Approximating Continuous Functions on Persistence Diagrams for Machine Learning Tasks

Elizabeth Munch, Michigan State University

Many machine learning tasks can be boiled down to the following idea: Approximate a continuous function defined on a topological space (the ``ground truth'') given the function values (or approximate function values) on some subset of the points. This formulation has been well studied Euclidean data; however, more work is necessary to extend these ideas to arbitrary topological spaces.

In this talk, we focus on the task of classification and regression on the space of persistence diagrams endowed with the bottleneck distance, $(\mathcal{D},d_B)$. These objects arise in the field of Topological Data Analysis as a signature which gives insight into the underlying structure of a data set. The issue is that the structure of $(\DD,d_B)$ is not directly amenable to the application of existing machine learning theories. In order to properly create this theory, we will give a full characterization of compact sets in $\mathcal{D}$; provide simple, exemplar functions for vectorization of persistence diagrams; and show that, in practice, this method is quite successful in classification and regression tasks on several data sets of interest.