

Detecting Changes in Slope With an L_0 Penalty

Bastien LE CHENADEC bastien.le-chenadec@eleves.enpc.fr
Sofiane EZZEHI sofiane.ezzehi@eleves.enpc.fr

January 3, 2024

What is expected for these mini-projects? *The goal of the exercise is to read (and understand) a research article, implement it (or find an implementation), test it on real data and comment on the results obtained. Depending on the articles, the task will not always be the same: some articles are more theoretical or complex, others are in the direct line of the course, etc... It is therefore important to balance the exercise according to the article. For example, if you have reused an existing implementation, it is obvious that you will have to develop in a more detailed way the analysis of the results, the influence of the parameters etc... Do not hesitate to contact us by email if you wish to be guided.*

The report *The report must be at most FIVE pages and use this template (excluding references). If needed, additional images and tables can be put in Appendix, but must be discussed in the main document. The report must contain a precise description of the work done, a description of the method, and the results of your tests. Please do not include source code! The report must clearly show the elements that you have done yourself and those that you have reused only, as well as the distribution of tasks within the team (see detailed plan below.)*

The source code *In addition to this report, you will have to send us a Python notebook allowing to launch the code and to test it on data. For the data, you can find it on standard sites like Kaggle, or the site <https://timeseriesclassification.com/> which contains a lot of signals!*

The oral presentations *They will last 10 minutes followed by 5 minutes of questions. The plan of the defense is the same as the one of the report: presentation of the work done, description of the method and analysis of the results.*

Deadlines *Two sessions will be available :*

- **Session 1**
 - Deadline for report: December 18th (23:59)
 - Oral presentations: December 20th and 22th (precise times TBA)
- **Session 2**
 - Deadline for report: January 9th (23:59)
 - Oral presentations: January, 11th and 12th (precise times TBA)

1 Introduction and contributions

The Introduction section (indicative length : less than 1 page) should detail the scientific context of the article you chose, as well as the task that you want to solve (especially if you apply it on novel data). **The last paragraph of the introduction must contain the following information:**

- Repartition of work between the two students
- Use of available source code or not, percentage of the source code that has been reused, etc.
- Use of existing experiments or new experiments (e.g. test of the influence of parameter that was not conducted in the original article, application of the method on a novel task/data set etc.)
- Improvement on the original method (e.g. new pre/post processing steps, grid search for optimal parameters etc.)

2 Method

The Method section (indicative length : 1 to 2 pages) should describe the mathematical aspects of the method in a summarized manner. Only the main steps that are useful for understanding should be highlighted. If relevant, some details on implementation can be provided (but only marginally).

A word on notations : if we have a sequence $x = x_1, \dots, x_n$, we denote $x[s : t]$ the subsequence x_s, \dots, x_t .

2.1 Model

Let $y = y_1, \dots, y_n$ be successive data points in time. The goal of this method is to find m change-points τ_1, \dots, τ_m such that the data is divided in $m + 1$ segments. We let $\tau_0 = 0$ and $\tau_{m+1} = n$; the segment j is then $y_{\tau_{j-1}+1}, \dots, y_{\tau_j}$.

The method considers a piecewise linear model to fit the data. We denote ϕ_{τ_j} the value taken by the model at time τ_j . Under these assumptions, the model writes :

$$\forall 0 \leq i \leq m, \forall \tau_i + 1 \leq t \leq \tau_{i+1}, \quad Y_t = \phi_{\tau_i} + \frac{\phi_{\tau_{i+1}} - \phi_{\tau_i}}{\tau_{i+1} - \tau_i}(t - \tau_i) + Z_t \quad (1)$$

where $(Z_t)_{1 \leq t \leq n}$ is assumed to be a gaussian white noise with variance σ^2 .

2.2 Penalized cost approach

The method uses a penalized cost approach to find the changepoints. The cost function is a squared error loss, penalized by a term that depends on the segments length, and an L_0 penalty on the number of changepoints. The cost function naturally writes :

$$\sum_{i=0}^m \left[\frac{1}{\sigma^2} \sum_{t=\tau_i+1}^{\tau_{i+1}} \left(y_t - \phi_{\tau_i} - \frac{\phi_{\tau_{i+1}} - \phi_{\tau_i}}{\tau_{i+1} - \tau_i}(t - \tau_i) \right)^2 + h(\tau_{i+1} - \tau_i) \right] + \beta m \quad (2)$$

By introducing a segment cost, for fitting a segment that takes the value ϕ at time s and ψ at time t , on the data y_{s+1}, \dots, y_t , that writes :

$$C(y[s+1 : t], \phi, \psi) = \frac{1}{\sigma^2} \sum_{j=s+1}^t \left(y_j - \phi - \frac{\psi - \phi}{t - s}(j - s) \right)^2 \quad (3)$$

the penalized cost function can be rewritten as :

$$\min_{\substack{m \\ \tau_1, \dots, \tau_m \\ \phi_0, \dots, \phi_{m+1}}} \beta m + \sum_{i=0}^m \mathcal{C}(y[\tau_i + 1 : \tau_{i+1}], \phi_{\tau_i}, \phi_{\tau_{i+1}}) + h(\tau_{i+1} - \tau_i) \quad (4)$$

2.3 Dynamic programming

We want to solve the problem (4) by dynamic programming. Compared to other changepoints detection methods, the difficulty lies in the continuity constraint between segments. Thus the authors introduce $f^t(\phi)$ as the minimum cost for segmenting y_1, \dots, y_t in $k + 1$ segments conditionally on the model taking the value ϕ at time t :

$$f^t(\phi) = \min_{\substack{\tau_1, \dots, \tau_k \\ \phi_0, \dots, \phi_k}} \sum_{i=0}^{k-1} \mathcal{C}(y[\tau_i + 1 : \tau_{i+1}], \phi_{\tau_i}, \phi_{\tau_{i+1}}) + h(\tau_{i+1} - \tau_i) \\ + \mathcal{C}(y[\tau_k + 1 : t], \phi_{\tau_k}, \phi) + h(t - \tau_k) + \beta k \quad (5)$$

which they are able to express recursively :

$$f^t(\phi) = \min_{\phi', s} f^s(\phi') + \mathcal{C}(y[s + 1 : t], \phi', \phi) + h(t - s) + \beta \quad (6)$$

They also introduce $f_\tau^t(\phi)$ which is the minimum cost for segmenting y_1, \dots, y_t in $k + 1$ segments conditionally on the model taking the value ϕ at time t and the changepoints being $\tau = (\tau_0, \tau_1, \dots, \tau_k, \tau_{k+1})$. This quantity is a quadratic polynomial in ϕ and the authors are able to express its coefficients recursively. All of this allows them to compute $f^n(\phi)$ in $\mathcal{O}(n!/(n-m)!)$.

2.4 Pruning

Given the high complexity of the dynamic programming algorithm (exploring the space of all possible changepoints), the authors introduce multiple pruning techniques to reduce the complexity.

Functional pruning The authors introduce \mathcal{T}_t the set of all possible changepoint vectors at time t , and $\mathcal{T}_t^* = \{\tau \in \mathcal{T}_t \mid \exists \phi \in \mathbb{R}, f^t(\phi) = f_\tau^t(\phi)\}$ the set of changepoint vectors at time t that are optimal for some ϕ .

Theorem 1. Let $\tau \in \mathcal{T}_s$ such that $\tau \notin \mathcal{T}_s^*$. Then $\tau \notin \mathcal{T}_t^*$ for all $t > s$.

This theorem allows us to only explore $\hat{\mathcal{T}}_t = \{(\tau, s) \mid 0 \leq s \leq t - 1, \tau \in \mathcal{T}_s^*\}$ when computing $f^t(\phi)$.

Inequality based pruning Let $K = 2\beta + h(1) + h(n)$. Suppose that h is non-negative and non-decreasing.

Theorem 2. Let $\tau \in \mathcal{T}_s$ such that :

$$\min_{\phi} f_\tau^s(\phi) > K + \min_{\phi'} f^s(\phi')$$

Then $\tau \notin \mathcal{T}_t^*$ for all $t > s$.

This theorem allows to further prune the set $\hat{\mathcal{T}}_t$.

2.5 Algorithm

The two difficulties of the implementation are storing the coefficients of the polynomials $f_\tau^t(\phi)$, and implementing the pruning techniques. The authors provided relatively detailed pseudo-code for the algorithm, but we still found it difficult to implement.

The goal of this first algorithm is to only keep segmentations that are optimal for some values of ϕ . Segmentation costs are polynomials, so this algorithm starts from the optimum segmentation for $\phi = -\infty$ and iteratively finds the next optimum segmentation by looking at the points where the polynoms intersect.

Algorithm 1 Functional pruning algorithm

```

1: Input: Candidate segmentations  $\hat{\mathcal{T}}_t$ , costs  $f_\tau^t(\phi)$  for  $\tau \in \hat{\mathcal{T}}_t$ .
2: Initialize:  $\phi_{curr} = -\infty$   $\tau_{curr} = \arg \min_{\tau \in \hat{\mathcal{T}}_t} [f_\tau^t(\phi_{curr})]$   $\mathcal{T}_{temp} = \hat{\mathcal{T}}_t \setminus \{\tau_{curr}\}$   $\mathcal{T}_t^* = \{\tau_{curr}\}$ 
3: while  $\mathcal{T}_{temp} \neq \emptyset$  do
4:   for  $\tau \in \mathcal{T}_{temp}$  do
5:      $x_\tau = \min \{\phi : f_\tau^t(\phi) - f_{\tau_{curr}}^t(\phi) = 0 \ \& \ \phi > \phi_{curr}\}$ 
6:     if  $x_\tau = \emptyset$  then
7:        $\mathcal{T}_{temp} = \mathcal{T}_{temp} \setminus \{\tau\}$ 
8:    $\tau_{curr} = \arg \min_{\tau \in \mathcal{T}_{temp}} (x_\tau)$ 
9:    $\phi_{curr} = x_{\tau_{curr}}$ 
10:   $\mathcal{T}_t^* = \mathcal{T}_t^* \cup \{\tau_{curr}\}$ 
11: Output:  $\mathcal{T}_t^*$ 

```

Then we can define the full algorithm that uses both pruning techniques.

Algorithm 2 CPOP

```

1: Input: Data  $y = y_1, \dots, y_n$ , penalty  $\beta$  and segment cost  $h$ .
2: Initialize:  $\hat{\mathcal{T}}_1 = \{\{0\}\}$ 
3: for  $t = 1, \dots, n$  do ▷ Cost computation
4:   for  $\tau \in \hat{\mathcal{T}}_t$  do
5:     if  $\tau = \{0\}$  then ▷ No changepoint
6:        $f_\tau^t(\phi) = \min_{\phi'} \mathcal{C}(y[1:t], \phi', \phi) + h(t)$ 
7:     else ▷ Compute recursively
8:        $f_\tau^t(\phi) = \min_{\phi'} (f_{\tau_1, \dots, \tau_{k-1}}^{\tau_k}(\phi') + \mathcal{C}(y[\tau_k + 1:t], \phi', \phi) + h(t - \tau_k) + \beta)$ 
9:      $\mathcal{T}_t^* = \text{functional\_pruning}(\hat{\mathcal{T}}_t, f^t)$ 
10:     $\hat{\mathcal{T}}_{t+1} = \{\tau \in \hat{\mathcal{T}}_t \mid \min_{\phi} f_\tau^t(\phi) \leq \min_{\phi', \tau'} f_{\tau'}^t(\phi') + K\} \cup \{(\tau, t) \mid \tau \in \mathcal{T}_t^*\}$ 
11: Output:  $\arg \min_{\tau \in \hat{\mathcal{T}}_n} [\min_{\phi} f_\tau^n(\phi)]$ 

```

We note that we only need to store the coefficients for $f_{\tau_1, \dots, \tau_{k-1}}^{\tau_k}$.

3 Data

The Data section (indicative length : 1 page) should provide a deep analysis of the data used for experiment. In particular, we are interested here in your capacity to provide relevant and thoughtful feedbacks on the data and to demonstrate that you master some "data diagnosis" tools that have been dealt with in the lectures/tutorials.

4 Results

The Result section (indicative length : 1 to 2 pages) should display numerical simulations on real data. If you re-used some existing implementations, it is expected that this section develops new experiments that were not present in the original article. Results should be discussed not only based on quantitative scores but also on qualitative aspects. In particular (especially if your article focuses on black box methods), please provide some feedbacks whether the method was adapted to the data or not and whether the hypothesis behind the approach you used were validated or not.