8 TDA for financial time series: persistent homology and landscapes

Description. The goal of this project is to analyze the evolution of daily returns of four major US stock markets indices (DowJones, Nasdaq, Russell2000, SP500) over the period 1989 – 2016 using persistent homology following the approach proposed in [14]. A classical approach in TDA to extract topological features from multivariate time-series taking its values in $\mathbb{R}^d$ (here, since we are considering the evolution of four indices $d = 4$) consists in using a sliding window of fixed length w to generate a sequence of w points in $\mathbb{R}^d$. Using the Vietoris-Rips filtration, the persistence diagram of each of these point cloud is then computed and used as a topological feature for further analysis or processing of the initial data.

This project aims at reproducing the experiments of [14] and explore and discuss a few variants.

Tasks.

1. Download the paper [14] and the data from the following address: `http://geometrica.saclay.inria.fr/team/Fred.Chazal/Centrale2017.html`. Have a quick look at the whole paper [14] to get used to the considered problem and proposed approach, and a more careful reading of Sections 3.1 and 4.

2. Write a function to compute persistence landscapes. This function should take as input a persistence diagram dgm (in the Gudhi format), a dimension k, the endpoints x_{min}, x_{max} of an interval, the number nb_{nodes} of nodes of a regular grid on the interval $[x_{min}, x_{max}]$ and a number of landscapes nb_{ld}, and output a $nb_{ld} \times nb_{nodes}$ array storing the values of the first nb_{ld} landscapes of dgm on the node of the grid.

3. Use the landscape function to run the experiments done in Section 4 of [14] but taking windows of length $w = 40$ and $w = 80$ and $w = 120$. Compare your results to the ones provided in the paper (are they very similar).

4. Propose and experiment another method, than just computing the norm of landscapes (e.g. consecutive bottleneck distances between persistence diagrams, norm of the difference between consecutive landscapes,...). Briefly discuss and compare your results to the ones of Section 4 in [14].

Software. For persistent homology computations, the use of the GUDHI library (`http://gudhi.gforge.inria.fr/`) is strongly recommended (C++ or Python version). Alternately, you can use the R package TDA : `https://cran.r-project.org/web/packages/TDA/index.html`

Contact. Frédéric Chazal: frederic.chazal@inria.fr