Assignment 1 (ML for TS) - MVA

Soël Megdoud Soel.Megdoud@gmail.com
Julien Delavande delavandejulien@gmail.com

October 28, 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 28th 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

$$\lambda\text{max} = |X^T y|\infty \tag{2}$$

Let's consider the KKT conditions for the Lasso problem, which state that the gradient of the objective function with respect to β must be zero. This leads to the following system of equations:

$$\frac{\partial f}{\partial \beta} = 0 \quad \text{which is equivalent to} \quad -X^T(y - X\beta) + \lambda, \text{sgn}(\beta) = \lambda, \text{sgn}(\beta) - X^T y + X^T X\beta = 0$$

We will now assume $\lambda > \lambda\text{max}$, where:

$$\lambda\text{max} = |X^T y|\infty$$

We will show by contradiction that $\beta = 0$ is the unique solution to the Lasso problem when $\lambda > \lambda\text{max}$. Suppose there exists an optimal solution $\beta \neq 0$.

Since $\beta \neq 0$, there exists some index i such that $\beta_i \neq 0$. Without loss of generality, let's assume $\beta_i > 0$ the case for $\beta_i < 0$ is similar.

We have:

$$\lambda, \text{sgn}(\beta_i) - X^T y_i + X^T X\beta_i = \lambda - X^T y_i + (X^T X\beta)i$$

Now, since we assumed $\lambda > |X^T y|\infty$, we have $\lambda > X^T y_i$. Therefore:

$$\lambda - X^T y_i > 0$$

Thus, the equation becomes:

$$\lambda - X^T y_i + X^T X\beta_i > X^T X\beta_i > 0$$

This implies that the KKT conditions are not satisfied for this non-zero β and that $\beta = 0$ is the unique solution for these λ.

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:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{3}$$

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

Answer 2

Given the convolutional dictionary learning problem:

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

For a fixed dictionary, each convolution $\mathbf{z}_k * \mathbf{d}_k$ is written as $D_k \mathbf{z}_k$, where $D_k \in \mathbb{R}^{n \times (n-L+1)}$ is a convolution matrix formed by sliding $\mathbf{d}_k$. The objective becomes:

$$\min_{\mathbf{z}} \|\mathbf{x} - D\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1$$

with $D = [D_1, \ldots, D_K]$, which is a Lasso regression. The threshold $\lambda_{\max} = \|D^\top \mathbf{x}\|_\infty$ ensures that for $\lambda > \lambda_{\max}$, the solution is $\mathbf{z} = 0$.

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 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:
$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{4}$$

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

For a Gaussian white noise process X_n, we have:

$$\mathbb{E}(X_n) = 0, \quad \text{and} \quad \mathbb{E}(X_n X_{n+\tau}) = \gamma(\tau) = \sigma^2 \delta(\tau),$$

where σ^2 is the variance of the process and $\delta(\tau)$ is the Dirac delta function.

Thus, the autocovariance function is:

$$\gamma(\tau) = \begin{cases} \sigma^2 & \text{if } \tau = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The power spectrum is the Fourier transform of the autocovariance function:

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

Since $\gamma(\tau) = \sigma^2 \delta(\tau)$, this simplifies to:

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

Thus, the power spectrum of white noise is constant across all frequencies, which is why it is called "white," analogous to white light containing all frequencies in equal measure.

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{5}$$

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

The sample autocovariance is defined as:

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

We want to show that $\hat{\gamma}(\tau)$ is biased but asymptotically unbiased. The true autocovariance is:

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

Now, consider the expectation of the sample autocovariance:

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

Thus, we have:

$$\mathbb{E}[\hat{\gamma}(\tau)] = \left(1 - \frac{\tau}{N}\right)\gamma(\tau).$$

This shows that $\hat{\gamma}(\tau)$ is a biased estimator because $\mathbb{E}[\hat{\gamma}(\tau)] \neq \gamma(\tau)$. However, as $N \to \infty$, the bias term $\frac{\tau}{N} \to 0$, making $\hat{\gamma}(\tau)$ **asymptotically unbiased**.

A simple way to de-bias this estimator is to scale $\hat{\gamma}(\tau)$ by $\frac{N}{N-\tau}$, yielding the corrected estimator:

$$\hat{\gamma}_{\text{debias}}(\tau) = \frac{N}{N-\tau}\hat{\gamma}(\tau).$$

Question 5

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

$$J(f) := (1/\sqrt{N})\sum_{n=0}^{N-1}X_n e^{-2\pi i f n/f_s} \tag{6}$$

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

- Step 1: Periodogram as a Function of Autocovariances

  The discrete Fourier transform (DFT) of the process $\{X_n\}_n$ is given by:

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

  The periodogram is the collection of values $|J(f_k)|^2$ for $k = 0, 1, \ldots, N/2$, where $f_k = f_s k/N$. Using the inverse Fourier transform, we can express $|J(f_k)|^2$ as a function of the sample autocovariances:

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

This shows that the periodogram is essentially the discrete Fourier transform of the sample autocovariances.

- Step 2: Asymptotic Unbiasedness of the Periodogram

  Let $f^{(N)}$ denote the Fourier frequency closest to f. We want to show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of the true power spectrum $S(f)$.

  Since $\hat{\gamma}(\tau)$ is asymptotically unbiased for $\gamma(\tau)$, and the periodogram is the Fourier transform of $\hat{\gamma}(\tau)$, it follows that the periodogram $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

  For large N, we have:

$$\mathbb{E}[|J(f^{(N)})|^2] \to S(f),$$

  which proves the asymptotic unbiasedness of the periodogram as an estimator of the power spectrum $S(f)$.

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?
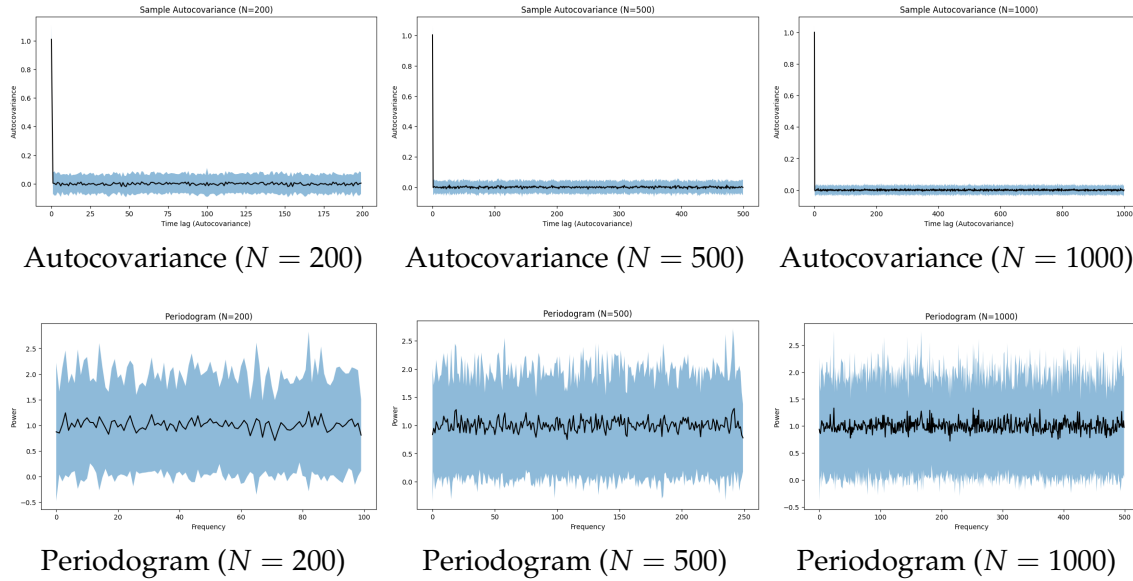
Add your plots to Figure 1.

Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

Answer 6

For both the sample autocovariances and periodograms, here are the key observations from the plots:

Sample Autocovariances:

- For each value of N (200, 500, 1000), the sample autocovariances have a sharp peak at $\tau = 0$, which is consistent with the fact that Gaussian white noise is uncorrelated.

- As N increases, the variance of the sample autocovariances decreases, which leads to a tighter confidence interval around the mean. This suggests that the estimate of the autocovariance becomes more reliable as the number of samples grows.

- The values of the autocovariances at lags $\tau \neq 0$ are close to zero, which aligns with the theoretical property of white noise being uncorrelated across time.

Periodograms:

- The periodograms show that for all values of N, the power is roughly constant across frequencies, which matches the expected flat power spectral density of Gaussian white noise.

- As N increases, the standard deviation of the periodogram decreases, meaning that the estimate of the power at each frequency becomes more stable with more data.

- Even though the periodogram is asymptotically unbiased, the variability (especially for smaller N) is quite high, which is reflected in the large fluctuations observed in the plots. This demonstrates the erratic behavior of the periodogram for small sample sizes.

In summary, both the sample autocovariance and periodogram plots become more reliable and less variable as N increases, which confirms the consistency of these estimators for large N.

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 $\tau > 0$

$$\text{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{7}$$

(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

- *Proof.* Let $X = \{X_n\}$ be a wide-sense stationary Gaussian process with zero mean, i.e., $\mathbb{E}[X_n] = 0$ for all n. The empirical estimator of the autocovariance at lag $\tau > 0$ is:

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

First, note that:

$$\mathbb{E}[\hat{\gamma}(\tau)] = \gamma(\tau),$$

since $\mathbb{E}[X_n X_{n+\tau}] = \gamma(\tau)$ due to stationarity.

To compute the variance $\text{var}(\hat{\gamma}(\tau))$, we have:

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

where

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

Using the hint for jointly Gaussian zero-mean variables, we get:

$$\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] = \gamma^2(\tau) + \gamma^2(n-m) + \gamma(n-m-\tau)\gamma(n-m+\tau).$$

Thus,

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

Let $k = n - m$. The number of pairs (n, m) such that $n - m = k$ is $(N - \tau - |k|)$ for $|k| \leq N - \tau - 1$. Therefore,

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

Subtracting $\gamma^2(\tau)$, we obtain:

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

Rewriting, we get:

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

$\square$

- Since $\gamma(k)$ is bounded and the term $\frac{\tau + |k|}{N} \to 0$ as $N \to \infty$, we have:

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

By the Chebyshev inequality, we have:

$$\forall \epsilon > 0 \quad P(|\hat{\gamma}(\tau) - \gamma(\tau)| \geq \epsilon) \leq \frac{\text{Var}(\hat{\gamma}(\tau))}{\epsilon^2}$$

Therefore, $\hat{\gamma}(\tau)$ converges in probability to $\gamma(\tau)$, and thus it is a consistent estimator.

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

Proof. **(a) Mean and Variance of $A(f_k)$ and $B(f_k)$:**

Given X_n are independent Gaussian random variables with mean zero and variance σ^2, we have:

$$A(f_k) = \sum_{n=0}^{N-1} X_n \cos\left(-2\pi f_k n / f_s\right),$$

9

$$B(f_k) = \sum_{n=0}^{N-1} X_n \sin\left(-2\pi f_k n / f_s\right),$$

where $f_k = \frac{f_s k}{N}$, for $k = 0, 1, \ldots, N/2$.

Mean:

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

$$\mathbb{E}[B(f_k)] = \sum_{n=0}^{N-1} \mathbb{E}[X_n] \sin\left(-2\pi f_k n / f_s\right) = 0.$$

Variance:

$$\mathrm{Var}[A(f_k)] = \mathbb{E}[A(f_k)^2] = \sum_{n=0}^{N-1} \mathbb{E}[X_n^2] \cos^2\left(2\pi k n / N\right)$$

$$= \sigma^2 \sum_{n=0}^{N-1} \cos^2\left(2\pi k n / N\right) = \frac{\sigma^2 N}{2},$$

since $\sum_{n=0}^{N-1} \cos^2\left(2\pi k n / N\right) = N/2$.

Similarly,

$$\mathrm{Var}[B(f_k)] = \frac{\sigma^2 N}{2}.$$

(b) Distribution of $|J(f_k)|^2$:

Since $A(f_k)$ and $B(f_k)$ are linear combinations of independent Gaussian variables, they are jointly Gaussian. The covariance between $A(f_k)$ and $B(f_k)$ is:

$$\mathrm{Cov}[A(f_k), B(f_k)] = \mathbb{E}[A(f_k) B(f_k)]$$

$$= \sum_{n=0}^{N-1} \mathbb{E}[X_n^2] \cos\left(2\pi k n / N\right) \sin\left(2\pi k n / N\right)$$

$$= \sigma^2 \sum_{n=0}^{N-1} \frac{\sin\left(4\pi k n / N\right)}{2} = 0,$$

since the sum of $\sin\left(4\pi k n / N\right)$ over one or multiple periods is zero.

Thus, $A(f_k)$ and $B(f_k)$ are independent Gaussian random variables with mean zero and variance $\sigma^2 N / 2$.

The periodogram at frequency f_k is:

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

Define $Z = \frac{\sqrt{2}}{\sigma\sqrt{N}}(A(f_k) + iB(f_k))$, then Z is a complex Gaussian random variable with independent standard normal real and imaginary parts. Therefore,

$$|Z|^2 = \left(\frac{\sqrt{2}}{\sigma\sqrt{N}} \right)^2 (A(f_k)^2 + B(f_k)^2) = \frac{2}{\sigma^2 N}(A(f_k)^2 + B(f_k)^2) \sim \chi_2^2,$$

where χ_2^2 denotes the chi-squared distribution with 2 degrees of freedom.

Thus,

$$|J(f_k)|^2 = \frac{\sigma^2}{N} \cdot \frac{|Z|^2}{2} \sim \frac{\sigma^2}{N} \cdot \mathrm{Exp}(1),$$

since $\chi_2^2/2$ follows an exponential distribution with mean 1.

(c) Variance of $|J(f_k)|^2$ and Non-Consistency:

The mean and variance of $|J(f_k)|^2$ are:

$$\mathbb{E}[|J(f_k)|^2] = \frac{\sigma^2}{N},$$

$$\mathrm{Var}[|J(f_k)|^2] = \left(\frac{\sigma^2}{N}\right)^2 \mathrm{Var}\left[\frac{\chi_2^2}{2}\right] = \left(\frac{\sigma^2}{N}\right)^2.$$

The variance does not tend to zero relative to the mean as $N \to \infty$:

$$\frac{\mathrm{Var}[|J(f_k)|^2]}{(\mathbb{E}[|J(f_k)|^2])^2} = 1.$$

Since the variance of $|J(f_k)|^2$ does not vanish as N increases, the periodogram $|J(f_k)|^2$ is not consistent.

(d) Covariance Between $|J(f_k)|^2$ and $|J(f_l)|^2$:

The covariance between $|J(f_k)|^2$ and $|J(f_l)|^2$ for $k \neq l$ is non-zero due to the overlapping sums in $A(f_k)$ and $A(f_l)$, and similarly for $B(f_k)$ and $B(f_l)$. This leads to:

$$\mathrm{Cov}[|J(f_k)|^2, |J(f_l)|^2] = \frac{\sigma^4}{N^2} \left| \sum_{n=0}^{N-1} e^{-2\pi i (f_k - f_l) n / f_s} \right|^2.$$

This covariance does not vanish as $N \to \infty$, causing the periodogram values at different frequencies to be correlated. This correlation leads to the erratic behavior observed in the periodogram, as fluctuations at one frequency affect those at others, preventing the periodogram from converging smoothly as N increases. □

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

When applying Bartlett's procedure, we divide the signal into $K = 5$ equal sections, compute the periodogram for each section, and then average them to reduce the variance. This method effectively smooths the periodogram and makes it more consistent, as the variance of the estimate is reduced by a factor of K.

Observations:

- The Bartlett periodograms for $N = 200, 500, 1000$ show a clear reduction in variance compared to the original periodogram without Bartlett's procedure.

- As N increases, the variance reduction becomes more noticeable, with smoother periodograms and reduced fluctuations.

- Despite the reduction in variance, some fluctuations remain, but they are much less erratic compared to the regular periodogram.

- Overall, Bartlett's procedure provides a much more stable estimate of the power spectrum, making it a consistent estimator for large N.
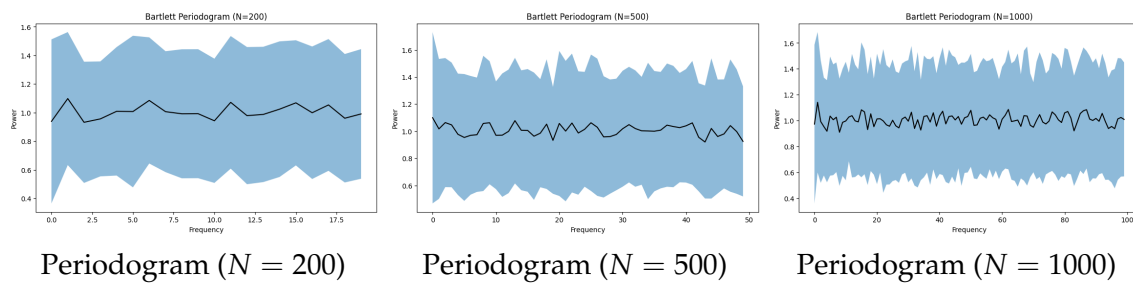

Periodogram ($N = 200$)   Periodogram ($N = 500$)   Periodogram ($N = 1000$)

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

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

After evaluating the different values of k (specifically, $k = 1, 2, 3, 4, 5, 7, 9$), we found that the optimal number of neighbors was $k = 3$, which yielded the highest average F1-score of approximately 0.88 on the validation sets.

Using this optimal number of neighbors, we retrained the k-NN classifier on the entire training set and evaluated its performance on the test set. The F1-score on the test set was approximately 0.45.

Comment: The drop in F1-score (0.88 on validation vs. 0.45 on test) likely arises from multiple factors: potential overfitting to training patterns, the high prevalence of class 1 in the test set (82.95%), and the complexity of DTW in handling diverse patterns. Adjusting class weights or incorporating more representative patterns could improve test performance.
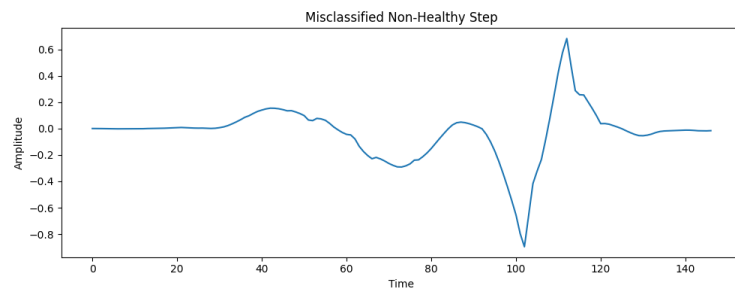
Question 11

Display on Figure 3 a badly classified step from each class (healthy/non-healthy).

Answer 11

We identified misclassified steps from both classes and plotted them in Figure 3.
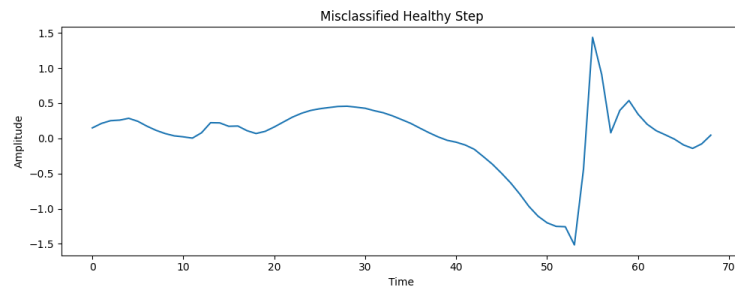
Badly classified healthy step

Badly classified non-healthy step

Figure 3: Examples of badly classified steps.