Arya Amarnath

Max Colbert

Bobby Estrada

Hilena Shekour

12/05/2023

Sliding Window Embedding and Persistence

Recently Sliding Window Embedding and Persistence has become a hot topic in the Data Science and Machine Learning field. In this project we were able to explore the definition, use and implement it using real world example data. Sliding window embedding is a technique used to turn time series data into point cloud data, which can then be analyzed using various mathematical and statistical methods. A time series is a sequence of points (xt) = (xt, xt+1, xt+2, . . .) where each point, xt, is within a, finite-dimensional, vector space, V. A point cloud is a series of points in space.We can relate sliding window embedding to time-series analysis, representing a signal of interest as a real-valued function f that maps time to the real numbers. By taking windows of the time series data, we can create representations of the signal in the form of vectors.

Persistence is also a very useful tool that allows us to extract important information from the data and quantify topological shapes.We can extract the information that filtrations made from our data to merge the data together over layers, finding new topological features within the shape. One filtration method is simplicial complexes, composed of simplices, that create a 'homology vector space'. By constructing simplicial complexes and using matrix operators, eigenvalues, and eigenvectors, we can analyze the properties of the data. In this project, we were able to analyze the persistence for gravitational waves. Gravitational waves are ripples in the fabric of spacetime caused by the acceleration of massive objects, such as black holes or neutron stars that are are incredibly faint and difficult to detect and carry important information about the astrophysical events that produced them.The first detection was made in 2015 by the Laser Interferometer Gravitational-Wave Observatory. Using topological methods to analyze the persistence of gravitational wave detections, we can gain new insights into the behavior and properties of these fascinating astrophysical phenomena.Sliding Window Embedding and Persistence have a wide range of applications in various fields, including finance, biology, and natural language processing. The mathematical techniques involved in the analysis of sliding window embedding and persistence are complex but provide insights into complex systems that would otherwise be difficult to analyze.

For the code example portion of our presentation, we utilized several python libraries to demonstrate sliding window embedding. The SingleTakensEmbedding function from the Giotto TDA performed the main operations for creating the sliding window embedding. We used numpy to create some synthetic data that was periodic, and aperiodic, to demonstrate the embedding technique. The periodic function created a loop, while the aperiodic function created a much more complex shape. This data was then used to create persistence diagrams in order to extract more useful information about the data. The same process was repeated for the gravitational wave portion of the code. As you can see in the workbook, we use the sliding window embedding and persistence diagrams to train a random forest classifier. We were able to achieve around a 70-75% accuracy with our model, which demonstrates that the model was able to identify the gravitational waves from the noise with moderate success. This also demonstrates that the sliding window embeddings, along with persistence diagrams were able to extract data that was actually useful in classification.

Sliding window embeddings are used in other areas for classification as well. Certain natural language models use sliding window embeddings to help with grammatically poor inputs. This technique can also be useful for named entity recognition in natural language text, such as names or places of importance. Anywhere that data is on a time series, or can be treated as such, sliding window embeddings are a useful tool for any data scientist to have access to.

References

1) https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8948044

2) https://ieeexplore.ieee.org/document/4670007

3) https://link.springer.com/article/10.1007/s10208-014-9206-z

4) https://arxiv.org/pdf/1910.08245.pdf

5) https://www.youtube.com/watch?v=5DN68XkTnIY