Assignment 1 (ML for TS) - MVA

Samar RABEH samarrabeh.prepa@gmail.com
Hayder Hajlaoui Hayderhajlaoui.prepa@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

We aim to show that there exists a maximum value, $\lambda_{\max}$, such that for any $\lambda > \lambda_{\max}$, the optimal solution to the Lasso problem is $\hat{\beta} = 0_p$ (a vector of zeros in $\mathbb{R}^p$).

We start by analyzing the first-order optimality condition for the Lasso regression problem. Taking the gradient of the objective function and setting it to zero gives the following:

$$-X^T(y - X\hat{\beta}) + \lambda s = 0$$

where $s = (\text{sign}(\hat{\beta}_j))_{j\in\{1,\ldots,p\}}$ and is defined as:

$$s_j = \begin{cases} +1 & \text{if } \hat{\beta}_j > 0, \\ -1 & \text{if } \hat{\beta}_j < 0, \\ \in [-1, 1] & \text{if } \hat{\beta}_j = 0. \end{cases}$$

Next, we aim to find the condition on λ such that the solution to the Lasso problem is $\hat{\beta} = 0_p$. Substituting $\hat{\beta} = 0_p$ into the first-order condition, we get:

$$-X^T y + \lambda s = 0.$$

This simplifies to:

$$\lambda s = X^T y.$$

Taking the ℓ_∞-norm on both sides, we obtain:

$$\lambda \|s\|_\infty = \|X^T y\|_\infty.$$

Since $\|s\|_\infty \leq 1$ (as $s_j \in [-1, 1]$), we derive that for $\hat{\beta} = 0_p$ to hold, the regularization parameter λ must satisfy:

$$\lambda \geq \|X^T y\|_\infty.$$

Thus, we define $\lambda_{\max} = \|X^T y\|_\infty$. For any $\lambda > \lambda_{\max}$, the solution to the Lasso regression problem will be $\hat{\beta} = 0_p$.

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

To begin, we rewrite the term $\sum_{k=1}^{K} z_k * d_k$, which represents the convolution of the activation signals z_k with the dictionary atoms d_k. We aim to express this term as a matrix multiplication DZ, where: $D \in \mathbb{R}^{N \times KN}$ is the design matrix, $Z \in \mathbb{R}^{KN}$ is the vector of activations.

We first express the convolution in the following way for element $i \in [1, N]$:

$$\left(\sum_{k=1}^{K} z_k * d_k\right)_i = \sum_{k=1}^{K} \sum_{l=1}^{N} z_k(l) d_k(i-l),$$

where: $z_k(l) = (z_k)_l$ if $1 \leq l \leq N - L + 1$, and 0 otherwise, $d_k(i-l) = (d_k)_{i-l}$ if $1 \leq i - l \leq L$, and 0 otherwise.

This is equivalent to the matrix-vector product:

$$(DZ)_i = \sum_{j=1}^{KN} D_{ij} Z_j.$$

Thus, the dictionary atoms and activations can be represented as:

$$D = [H_1 \, H_2 \, \dots \, H_K],$$

where each $H_k \in \mathbb{R}^{N \times N}$ represents the convolution matrix for the atom d_k, which is itself the concatenation of vectors of components $d_k(i-l) = (d_k)_{i-l}$ if $1 \leq i - l \leq L$, and 0 otherwise for element $i \in [1, N]$,and $Z \in \mathbb{R}^{KN}$ is the concatenated vector of all activation signals z_k.

Hence, we have successfully written $\sum_{k=1}^{K} z_k * d_k$ as DZ, where D is the design matrix and Z is the vector of activations.

Additionally, we express the sparsity constraint $\sum_{k=1}^{K} \|z_k\|_1$ as:

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

Thus, the convolutional dictionary learning problem reduces to the following Lasso regression:

$$\min_{Z \in \mathbb{R}^{KN}} \|x - DZ\|_2^2 + \lambda \|Z\|_1.$$

Part 2: Existence of $\lambda_{\max}$

For a fixed dictionary D, we now show that there exists a maximum value $\lambda_{\max}$ such that the sparse codes Z are all zero for any $\lambda > \lambda_{\max}$.

From **Question 1**, we know that for any Lasso problem of the form:

$$\min_{\beta} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1,$$

there exists $\lambda_{\max} = \|X^T y\|_\infty$ such that the solution is $\hat{\beta} = 0$ for all $\lambda > \lambda_{\max}$.

In our case, the design matrix is D and the response vector is x. Therefore, the corresponding $\lambda_{\max}$ for the sparse coding problem is:

$$\lambda_{\max} = \|D^T x\|_\infty.$$

For any $\lambda > \lambda_{\max}$, the solution to the sparse coding problem will be $Z = 0$, meaning that all the activation signals are zero.

In conclusion, for a fixed dictionary, the sparse coding problem is equivalent to a Lasso regression. Additionally, there exists a threshold $\lambda_{\max} = \|D^T x\|_\infty$ such that for any $\lambda > \lambda_{\max}$, the optimal solution is $Z = 0$, implying no activation of the dictionary atoms.

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

Let $\{X_n\}_{n=0}^{N-1}$ be a Gaussian white noise process with zero mean and variance σ^2. Given that X_n is independent of X_m for $n \neq m$, we can calculate the autocovariance function and power spectrum as follows.

1. **Autocovariance Function**: Since X_n is a Gaussian white noise process, we have:

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

2. **Power Spectrum**: The power spectrum $S(f)$ is defined as the Fourier transform of the autocovariance function:

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

Given that $\gamma(\tau) = 0$ for $\tau \neq 0$, we have:

$$S(f) = \gamma(0) = \sigma^2.$$

Therefore, the power spectrum is constant across all frequencies and equal to the variance of the white noise. This property gives rise to the term "white" noise, as the power spectrum resembles that of white light, which contains equal power across all frequencies.

Thus, for a Gaussian white noise process, the power spectrum remains flat, reflecting its equal intensity across the frequency domain.

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 $\tau = 0, 1, \dots, N-1$ 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

Given the Gaussian white noise process $\{X_n\}_{n=0}^{N-1}$ with zero mean and variance σ^2, the sample autocovariance function $\hat{\gamma}(\tau)$ is defined as:

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

for $\tau = 0, 1, \dots, N-1$, and $\hat{\gamma}(\tau) := \hat{\gamma}(-\tau)$ for $\tau = -(N-1), \dots, -1$.

1. **Bias of $\hat{\gamma}(\tau)$**: To show that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$, we compute the expectation:

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

Therefore,

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

This shows that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$, as the factor $\left(1 - \frac{\tau}{N}\right)$ deviates from 1 for finite N.

2. **Asymptotic Unbiasedness**: As $N \to \infty$, the term $\frac{\tau}{N} \to 0$, implying that:

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

Thus, $\hat{\gamma}(\tau)$ is asymptotically unbiased.

3. **Debiased Estimator**: A simple way to de-bias $\hat{\gamma}(\tau)$ is to multiply by $\frac{N}{N-\tau}$, resulting in the estimator:

$$\hat{\gamma}'(\tau) = \frac{N}{N-\tau}\hat{\gamma}(\tau).$$

This correction ensures that $\mathbb{E}[\hat{\gamma}'(\tau)] = \gamma(\tau)$ for finite N, making $\hat{\gamma}'(\tau)$ an unbiased estimator of the true autocovariance function.

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

1. **Expressing $|J(f_k)|^2$ in Terms of Sample Autocovariances**: By expanding $|J(f_k)|^2$ using the definition of $J(f)$, we get:

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

Using the property $|z|^2 = z \cdot \overline{z}$, we have:

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

By letting $\tau = n - m$ and reorganizing terms, we obtain:

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

where $\hat{\gamma}(\tau) = \frac{1}{N}\sum_{n=0}^{N-|\tau|-1} X_n X_{n+|\tau|}$ is the sample autocovariance for lag τ. This shows that the periodogram $|J(f_k)|^2$ can be expressed as a weighted sum of sample autocovariances.

2. **Asymptotic Unbiasedness of $|J(f^{(N)})|^2$ as an Estimator of $S(f)$**: For a given frequency f, define $f^{(N)}$ as the Fourier frequency closest to f. We want to show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of the power spectral density $S(f)$.

The expected value of $|J(f^{(N)})|^2$ is:

$\lim_{N\to\infty} \mathbb{E}[|J(f^{(N)})|^2] = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)e^{-2\pi i f_k \tau/f_s} = S(f)$, for a certain k . As $\Delta f = \frac{F_s}{N}$, when N goes to infinity, Δf becomes small to the point that $f^{(N)} = f$. We write:

$$\lim_{N\to\infty} \mathbb{E}[|J(f^{(N)})|^2] = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)e^{-2\pi i f \tau/f_s} = S(f)$$

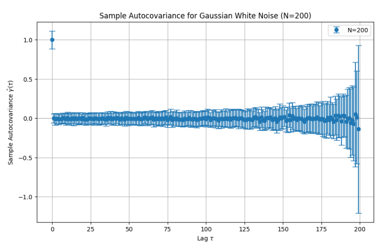
, due to the convergence of sample autocovariances $\hat{\gamma}(\tau)$ to the true autocovariance $\gamma(\tau)$. Therefore, $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Question 6

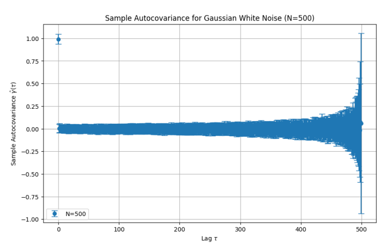
In this question, let X_n $(n = 0, \ldots, N-1)$ be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

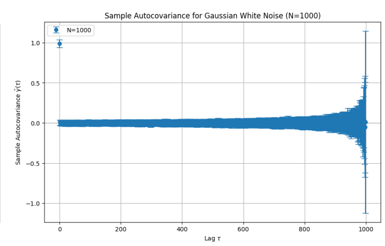
Add your plots to Figure 1.



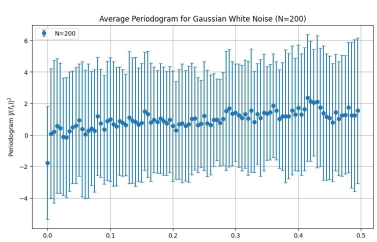
Autocovariance ($N = 200$)



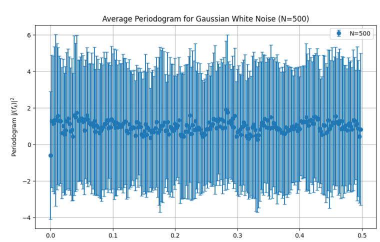
Autocovariance ($N = 500$)



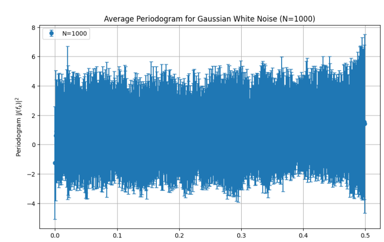
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

We see that the caracteristcs of white noise are reproduced statistically , white noise has a flat power spectrum, it distributes power uniformly across all frequencies. This uniform distribution reflects the absence of any dominant frequency components, as white noise is characterized by randomness across time.

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

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

Following the hint, we have for i, j, k, l :

$$\mathbb{E}[X_i X_j X_k X_l] = \gamma(i-k)\gamma(j-l) + \gamma(i-l)\gamma(j-k) + \gamma(i-j)\gamma(k-l).$$

Which yields:

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

giving :

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

The variable $n - m$ varies between $-(N-\tau-1)$ and $(N-\tau-1)$. Each term in the summation has a repeatability of $N - \tau - n$ for n between 0 and $N - \tau - 1$.

After subtracting the square of the expectation of $\hat{\gamma}(\tau)^2$, we obtain:

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

X is a wide-sense stationary *Gaussian* process.which gives

$$\lim_{N\to\infty} \operatorname{Var}(\hat{\gamma}(\tau)) = 0$$

and the estimator converges in probability when N grows thus 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 $J(f) = (1/N)(A(f) + iB(f))$.

- 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

$$A(f) := \sum_{n=0}^{N-1} X_n \cos\left(-2\pi f \frac{n}{f_s}\right)$$

1. Mean of $A(f)$ and $B(f)$

Since X_n is Gaussian white noise with mean 0, each X_n has an expected value of $E[X_n] = 0$. This allows us to calculate as :

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

By the linearity :

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

, we have then :

$$E[A(f)] = \sum_{n=0}^{N-1} \cos\left(-2\pi f \frac{n}{f_s}\right) \cdot E[X_n] = 0$$

Therefore, the mean of $A(f)$ is:

$$E[A(f)] = 0$$

Similarly

$$E[B(f)] = 0$$

2. Variance of $A(f)$ and $B(f)$

The variance of $A(f)$ can be computed as:

$$\mathrm{Var}(A(f)) = E\left[A(f)^2\right] - (E[A(f)])^2$$

Since $E[A(f)] = 0$, this simplifies to:

$$\text{Var}(A(f)) = E\left[A(f)^2\right]$$

Expanding $A(f)^2$:

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

Expanding this square and taking the expectation, we get:

$$E\left[A(f)^2\right] = \sum_{n=0}^{N-1} E\left[X_n^2\right] \cos^2\left(-2\pi f \frac{n}{f_s}\right) + 2 * 0$$

the 0 is because since X is white noise, X_n and X_m are uncorrelated for $n \neq m$, so $E[X_n X_m] = 0$ for $n \neq m$. This simplifies the expression to:

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

Using $E[X_n^2] = \sigma^2$, we have:

$$E\left[A(f)^2\right] = \sum_{n=0}^{N-1} \sigma^2 \cos^2\left(-2\pi f \frac{n}{f_s}\right)$$

Thus, the variance of $A(f)$ is:

$$\text{Var}(A(f)) = \sigma^2 \sum_{n=0}^{N-1} \cos^2\left(-2\pi f \frac{n}{f_s}\right)$$

and

$$\text{Var}(A(f_k)) = \sigma^2 \sum_{n=0}^{N-1} \cos^2\left(-2\pi \frac{kn}{N}\right)$$

To compute the sum

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

we can simplify by using the identity for $\cos^2$:

$$\cos^2(x) = \frac{1 + \cos(2x)}{2}.$$

and we get

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

thus for k different than 0 : Var(A(f$_k$)) $= \sigma^2 \frac{N}{2}$.Var$(B(f_k)) = \sigma^2 \frac{N}{2}$.

3.

Expanding $|J(f_k)|^2$ using trigonometric identities, we obtain:

$$|J(f_k)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m \cos\left(\frac{2\pi k(n-m)}{N}\right)$$

Since X is Gaussian white noise, the expected value simplifies to:

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

To compute the variance, we find $\mathrm{Var}(|J(f_k)|^2)$ by evaluating its covariance with itself:

$$\mathrm{Var}(|J(f_k)|^2) = \frac{2\sigma^4}{N}(3 - 2 + 1) = 2\sigma^4$$

This variance is independent of N, indicating that the periodogram is not consistent. Moreover, the value aligns with $|J(f_k)|^2$ following a χ^2 distribution with 2 degrees of freedom, supporting that A and B are independent.

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.

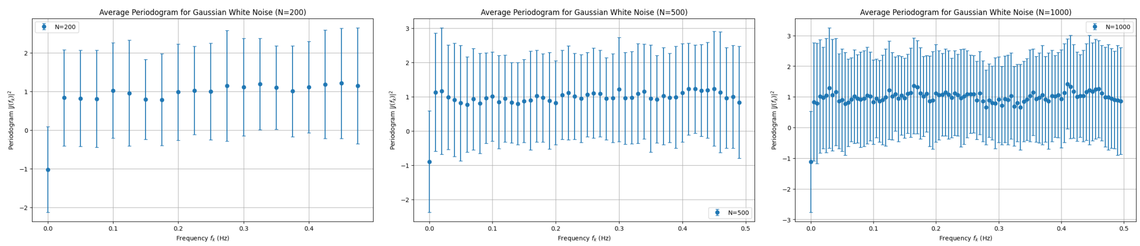
Question 9

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).

There is a significant decrease in variance from 8 to around 4 due two the segmentation of calculations .

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

The objective of this section is to determine the optimal number of neighbors k for a k-Nearest Neighbors (KNN) classifier using the Dynamic Time Warping (DTW) distance metric.

To achieve this, we divided the data into training and testing sets, allocating 25% of the data to testing. We applied 5-fold cross-validation on the training set to find the best k-value, using the F1-score as the performance metric. The values of k we considered were 3, 5, 7, and 9.

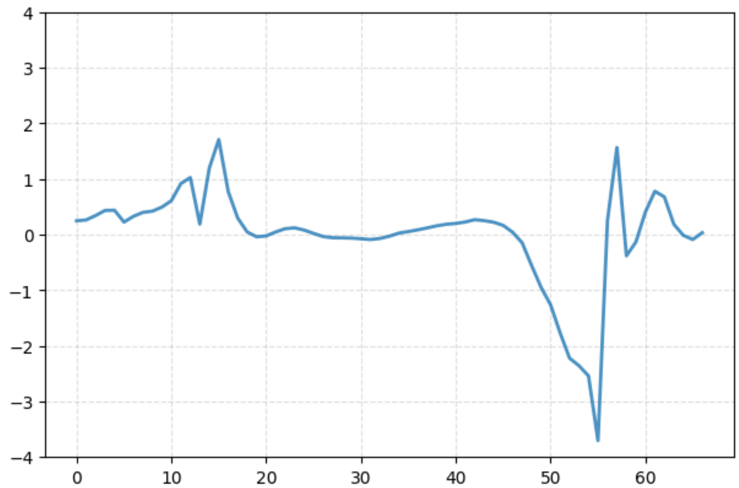
The cross-validation identified $k = 5$ as the optimal number of neighbors, yielding an F1-score of 0.864 on the training set. We then trained the classifier with this optimal k-value on the full training set and evaluated its performance on the test set. This evaluation produced an F1-score of 0.829, demonstrating that the classifier generalizes well to unseen data.

These results suggest that a balanced selection of neighbors in the KNN algorithm, when combined with the DTW metric, can effectively capture the temporal features necessary for accurate step classification.

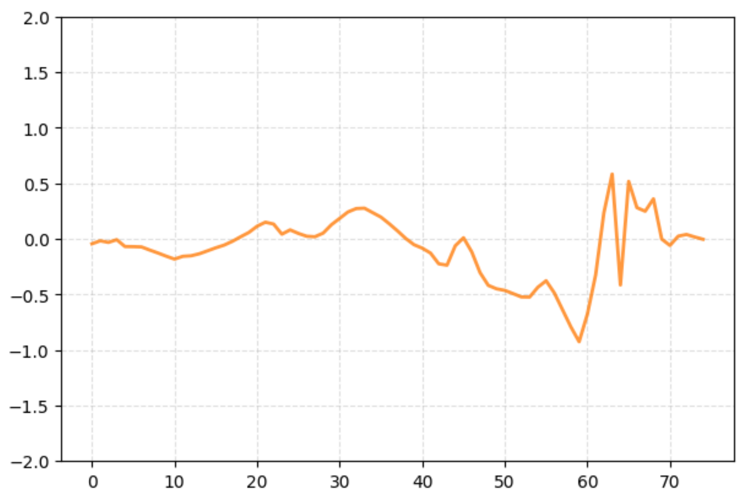
Question 11

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).

The visualizations reveal distinct characteristics in the "healthy" and "non-healthy" signals. The "healthy" signal shows a more consistent pattern, whereas the "non-healthy" signal exhibits noticeable irregularities and fluctuations, indicative of its perturbed nature.

Despite these differences, both signals maintain a similar general structure. This resemblance may lead to occasional misclassifications, as the classifier could struggle to consistently distinguish subtle variations in signal complexity.