Updating the R Package 'TDA'

Ben Holmgren

R Package Background

R interface for the efficient algorithms of the C++ libraries 'GUDHI', 'Dionysus' and 'PHAT'.

Implements the methods in Fasy et al. (2014) and Chazal et al. (2014) for analyzing statistical significance of persistent homology features.

Functions

Self generating documentation for the Distance Function, the Distance to Measure Function, the k-nearest neighbor density estimator, the kernel density estimator, and rips diag via the roxygen2 package for R documentation.

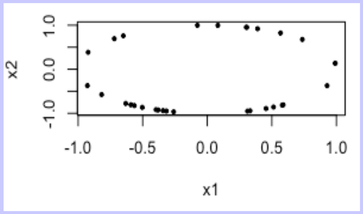
Online Tutorials

Generated online documentation providing tutorials on kernel density estimation, persistence diagrams, and rips filtration using Jekyll

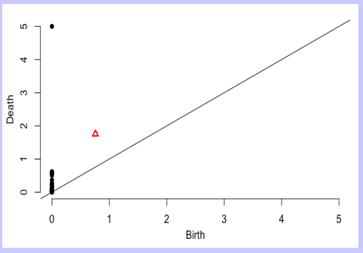
Introduction to Persistent Homology

In TDA, a core approach to understanding the topological structure of data is through persistent homology, which quantifies the persistence of features in a family of nested topological spaces.

When applying any of the filtrations used in TDA, topological features will emerge and disappear, so called births and deaths.
As a filtration is taking place, we can plot at what value of a parameter t that features are born and die.



For example, for a point cloud along the unit circle, we can calculate the persistence diagram resulting from a Rips filtration.



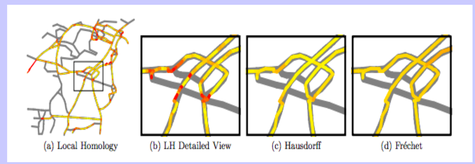
Note that all features are born at time 0, since the Rips filtration begins with all vertices simultaneously. Features die respective to when they are engulfed by the growing simplicial complex, and one feature dies at time infinity. This is the simplex which is the final remaining feature in the simplicial complex.

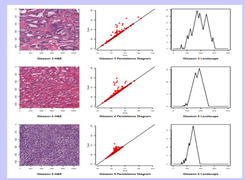
From this, we can gain the persistence of each feature in the data set. The further it is above the curve representative of time on bothaxes, the greater that feature's persistence. High persistence features are generally thought to be statistically significant, while lower persistence features more representatie of statistical noise.

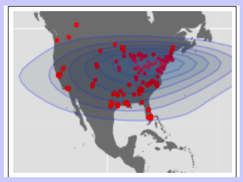
Motivation

Topological Data Analysis (TDA) is a set of methods to find topological structure in data, which is generally represented by point clouds. Due to the work of Dr. Fasy and Dr. Millman at Montana State University along with their collaborators, the R package 'TDA' was created to actually compute topological features from a data set.

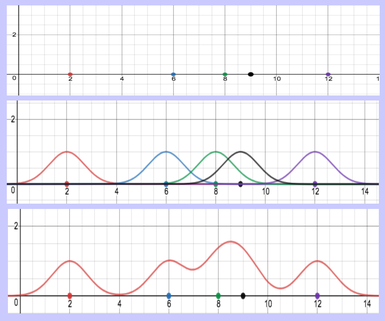
With recent developments in computational topology which allow for the computation of important functions in TDA, a gap has emerged between theoretical understanding in TDA and actual widespread implementation. My goal has been to mitigate the divide between the theoretical side of TDA and its actual widespread use, and I've been able to do so through improving the documentation within the R package, as well as by creating online tutorials on important topics in TDA..



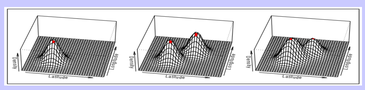




Kernel Density Estimation



To represent point cloud data effectively,a primary method in TDA is by using kernels. Commonly, if Gaussian curves are inserted above points and summed together, a helpful representation of data is gained.



Applications

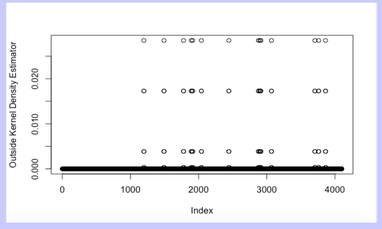
Nonparametric Kernel Estimators for Image Classification
Schneider et al.

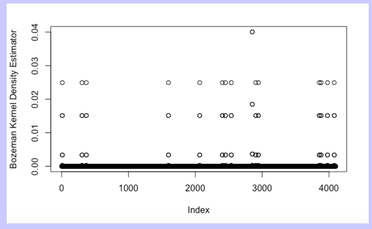
Kernel Density Estimation for Text-Based Geolocation
Hulden et al.

A Review of Kernel Density Estimation with Applications to Econometrics
Diaz, Zambom

Closer to home

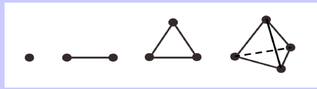
For the sake of a tutorial true to the community surrounding Montana State University, we can conduct a kernel density estimate on top runners in last year's Bridger Ridge run, discovering kde values which favor runners from Bozeman to runners outside of Bozeman in the top 30 places

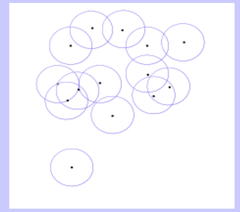


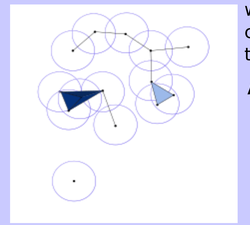


Vietoris Rips Filtration

To begin, recall a simplex, a triangle of n dimensions







Considering a point cloud, we can then extend circles with equivalent radii from each point, and add an edge between every pair of points whose circles intersect. In the end, this constructs a simplicial complex.

By changing the radii of these circles, we then gain a series of simplicial complexes, which can be understand topologically with persistent homology

Rips filtrations remain true to a point cloud regardless of direction, though short-comings exist in its viability over large data

Applications

Shape reconstruction

''Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes'
Attali et al.

"SimBa: An Efficient Tool for Approximating Rips-filtration Persistence via Simplicial Batch-collapse"
Dey et al.

Future Work

For the conclusion of the academic year and through the summer, I plan to continue activity with the R TDA package, as well as continuing to generate more online tutorials through jekyll on the site I've created for this purpose, located at:

https://comptag.github.io/rpackage_tutorials/

I have also been attending weekly seminars hosted by the CompTAG group at MSU throughout the academic year, and have found interest in numerous topics, most recently in a variation to the art gallery problem, attempting to extend an NP-Complete proof into 3 dimensions. I also plan to continue to think about this and other geometric optimization problems.

Works Cited

Fasy, B T (2017, August). Research Statement. Retrieved from https : //www.cs.montana.edu/brittany/research/fasy − brittany − rsrch − stmt.pdf

Fasy, B T, Kim, J, Lecci, F, Maria, C, Millman, D L, Rouvreau, V (2018). Introduction to the R package TDA. The Comprehensive R Archive Network, 1-24.

Hulden, Mans, et al. “Kernel Density Estimation for Text-Based Geolocation.” AAAI Publications, Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI Conference on Artificial Intelligence, 9 Feb. 2015, www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/10034.

Mahmuda Ahmed , Brittany Terese Fasy , Carola Wenk, Local persistent homology based distance between maps, Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, November 04-07, 2014, Dallas, Texas

“R Logo.” R, www.r-project.org/logo/.

Reiterer, Andreas. “Create Your Own Blog with Jekyll.” Codeburst, Codeburst, 24 Aug. 2017, codeburst.io/create-your-own-blog-with-jekyll-c0abc3ef7ddd.

Shah, Het. “Flock of Birds Makes Amazing Patterns in the Sky! | Evening Time | Kankaria Lake | Ahmedabad |.” YouTube, YouTube, 26 Feb. 2017, www.youtube.com/watch?v=SRBP7uKuVaQ.

Zanin, Adriano, and Ronaldo Dias. “A Review of Kernel Density Estimation with Applications to Econometrics.” ArXiv.org, Cornell University, 12 Dec. 2012, arxiv.org/abs/1212.2812.

Research reported in this poster was supported by the Undergraduate Scholars Program at Montana State University. Throughout this project, I have been mentored by Dave Millman and Brittany Fasy