Persistence Images: Machine Learning on the Shape of Data

Tegan Emerson, Naval Research Laboratory

Over recent years tools in computational topology have become more efficient and capable of dealing with larger data sets. As the popularity and scalability of these tools has increased so has the need for techniques to leverage the computed topological structures within machine learning. We will describe two of the first tools that were used towards this end: the bottleneck and Wasserstein distances and persistence landscapes. Next, we will present Persistence Images (PIs): a stable vector-based representation of persistent homology. Comparison of PIs to both distance based classifiers and persistence landscapes on a toy data set will be provided. Additionally, we will present results on synthetic data sets for parameter identification in nonlinear systems. Finally, we will conclude with results on a real world application arising from the desire to differentiate between the brain structures of Schizophrenics, their asymptomatic siblings, and healthy controls. The presented results leverage feature selection via sparse support vector machines as well as the connection between pixels of the persistence image and critical regions of the birth-persistence plane.