

Machine Learning for Time Series

Assignment 1

Janis Aiad janisaiad.ja@gmail.com
Jordi Baroja baroja.jordi@gmail.com

Instructor: Prof. Laurent Oudre
Deadline: 09/11/2025

ACADEMIC YEAR: 2025–2026

Contents

1 Convolution dictionary learning	2
2 Spectral feature	3
3 Data study	9
3.1 General information	9
3.2 Step classification with the dynamic time warping (DTW) distance	10

1 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 (1) is $\mathbf{0}_p$ (a p -dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

$$\lambda_{\max} = \max_{j \in \{1, \dots, p\}} \left| \sum_{i=1}^n X_{ij} y_i \right| \quad (2)$$

Proof: Let $f(\beta) = \frac{1}{2} \|y - X\beta\|_2^2$. The function is twice differentiable everywhere, and the Hessian is $\nabla^2 f(\beta) = X^T X \succeq 0$. So f is convex. We know the function satisfies, for every pair of points (x, y) : $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$, and so by letting $y = \beta$, $x = 0$, and computing the gradient of the function at 0 which is $\nabla f(0) = -X^T y$ we get the identity:

$$f(\beta) \geq f(0) - \langle X^T y, \beta \rangle \quad (3)$$

Now, if we let $\lambda > \lambda_{\max}$, we have that $\lambda |\beta_i|$ is strictly greater than the absolute value of the i 'th component of the dot product (this is true by definition of λ_{\max}), and so by adding all the terms we get

$$f(\beta) + \lambda \|\beta\|_1 > f(0) \quad \forall \beta \in \mathbb{R}^p$$

This proves what we wanted.

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:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k} \left\| \mathbf{x} - \sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{z}_k\|_1 \quad (4)$$

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

To proof the first statement, we will formulate the problem as a minimization over $z \in \mathbb{R}^{K(N-L+1)}$, where this vector is obtained from concatenating the K vectors $z_k \in \mathbb{R}^{N-L+1}$. We notice first that with this notation we have

$$\sum_{k=1}^K \|z_k\|_1 = \|z\|_1$$

To put this problem in the lasso regression form, we notice that the convolution $z_k * d_k = M_{d_k} z_k$, where $M_{d_k} \in \mathbb{R}^{N \times (N-L+1)}$ is the matrix obtained by reversing d_k and putting one more coefficient at each row in a circular way.

$$M_{d_k} = \begin{bmatrix} (d_k)_L & 0 & \cdots & \cdots & 0 \\ (d_k)_{L-1} & (d_k)_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & (d_k)_1 \end{bmatrix}$$

Now, if we take the vector $y = \sqrt{2}x$ and we take the matrix $M \in \mathbb{R}^{N \times K(N-L+1)}$ as

$$M = \sqrt{2} [M_{d_1} \mid M_{d_2} \mid \cdots \mid M_{d_K}]$$

We have that the problem is indeed a lassso regression minimization one.

The second statement is an immediate consequence of the first question, applied to the vector y and the matrix M .

2 Spectral feature

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) be a Gaussian white noise process with zero mean and variance σ^2 . It is weakly stationary (i.i.d.).

The autocovariance function is

$$\gamma(\tau) = \mathbb{E}[X_n X_{n+\tau}] = \begin{cases} \sigma^2 & \text{if } \tau = 0, \\ 0 & \text{otherwise,} \end{cases} \quad \text{so } \gamma(\tau) = \sigma^2 \delta_{\tau 0}.$$

The power spectrum (defined as $S(f) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau / f_s}$) is then

$$S(f) = \sum_{\tau} \sigma^2 \delta_{\tau 0} e^{-2\pi i f \tau / f_s} = \sigma^2,$$

which is constant (flat) for all frequencies f —so the term "white."

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

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

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

Taking the expectation of $\hat{\gamma}(\tau)$ for $\tau \geq 0$:

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

Since $\mathbb{E}[\hat{\gamma}(\tau)] = \frac{N-\tau}{N} \gamma(\tau) \neq \gamma(\tau)$ for $\tau > 0$, the estimator is biased. However, as $N \rightarrow \infty$, we have $\lim_{N \rightarrow \infty} \frac{N-\tau}{N} = 1$, so $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\gamma}(\tau)] = \gamma(\tau)$, showing it is asymptotically unbiased.

A simple way to de-bias the estimator is to use the normalization factor $1/(N-\tau)$ instead of $1/N$, giving:

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

which satisfies $\mathbb{E}[\tilde{\gamma}(\tau)] = \gamma(\tau)$ for all finite N and τ .

Question 5

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

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

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

Part 1: Express $|J(f_k)|^2$ via the sample autocovariances using DFT isometry.
 Let $f_k = f_s k / N$ and $J(f_k) = (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i k n / N}$. By Parseval/Plancherel (the DFT is unitary), products in time correspond to convolutions in frequency and vice versa. Compute

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

With the change of variables $\tau = n - m$, the admissible pairs contribute

$$|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1} \underbrace{\left(\frac{1}{N} \sum_{n=\max(0, \tau)}^{\min(N-1, N-1+\tau)} X_n X_{n-\tau} \right) e^{-2\pi i k \tau / N}}_{\hat{\gamma}(|\tau|) \text{ (by the given definition)}}.$$

So, with the biased sample autocovariance $\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}$ for $\tau \geq 0$ and $\hat{\gamma}(-\tau) := \hat{\gamma}(\tau)$, we obtain the finite identity:

$$|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(|\tau|) e^{-2\pi i k \tau / N}, \quad k = 0, \dots, \frac{N}{2}.$$

In words: the periodogram is the DFT of the (finite-lag, biased) sample autocovariance. This follows from the convolution/correlation theorem and the isometry of the DFT.

Part 2: Asymptotic unbiasedness of $|J(f^{(N)})|^2$ for $S(f)$, $f > 0$.

Taking expectations and using stationarity,

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

From the previous question, for fixed τ one has $\mathbb{E}[\hat{\gamma}(|\tau|)] = \frac{N-|\tau|}{N} \gamma(\tau)$, so $\mathbb{E}[\hat{\gamma}(|\tau|)] \rightarrow \gamma(\tau)$ as $N \rightarrow \infty$. Since $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$ (absolute summability) and $f^{(N)} \rightarrow f$ (Fourier grid spacing $f_s/N \rightarrow 0$), dominated convergence yields

$$\lim_{N \rightarrow \infty} \mathbb{E}[|J(f^{(N)})|^2] = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi i f \tau / f_s} = S(f), \quad f > 0.$$

So, the periodogram at the closest Fourier frequency is an asymptotically unbiased estimator of the power spectrum.

Question 6

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

- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs τ) for 100 simulations of X . Plot the average value as well as the average \pm , the standard deviation. What do you observe?
- For $N \in \{200, 500, 1000\}$, compute the *periodogram* ($|J(f_k)|^2$ vs f_k) for 100 simulations of X . Plot the average value as well as the average \pm , the standard deviation. What do you observe?

Add your plots to Figure 1.

Answer 6

What we can see in the autocovariance plots is that in the case of white noise it is just the Dirac function, as it can be theoretically shown easily. The estimation have also this shape, and an improtant thing we observe is that the variance indeed decreases as N increases, empirically showing what we have theoretically established.

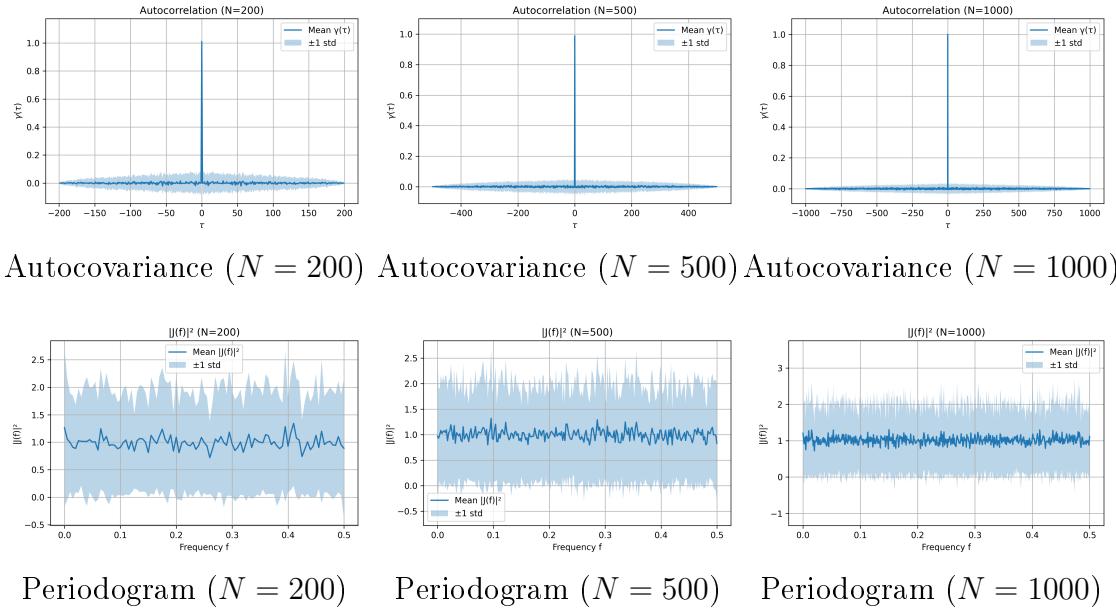


Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

In the case of periodograms, we can see that in mean they are close to being constant and equal to 1, as it can be shown is the case for white noise, but in this case the variance does not decrease as N increases. In fact, as we are computing more frequencies as N increases the graph is less smooth for big values of N .

Question 7

We want to show that the estimator $\hat{\gamma}(\tau)$ is consistent, ie 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 $\tau > 0$

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

(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

Part 1: Compute $\text{var}(\hat{\gamma}(\tau))$ for $\tau > 0$.

For $\tau \geq 0$, the variance is

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

From the previous exercise, $\mathbb{E}[\hat{\gamma}(\tau)] = \frac{N-\tau}{N} \gamma(\tau)$, so we need to compute

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

For centered jointly Gaussian variables $\{Y_1, Y_2, Y_3, Y_4\}$, Isserlis' theorem (Gaussian moment factorization) gives:

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

Applying this to $\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}]$ and using stationarity $\mathbb{E}[X_n X_m] = \gamma(m-n)$:

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

That is,

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

Changing variables to $k = m-n$ (so k ranges from $-(N-\tau-1)$ to $N-\tau-1$), and counting admissible pairs for fixed k :

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

Noticing that $\sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|)\gamma^2(\tau) = \frac{(N-\tau)^2}{N^2}\gamma^2(\tau)$, we get

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

So,

$$\begin{aligned} \text{var}(\hat{\gamma}(\tau)) &= \mathbb{E}[(\hat{\gamma}(\tau))^2] - \left(\frac{N-\tau}{N}\gamma(\tau) \right)^2 \\ &= \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|) [\gamma^2(k) + \gamma(k+\tau)\gamma(k-\tau)] \\ &= \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|k|}{N} \right) [\gamma^2(k) + \gamma(k+\tau)\gamma(k-\tau)], \end{aligned}$$

which is the desired expression after relabeling $k \rightarrow n$.

Part 2: Consistency.

For consistency, we need $\text{var}(\hat{\gamma}(\tau)) \rightarrow 0$ as $N \rightarrow \infty$. From absolute summability $\sum_{k \in \mathbb{Z}} |\gamma(k)| < \infty$, we have $\sum_{k \in \mathbb{Z}} |\gamma^2(k) + \gamma(k+\tau)\gamma(k-\tau)| \leq C < \infty$ for some constant C .

For fixed k , the weighting $(1 - \frac{\tau+|k|}{N}) \rightarrow 1$ as $N \rightarrow \infty$. Bounding the terms by dominated convergence and using that the sum has $O(N)$ terms while each term is $O(1/N)$:

$$\text{var}(\hat{\gamma}(\tau)) \leq \frac{C}{N} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Since $\mathbb{E}[\hat{\gamma}(\tau)] \rightarrow \gamma(\tau)$ (asymptotic unbiasedness) and $\text{var}(\hat{\gamma}(\tau)) \rightarrow 0$, we conclude that $\hat{\gamma}(\tau)$ converges in mean square (and so in probability) to $\gamma(\tau)$, i.e., $\hat{\gamma}(\tau)$ 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 $f = f_0, f_1, \dots, f_{N/2}$ 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

Means and variances of $A(f_k)$ and $B(f_k)$. Let $\theta_{k,n} = 2\pi f_k n / f_s = 2\pi k n / N$. For Gaussian white noise $X_n \sim \mathcal{N}(0, \sigma^2)$ i.i.d.,

$$\mathbb{E}[A(f_k)] = \mathbb{E}[B(f_k)] = 0, \quad \text{var}(A(f_k)) = \sigma^2 \sum_{n=0}^{N-1} \cos^2 \theta_{k,n}, \quad \text{var}(B(f_k)) = \sigma^2 \sum_{n=0}^{N-1} \sin^2 \theta_{k,n}.$$

For Fourier frequencies $k \in \{1, \dots, \frac{N}{2}-1\}$ one has $\sum_{n=0}^{N-1} \cos^2 \theta_{k,n} = \sum_{n=0}^{N-1} \sin^2 \theta_{k,n} = N/2$ and $\sum_{n=0}^{N-1} \sin \theta_{k,n} \cos \theta_{k,n} = 0$, we can see that in the following way:

$$0 = \sum_{n=0}^{N-1} e^{2i\theta_{k,n}} = \sum_{n=0}^{N-1} (\cos \theta_{k,n} + i \sin \theta_{k,n})^2 = \sum_{n=0}^{N-1} \cos^2 \theta_{k,n} - \sum_{n=0}^{N-1} \sin^2 \theta_{k,n} + 2i \sum_{n=0}^{N-1} \sin \theta_{k,n} \cos \theta_{k,n}$$

This implies the first two terms are equal, and by adding them together we get the value, and the last sum must be 0. So we get:

$$\text{var}(A(f_k)) = \text{var}(B(f_k)) = \frac{N\sigma^2}{2}, \quad \text{cov}(A(f_k), B(f_k)) = 0.$$

For $k = 0$ and (when N is even) $k = \frac{N}{2}$, $\sin \theta_{k,n} \equiv 0$ and $\cos^2 \theta_{k,n} \equiv 1$, so

$$\text{var}(A(f_0)) = \text{var}(A(f_{N/2})) = N\sigma^2, \quad \text{var}(B(f_0)) = \text{var}(B(f_{N/2})) = 0.$$

Distribution of $|J(f_k)|^2$. With $J(f_k) = (1/\sqrt{N})(A(f_k) + iB(f_k))$, for $k \in \{1, \dots, \frac{N}{2}-1\}$ we have

$$\text{Re } J(f_k), \text{Im } J(f_k) \stackrel{\text{indep.}}{\sim} \mathcal{N}\left(0, \frac{\sigma^2}{2}\right),$$

so

$$\frac{2}{\sigma^2} |J(f_k)|^2 \sim \chi_2^2 \iff |J(f_k)|^2 \stackrel{d}{=} \sigma^2 \text{Exp}(1),$$

and therefore

$$\mathbb{E}[|J(f_k)|^2] = \sigma^2, \quad \text{var}(|J(f_k)|^2) = \sigma^4.$$

For $k = 0$ (and $k = \frac{N}{2}$), $J(f_k)$ is real with $J(f_k) \sim \mathcal{N}(0, \sigma^2)$, hence

$$\frac{1}{\sigma^2} |J(f_k)|^2 \sim \chi_1^2, \quad \mathbb{E}[|J(f_k)|^2] = \sigma^2, \quad \text{var}(|J(f_k)|^2) = 2\sigma^4.$$

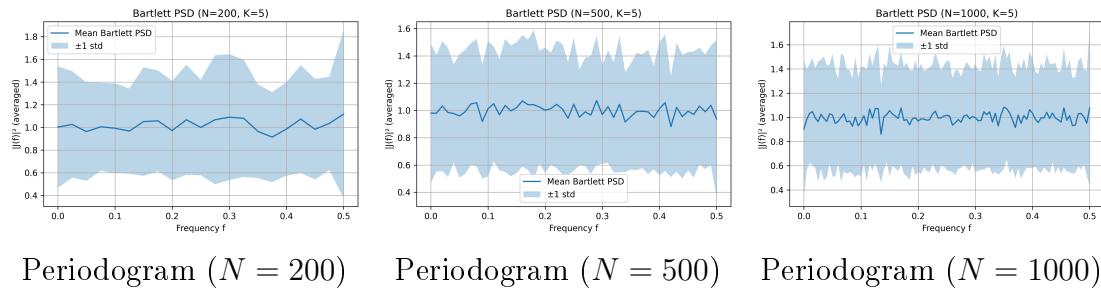


Figure 2: Bartlett's periodograms of a Gaussian white noise (see Question 9).

Non-consistency of the periodogram. Since $\text{var}(|J(f_k)|^2)$ does not decrease with N (it equals σ^4 for interior k and $2\sigma^4$ at $k = 0, \frac{N}{2}$), the periodogram is not consistent even though it is asymptotically unbiased.

Erratic behavior and covariance between ordinates. For Gaussian white noise, the DFT coefficients at distinct Fourier frequencies are independent (up to conjugate symmetry $J(f_{N-k}) = \overline{J(f_k)}$ for real signals), so for $k \neq \ell$ in $\{1, \dots, \frac{N}{2} - 1\}$,

$$\text{cov}(|J(f_k)|^2, |J(f_\ell)|^2) = 0.$$

Combined with large per-frequency variance, this independence explains the spiky, erratic appearance of the raw periodogram.

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 Bartlett's estimate (set $K = 5$). What do you observe?

Add your plots to Figure 2.

Answer 9

As we can see, the graphs now look very similar to the periodograms ones except that now the variance has decreased by a factor of 5, so the standard deviation decreases by a factor of $\sqrt{5}$. This makes the graphs look more smooth than before, but the shape remains the same.

3 Data study

3.1 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.

3.2 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 training the classifier for different values of $k = 1, 2, 3, 5, 7, 9, 11$, we have obtained that the best value is $k^* = 1$, meaning each new sample is classified according to the closest training example under the DTW distance. The F score we get is 0.88 which is quite high, so it could mean that DTW already provides a very sensitive similarity measure. However, taking into account that the training dataset is small, only 168 samples, and that the results with higher values of k do not decay too much it is possible that the model is overfitting. Indeed, when we train a classifier with $k = 1$ with the whole training dataset and we test it we only get an F score of 0.48, which is quite bad. We have checked and the training dataset contains roughly half healthy examples and half unhealthy ones, while in the test dataset we have an 83% of unhealthy, which makes it difficult to process, so the classifier has not fully learned the difference between the two types. We would suggest that we increase the size of the dataset and maybe add the other data on it, as of the 16 variables of the initial dataset we are only selecting 1.

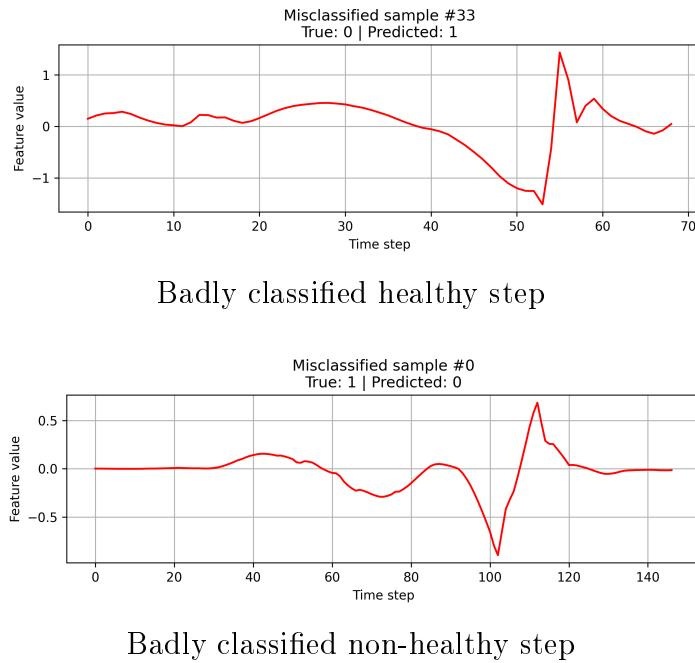


Figure 3: Examples of badly classified steps (see Question 11).

Question 11

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

Answer 11

Here we can see how both examples, one belonging to each type, visually look quite similar, although we have noticed that the unhealthy examples look more smooth in the first part of the graph. We conclude that there is still work to do in the improvement of performance of this model.