FLORIDA ATLANTIC UNIVERSITY

Department of Mathematics and Statistics
Florida Atlantic University

MS Thesis Defense

Nicole Abreu

Topological Machine Learning with Unreduced Persistence Diagrams

Wednesday, November 13, 3:00pm in SE 215

Advisor: Dr. Francis Motta
Co-Advisor: Dr. Parker Edwards

A common topological data analysis approach used in the experimental sciences involves creating machine learning pipelines that incorporate discriminating topological features derived from persistent homology (PH) of data samples, encoded in persistence diagrams (PDs) and associated topological feature vectors. Often the most computationally demanding step is computing PH through an algorithmic process known as boundary matrix reduction. In this work, we introduce several methods to generate topological feature vectors from unreduced boundary matrices. We compared the performance of classifiers trained on vectorizations of unreduced PDs to vectorizations of fully reduced PDs across several benchmark ML datasets. We discovered that models trained on PDs built from unreduced diagrams can perform on par and even outperform those trained on full-reduced diagrams. This observation suggests that machine learning pipelines which incorporate topology-based features may benefit in terms of computational cost and performance by utilizing information contained in unreduced boundary matrices. A common topological data analysis approach used in the experimental sciences involves creating machine learning pipelines that incorporate discriminating topological features derived from persistent homology (PH) of data samples, encoded in persistence diagrams (PDs) and associated topological feature vectors. Often the most computationally demanding step is computing PH through an algorithmic process known as boundary matrix reduction. In this work, we introduce several methods to generate topological feature vectors from unreduced boundary matrices. We compared the performance of classifiers trained on vectorizations of unreduced PDs to vectorizations of fully reduced PDs across several benchmark ML datasets. We discovered that models trained on PDs built from unreduced diagrams can perform on par and even outperform those trained on full-reduced diagrams. This observation suggests that machine learning pipelines which incorporate topology-based features may benefit in terms of computational cost and performance by utilizing information contained in unreduced boundary matrices.

Please contact Dr. Hongwei Long (hlong@fau.edu) for an electronic copy of the thesis.
ALL ARE CORDIALLY INVITED