MATHEMATICS FOR BIG DATA
TOPOLOGICAL DATA ANALYSIS

AUTHOR: LAWRENCE ASAMOAH ADU-GYAMFI
No.: 1484610
DATE: 30/05/2019

Table of Contents

1. INTRODUCTION

In this exercise we are provided with three (3) separate datasets of points with each possibly being a circle, a sphere or a torus.
The objective of this report is to present the details of steps leading to the conclusion of which of the datasets correspond to which shape or figure.

1.1.Overview of datasets

Each of the datasets consists of 450 observations(rows) with 3 corresponding variables (columns).
Figure 1-1 and Figure 1-2 show the preliminary plots of the datasets as they are; both in 2-dimensional and 3-dimensional views
Observing the different views and plots, we can already make assumptions as which dataset belongs to which shape. However, we will confirm this fully looking in details at the calculation of the persistence homology and the reviewing the corresponding diagrams.
To identify the sphere, we expect to see 1 prominent void persisting in the homology, while for the circle we expect to 1 prominent loop in the persistent homology. For the torus, we expect to see a lot variations of homology groups present in the persistent homology as the radius of the filter function is varied.

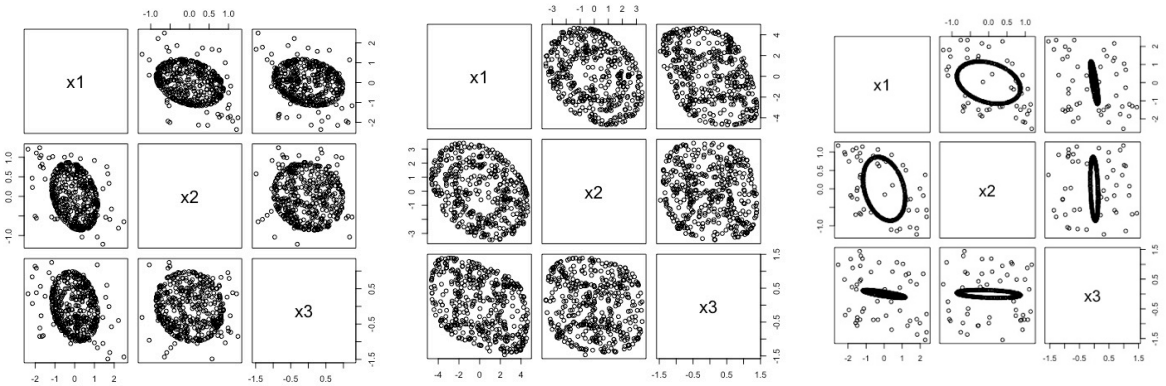

Figure 1-1: 2-D plots of datasets (from left: points1, points2, points3)

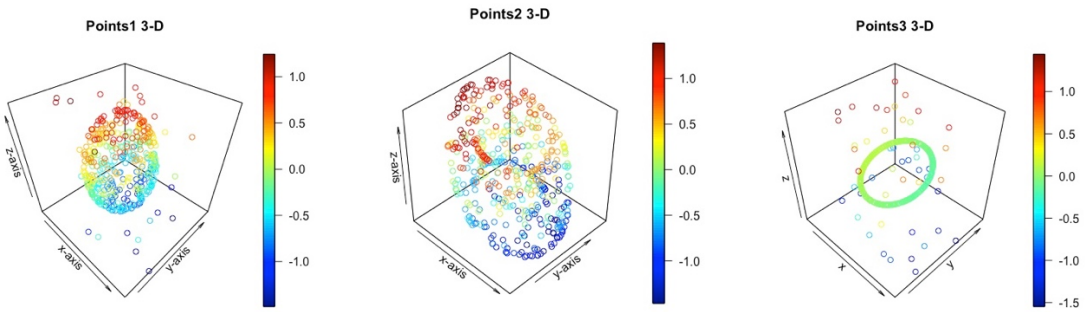

Figure 1-2: 3-D plots of datasets (from left: points1, points2, points3)

2. Methodology, Results, Conclusion

2.1.Methodology

The calculation has performed using the TDA package in the R programming language.

The kernel density estimator and vietoris rips function have been used for the filtrations and compared.
A 3-D grid is defined using the maximum values in the datasets and this is provided as an input parameter for the function in R.
The persistent homology is then calculated the results presented in a persistence diagram and barcode as well.

2.2.Results

From the persistence diagram of the first dataset (points1.csv), we observe a void (blue square) whose vertical distance from the reference diagonal line shows it is significant. This is further confirmed from the barcode diagram as well.
Using the vietoris rips filtering function also detects this void as well and the persistence of the void is much prominent with this approach.

The persistence homology of the second dataset(points2.csv) shows several homology groups being detected, however we notice that most of these do not persist for long. We do see the persistence of several points and loops, but the voids die off immediately after they show up. This phenomenon was detected by both functions.

For the third dataset(points3.csv), we notice the persistence of a loop in the homology. This shape was captured by both filters and is evident on the barcode diagrams as well.

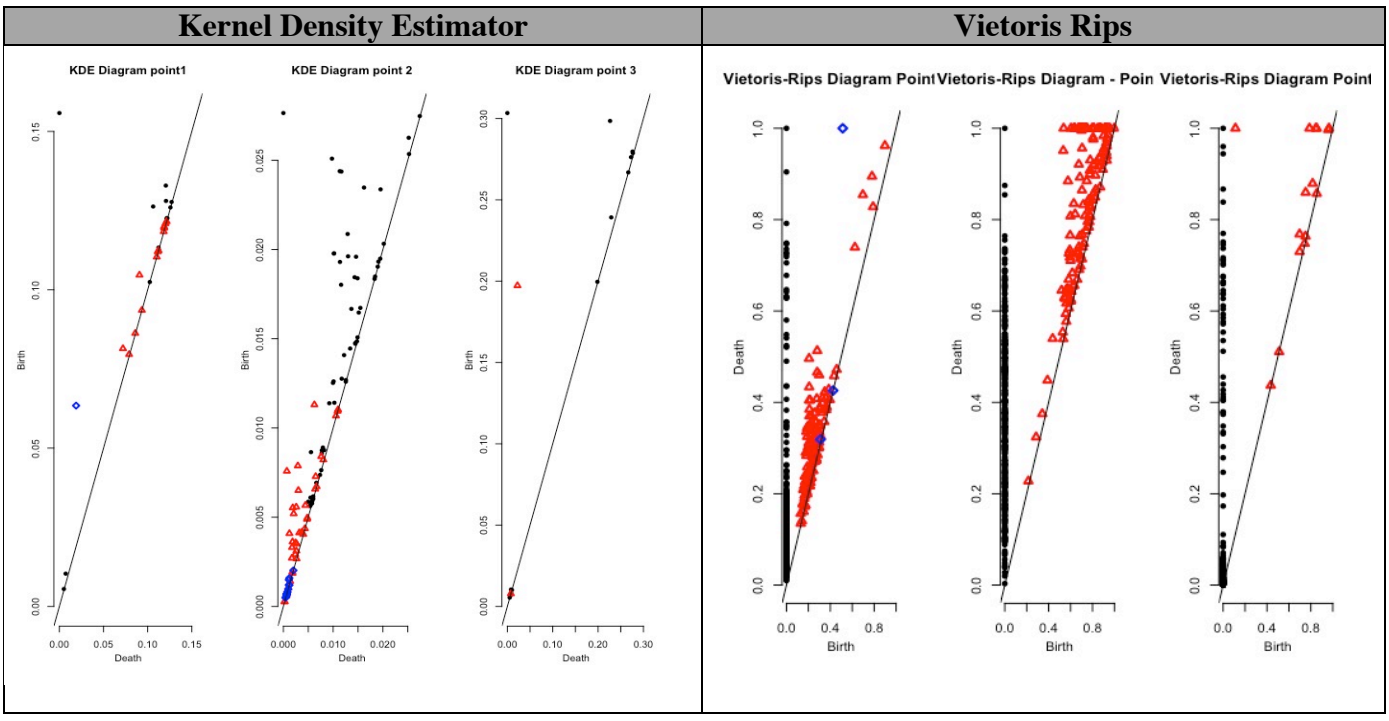

Figure 2-1: Persistent Diagram of Datasets (from left: points1, points2, points3)

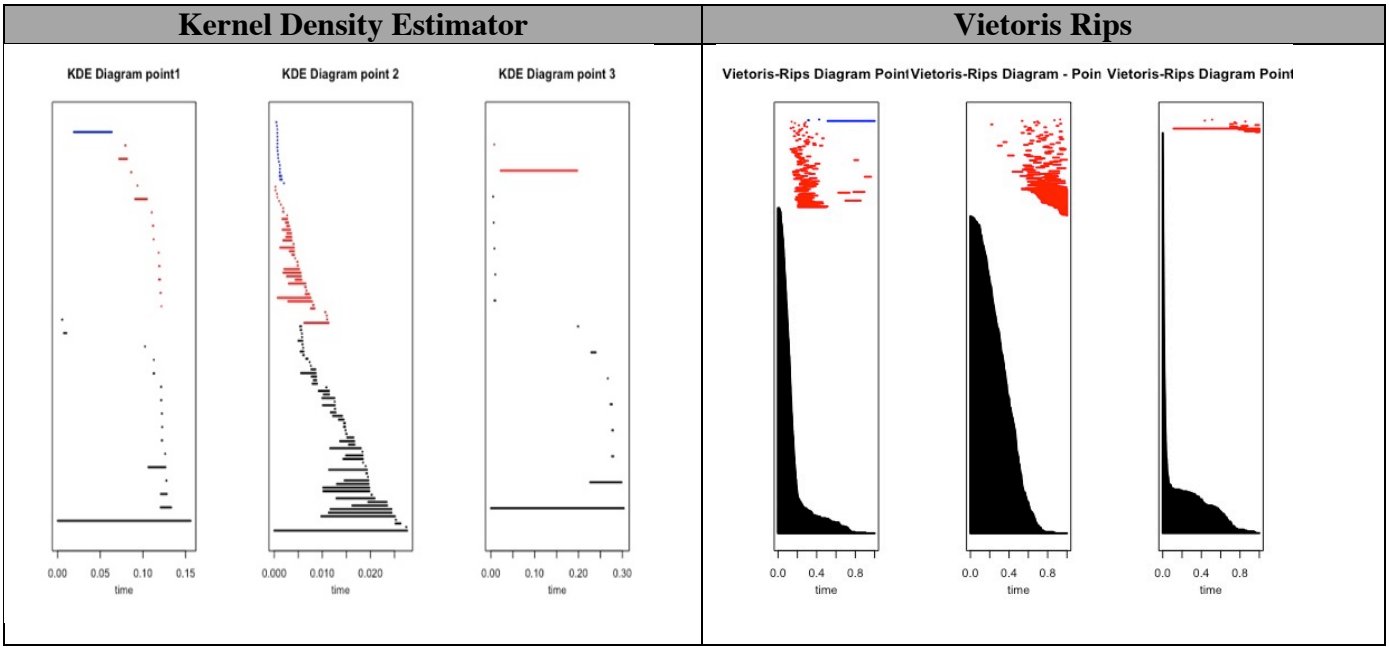

Figure 2-2: Barcode Diagram of Datasets (from left: points1, points2, points3)

2.3.Conclusion

From these observations the conclusion is presented below:

- Points1.csv – sphere
- Points2.csv – torus
- Points3.csv – circle

And these are confirmed by looking at the plots of the raw data in *Figure 1-1* and *Figure 1-2*.

2.1.References

1. **Topological Data Analysis Lectures** – *Albert Ruiz Cirera*
2. **Introduction to the R package TDA** - *Brittany T. Fasy, Jisu Kim, Fabrizio Lecci, Cl´ement Maria, David L. Millman, and Vincent Rouvreau In collaboration with the CMU TopStat Group*