CLUSTERING OF DISCRETE MEASURES VIA MEAN MEASURE QUANTIZATION WITH APPLICATIONS TO TOPOLOGICAL DATA ANALYSIS

Frédéric Chazal

INRIA Saclay

Abstract. In Topological Data Analysis (TDA), robust topological information commonly comes in the form of a set of so-called persistence diagrams that can be seen as discrete measures and are uneasy to use in generic machine learning frameworks. In this talk we will introduce a fast, learnt, unsupervised vectorization method for measures in Euclidean spaces and use it for clustering of distributions of discrete measures. The algorithm is simple and efficiently discriminates important space regions where meaningful differences to the ''mean'' measure arise. Applied to TDA problems, we will show that it is proven to be able to separate clusters of persistence diagrams. We will illustrate the strength and robustness of our approach on a few synthetic and real data sets.