $N \to \infty$, we have $\frac{N-|\tau|}{N} \to 1$ for each fixed τ, and the finite sum converges to the bilateral sum defining $S(f)$ evaluated at $f = f_k = f_s k/N$. Hence

$$\lim_{N\to\infty} \mathbb{E}\left[|J(f_k)|^2\right] = S(f_k).$$

Finally, for an arbitrary $f > 0$, let $f^{(N)}$ be the closest Fourier frequency f_k. Then $f^{(N)} \to f$ as $N \to \infty$, and by continuity of $S(f)$,

$$\lim_{N\to\infty} \mathbb{E}\left[|J(f^{(N)})|^2\right] = S(f).$$

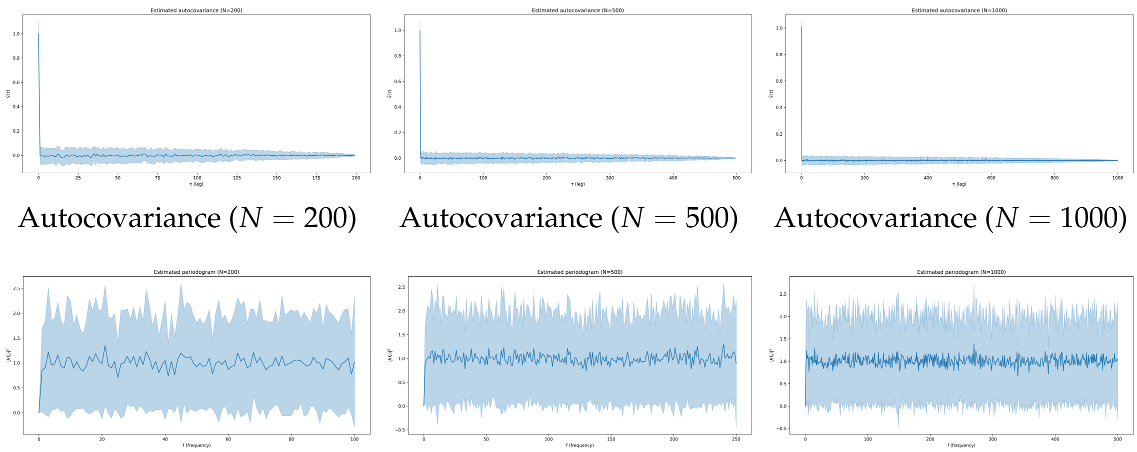
Therefore $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

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 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 all signal lengths $N \in \{200, 500, 1000\}$, the estimated sample autocovariances exhibit a sharp peak at lag $\tau = 0$, followed by values fluctuating around zero. This behavior is characteristic of white noise, whose samples are uncorrelated across time. The periodograms, representing $|J(f_k)|^2$ versus f_k, are approximately flat across all frequencies, reflecting the constant theoretical power spectral density of white noise. Averaging over 100 simulations smooths the mean spectrum, but the wide standard deviation band highlights that the periodogram is a highly variable estimator—asymptotically unbiased but not consistent. Increasing N densifies the frequency grid, yielding smoother curves but similar variance at each frequency.

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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{6}$$

  (Hint: if $\{Y_1, Y_2, Y_3, Y_4\}$ are four centered jointly Gaussian variables, then $\mathbb{E}[Y_1Y_2Y_3Y_4] = \mathbb{E}[Y_1Y_2]\mathbb{E}[Y_3Y_4] + \mathbb{E}[Y_1Y_3]\mathbb{E}[Y_2Y_4] + \mathbb{E}[Y_1Y_4]\mathbb{E}[Y_2Y_3]$.)

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

Answer 7

Let $\tau > 0$. We compute $\mathbb{E}[\hat{\gamma}^2(\tau)]$:

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

Using the hint, since we assume that X is a Gaussian process, we have:

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

We know that $\mathrm{Var}(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}^2(\tau)] - (\mathbb{E}[\hat{\gamma}(\tau)])^2$. From the previous exercise, we have $\mathbb{E}[\hat{\gamma}(\tau)] = \frac{N-\tau}{N}\gamma(\tau)$. Therefore,

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

By noticing that this term simplifies with the previous result, we can now express the variance as:

$$\mathrm{Var}(\hat{\gamma}(\tau)) = \frac{1}{N^2}\sum_{0\le n,m\le N-\tau-1} \gamma^2(n-m) + \gamma(n-m+\tau)\gamma(n-m-\tau)$$

The double summation only depends on the difference $k = n - m$. By changing the variables, we have for a given k, $(N-\tau-|k|)$ pairs (n, m). We get:

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

The obtained result is exactly what we were looking for.

From exercise 4, we know that $\lim_{N\to\infty} \mathbb{E}[\hat{\gamma}(\tau)] = \gamma(\tau)$, so the estimator is asymptotically unbiased.

We need to show that $\lim_{N\to\infty} \mathrm{Var}(\hat{\gamma}(\tau)) = 0$ to prove that $\hat{\gamma}(\tau)$ is consistent.

We bound $\mathrm{Var}(\hat{\gamma}(\tau))$:

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

Applying the Bienaymé-Tchebychev inequality to $\hat{\gamma}$ with $\epsilon > 0$:

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

Since we have proven that the variance tends to 0, we conclude that $\hat{\gamma}(\tau)$ is consistent.

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

Question 8

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

- 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

- Since $\{X_n\}$ is a Gaussian white noise, we have $\mathbb{E}[X_n] = 0$ and $\operatorname{Var}(X_n) = \sigma^2$. Let

$$A(f_k) = \sum_{n=0}^{N-1} X_n \cos\left(-\frac{2\pi kn}{N}\right), \quad B(f_k) = \sum_{n=0}^{N-1} X_n \sin\left(-\frac{2\pi kn}{N}\right).$$

  The mean of $A(f_k)$ is

$$\mathbb{E}[A(f_k)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n] \cos\left(-\frac{2\pi kn}{N}\right) = 0.$$

  For the variance,

$$\begin{aligned}
\operatorname{Var}[A(f_k)] &= \sum_{n=0}^{N-1} \operatorname{Var}(X_n) \cos^2\left(\frac{2\pi kn}{N}\right) \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos^2\left(\frac{2\pi kn}{N}\right) \\
&= \frac{\sigma^2}{2}\left(N + \Re\left(\sum_{n=0}^{N-1} e^{4i\pi kn/N}\right)\right) = \frac{N\sigma^2}{2}.
\end{aligned}$$

  Similarly,

$$\mathbb{E}[B(f_k)] = 0, \qquad \operatorname{Var}[B(f_k)] = \frac{N\sigma^2}{2}.$$

- The covariance between $A(f_k)$ and $B(f_k)$ is

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

  Thus $A(f_k)$ and $B(f_k)$ are uncorrelated, and since they are jointly Gaussian, they are independent.

  The periodogram value is

$$|J(f_k)|^2 = \frac{1}{N}\left(A^2(f_k) + B^2(f_k)\right).$$

  Since $A(f_k)$ and $B(f_k)$ are independent $\mathcal{N}(0, N\sigma^2/2)$ variables,

$$\frac{2|J(f_k)|^2}{\sigma^2} \sim \chi_2^2.$$

  Hence,

$$|J(f_k)|^2 \sim \sigma^2\,\mathrm{Exp}(1),$$

  for $0 < k < N/2$, while for $k = 0$ and $k = N/2$, $B(f_k) = 0$ and we have

$$|J(f_k)|^2 \sim \frac{\sigma^2}{2}\chi_1^2.$$

- From the exponential distribution,

$$\mathbb{E}[|J(f_k)|^2] = \sigma^2, \qquad \mathrm{Var}(|J(f_k)|^2) = \sigma^4.$$

  The variance does not depend on N. Thus, the periodogram $|J(f_k)|^2$ is not a consistent estimator of the true spectral density since its variance does not vanish as $N \to \infty$.

- The erratic behavior of the periodogram observed in Question 6 arises precisely from this non-vanishing variance.

- X_n is Gaussian white noise. $J(f_k)$ and $J(f_j)$ are jointly Gaussian, since they are linear combinations of jointly Gaussian variables X_n. In this case, for jointly Gaussian variables, being uncorrelated, i.e. $\mathrm{Cov}(J(f_k), J(f_j)) = 0$, implies that they are independent. We can therefore compute, for $j \neq j$:

$$\begin{aligned}\mathbb{E}[J(f_k)J(f_j)^*] &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathbb{E}[X_n X_m]e^{-i2\pi kn/N}e^{i2\pi jm/N} \\ &= \frac{\sigma^2}{N}\sum_{n=0}^{N-1}e^{-i2\pi(k-j)n/N} \\ &= 0\end{aligned}$$

Which is enough to prove that $J(f_k)$ and $J(f_j)$ are independent. This implies that $|J(f_k)|^2$ and $|J(f_j)|^2$ are also independent. We can thus explain the erratic behavior of the periodogram observed in Question 6 by the fact that each periodogram estimate $|J(f_k)|^2$ has a constant variance σ^4, which does not decrease with N. Moreover, the estimate at f_k being independent of the estimate at f_j for $k \neq j$ means that the value at one frequency does not give any information about the value at another, notably its neighbours. This means that at each frequency we will have high fluctuations independently of its neighbours, resulting in the erratic plot.

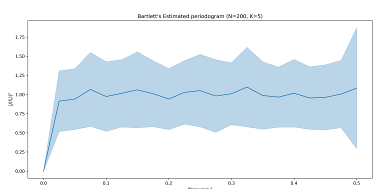
Question 9

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.

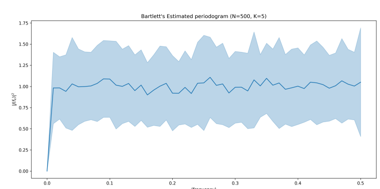
- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe?

Add your plots to Figure 2.

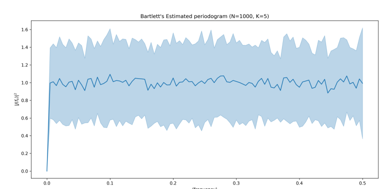
Answer 9



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

The standard deviation obtained using Bartlett's procedure is clearly much smaller than the one obtained in question 6. We have successfully reduced the variance of the estimator by taking the mean of several periodograms. Moreover, notice that the standard deviation is approximately 0.44, which corresponds to $\frac{1}{\sqrt{5}} \approx 0.447$. The standard deviation been divided by $\sqrt{K}$, where we set $K = 5$, and therefore the variance has indeed been 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 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.

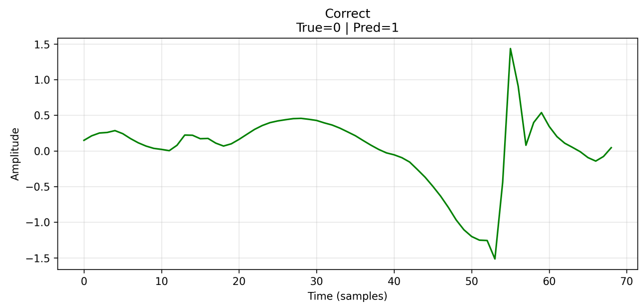
Answer 10

We find that the best numbers of neighbors is 2 or 4. Best F1 = 0.88 for k = 2; performance drops for larger k, showing that local neighborhood information is most discriminative in DTW space.

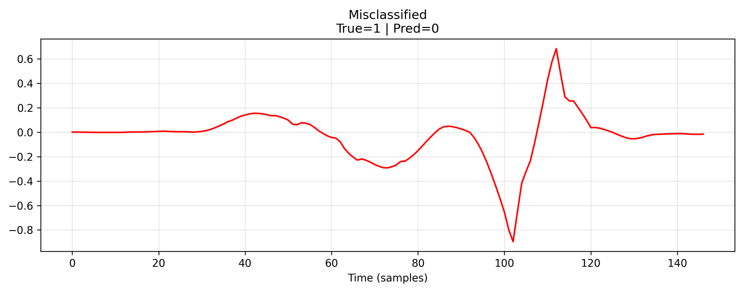
Question 11

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).