

An introduction to Topological Data Analysis

Wednesday, June 7, 2023 9:30 AM (1 hour)

Topological Data Analysis (TDA) is a growing field of research at the intersection of data science and computational geometry and topology. It has encountered key successes in several different applications (ranging from cancer subtype identification in bioinformatics to shape recognition in computer vision, just to name a few), and become the landmark product of several companies in the recent years. Indeed, many data sets nowadays come in the form of point clouds embedded in very large dimensional spaces, yet concentrated around low-dimensional geometric structures that need to be uncovered. Unraveling these structures is precisely the goal of TDA, which can build descriptors that can reliably capture geometric and topological information (connectivity, loops, holes, curvature, etc.) from the data sets without the need for an explicit mapping to lower-dimensional space. This is extremely useful since the hidden, non-trivial topology of many data sets can make it very challenging to perform well for classical techniques in data science and machine learning, such as dimensionality reduction.

In this talk, I will provide a global overview of TDA, by introducing its main descriptors and by presenting the theoretical guarantees that they enjoy. I will also show how they can be efficiently computed in practice with the dedicated, open-source library GUDHI, and describe some applications where TDA proved useful.

Presenter: CARRIÈRE, Mathieu