

Assignment 3 (ML for TS) - MVA 2022/2023

Zakaria EL FOUNOUN zakaria.el-founoun@student-cs.fr

May 21, 2023

1 Introduction

Objective. The goal is to implement (i) a signal processing pipeline with a change-point detection method and (ii) wavelets for graph signals.

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 Friday 7th April 11:59 PM.
- Rename your report and notebook as follows:
FirstnameLastname1_FirstnameLastname1.pdf and
FirstnameLastname2_FirstnameLastname2.ipynb.
For instance, LaurentOudre_CharlesTruong.pdf.
- Upload your report (PDF file) and notebook (IPYNB file) using this link:
<https://www.dropbox.com/request/rmETjrLAH9Li3pf8JvOt>.

2 Dual-tone multi-frequency signaling (DTMF)

In the last tutorial, we started designing an algorithm to infer from a sound signal the sequence of symbols encoded with DTMF.

Question 1

Finalize this procedure—in particular, find the best hyperparameters. Describe in 5 to 10 lines your methodology and the calibration procedure (give the hyperparameter values).

Answer 1

In our case the main hyperparameter we focused on is the number of change points or breakpoints. We used a for loop and tried different integers from 1 to 49, and the for loop iterates over these values to compute the approximation error for each possible number of breakpoints. The error is giving by MultivariateNormalCost as implemented in the last tutorial.

The effect of tuning this hyperparameter is to control the granularity of the segmentation. A smaller number of breakpoints will result in fewer segments with larger duration, whereas a larger number of breakpoints will result in more segments with smaller duration. The choice of the optimal number of breakpoints depends on the specific signal and the trade-off between segmentation accuracy and computational complexity. The KneeLocator class is used to automatically determine the optimal number of breakpoints based on the knee or elbow point in the approximation error curve.

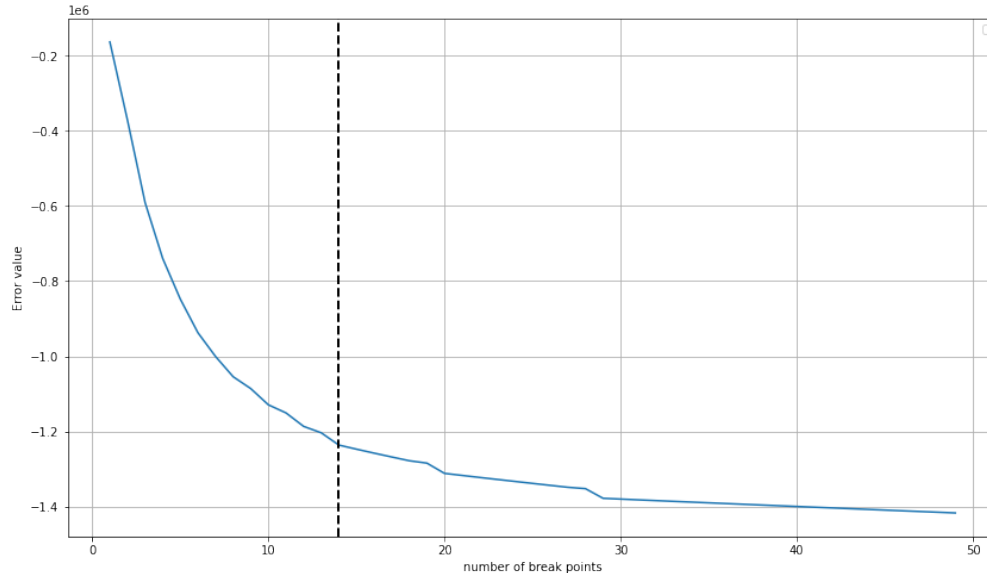


Figure 1: Error vs number of break points

The best break point is : 14

Question 2

What are the two symbolic sequences encoded in the provided signals?

Answer 2

- Sequence 1:
- Sequence 2:

3 Wavelet transform for graph signals

Let G be a graph defined a set of n nodes V and a set of edges E . A specific node is denoted by v and a specific edge, by e . The eigenvalues and eigenvectors of the graph Laplacian L are $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and u_1, u_2, \dots, u_n respectively.

For a signal $f \in \mathbb{R}^n$, the Graph Wavelet Transform (GWT) of f is $W_f : \{1, \dots, M\} \times V \longrightarrow \mathbb{R}$:

$$W_f(m, v) := \sum_{l=1}^n \hat{g}_m(\lambda_l) \hat{f}_l u_l(v) \quad (1)$$

where $\hat{f} = [\hat{f}_1, \dots, \hat{f}_n]$ is the Fourier transform of f and \hat{g}_m are M kernel functions. The number M of scales is a user-defined parameter and is set to $M := 9$ in the following. Several designs are available for the \hat{g}_m ; here, we use the Spectrum Adapted Graph Wavelets (SAGW). Formally, each kernel \hat{g}_m is such that

$$\hat{g}_m(\lambda) := \hat{g}^U(\lambda - am) \quad (0 \leq \lambda \leq \lambda_n) \quad (2)$$

where $a := \lambda_n / (M + 1 - R)$,

$$\hat{g}^U(\lambda) := \frac{1}{2} \left[1 + \cos \left(2\pi \left(\frac{\lambda}{aR} + \frac{1}{2} \right) \right) \right] \mathbb{1}(-Ra \leq \lambda < 0) \quad (3)$$

and $R > 0$ is defined by the user.

Question 3

Plot the kernel functions \hat{g}_m for $R = 1$, $R = 3$ and $R = 5$ (take $\lambda_n = 12$) on Figure 2. What is the influence of R ?

Answer 3

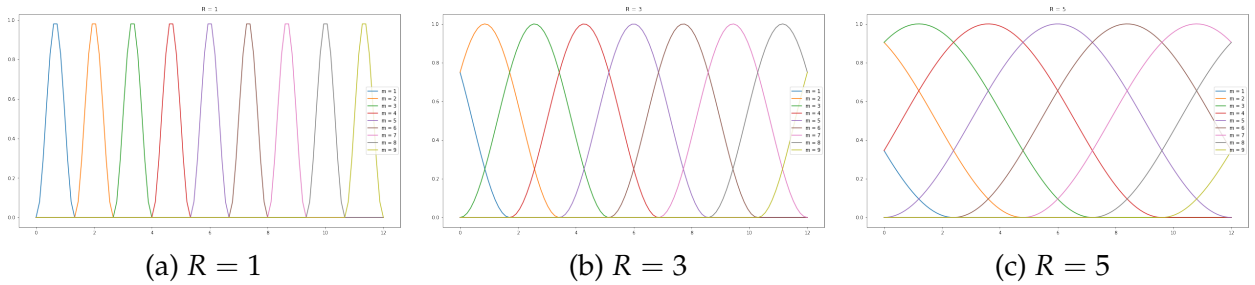


Figure 2: The SAGW kernels functions

Influence of R : The value of parameter R has two effects, it determines the degree of overlap between the shifted kernels. As R increases, the overlap between the kernels also increases and the the kernels become more wrapped around each other. The second effect is the width of the kernel, as R increased the width also increase.

We will study the Molene data set (the one we used in the last tutorial). The signal is the temperature.

Question 4

Construct the graph using the distance matrix and exponential smoothing (use the median heuristics for the bandwidth parameter).

- Remove all stations with missing values in the temperature.
- Choose the minimum threshold so that the network is connected and the average degree is at least 3.
- What is the time where the signal is the least smooth?
- What is the time where the signal is the smoothest?

Answer 4

The stations with missing values are:

'BATZ', 'BEG_MEIL', 'CAMARET', 'PLOUGONVELIN', 'RIEC SUR BELON', 'ST NAZAIRE-MONTOIR', 'PLOUAY-SA', 'VANNES-MEUCON', 'LANNAERO', 'PLOUDALMEZEAU', 'LANDIVISIAU', 'SIZUN', 'QUIMPER', 'OUESSANT-STIFF', 'LANVEOC', 'ARZAL', 'BREST-GUIPAVAS', 'BRIGNOGAN'

The threshold is equal to 0.83

The signal is the least smooth at 2014-01-21 06:00:00

The signal is the smoothest at 2014-01-24 19:00:00

Question 5

(For the remainder, set $R = 3$ for all wavelet transforms.)

For each node v , the vector $[W_f(1, v), W_f(2, v), \dots, W_f(M, v)]$ can be used as a vector of features. We can for instance classify nodes into low / medium / high frequency:

- a node is considered low frequency if the scales $m \in \{1, 2, 3\}$ contain most of the energy,
- a node is considered medium frequency if the scales $m \in \{4, 5, 6\}$ contain most of the energy,
- a node is considered high frequency if the scales $m \in \{6, 7, 9\}$ contain most of the energy.

For both signals from the previous question (smoothest and least smooth) as well as the first available timestamp, apply this procedure and display on the map the result (one colour per class).

Answer 5

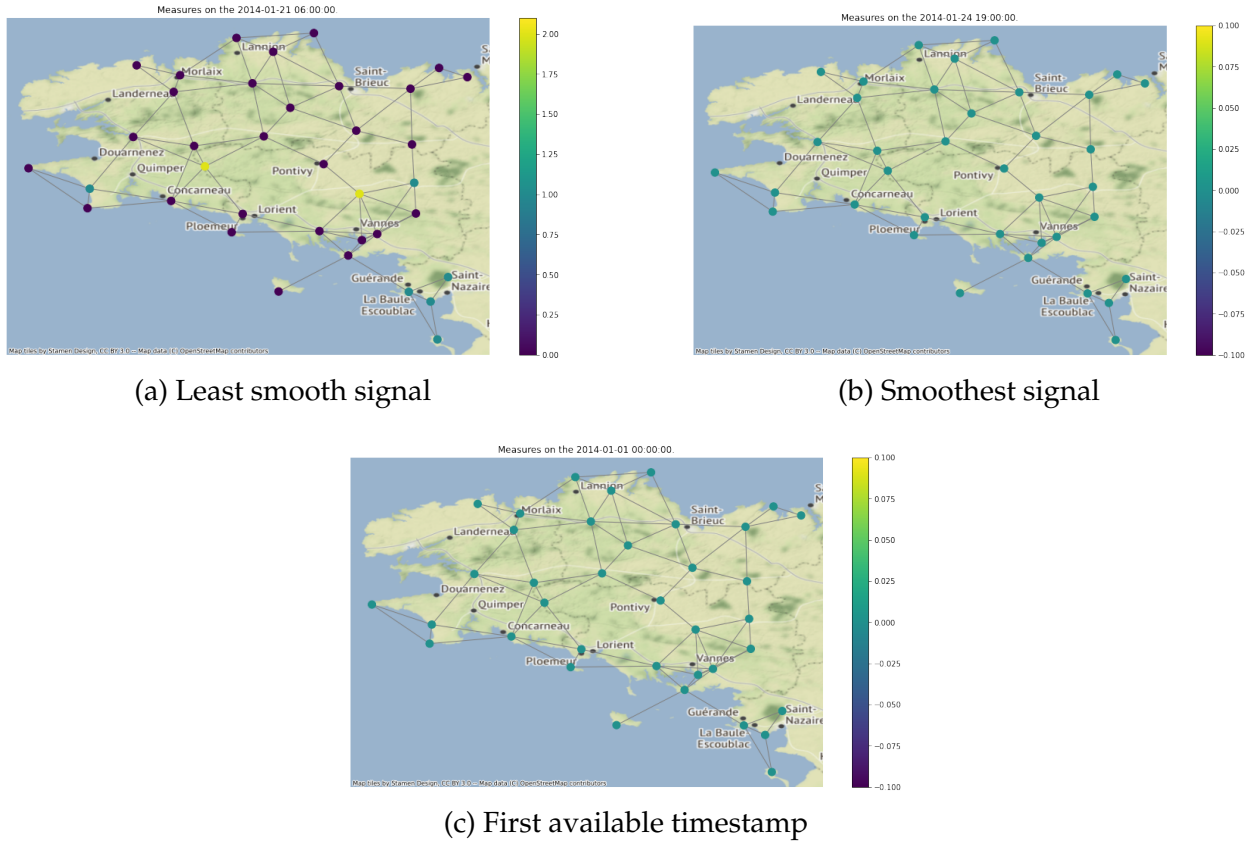


Figure 3: Classification of nodes into low / medium / high frequency

Question 6

Display the average temperature and for each timestamp, adapt the marker colour to the majority class present in the graph (see notebook for more details).

Answer 6

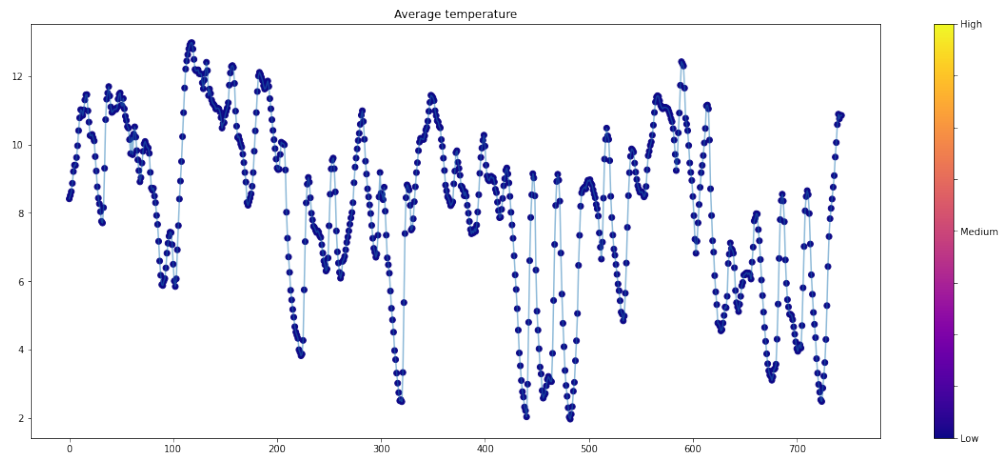


Figure 4: Average temperature. Markers' colours depend on the majority class.

Question 7

The previous graph G only uses spatial information. To take into account the temporal dynamic, we construct a larger graph H as follows: a node is now *a station at a particular time* and is connected to neighbouring stations (with respect to G) and to itself at the previous timestamp and the following timestamp. Notice that the new spatio-temporal graph H is the Cartesian product of the spatial graph G and the temporal graph G' (which is simply a line graph, without loop).

- Express the Laplacian of H using the Laplacian of G and G' (use Kronecker products).
- Express the eigenvalues and eigenvectors of the Laplacian of H using the eigenvalues and eigenvectors of the Laplacian of G and G' .
- Compute the wavelet transform of the temperature signal.
- Classify nodes into low/medium/high frequency and display the same figure as in the previous question.

Answer 7

- The Laplacian of H is the kronecker sum of the laplacian of G and the laplacian of G' , that can be also expressed by the Kronecker products as:

$$L_H = L_G \oplus L_{G'} = L_G \otimes I_n + L_{G'} \otimes I_m$$

where n and m are respectively the number of nodes in graph G and G' and L_A is the laplacian of A .

- The eigenvalues of the Laplacian of H are giving by: $\{\lambda_i + \mu_j, 1 \leq i \leq n, 1 \leq j \leq m\}$ and the corresponding eigenvectors are giving by the kronecker product of the eigenvectors of the laplacian of G and that of G' as $\{x_i \otimes y_j, 1 \leq i \leq n, 1 \leq j \leq m\}$ where λ_i (resp μ_j) are the eigenvalues of G (res of G') and x_i (res y_j) are the corresponding eigenvectors.

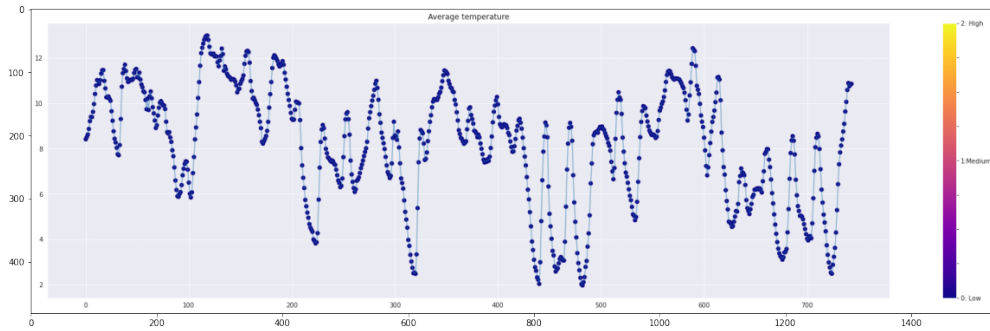


Figure 5: Average temperature. Markers' colours depend on the majority class.