Report about
TOPOLOGICAL DATA ANALYSIS
Three persistence diagrams and comments

GENÍS LÁINEZ MORENO,

Univertitat Autònoma de Barcelona. 24/05/2020

Abstract

This report is intended to make a description of 3 different given datasets and perform persistence diagrams on all of them in order to determine the topological shape of each one.

1 The dataset

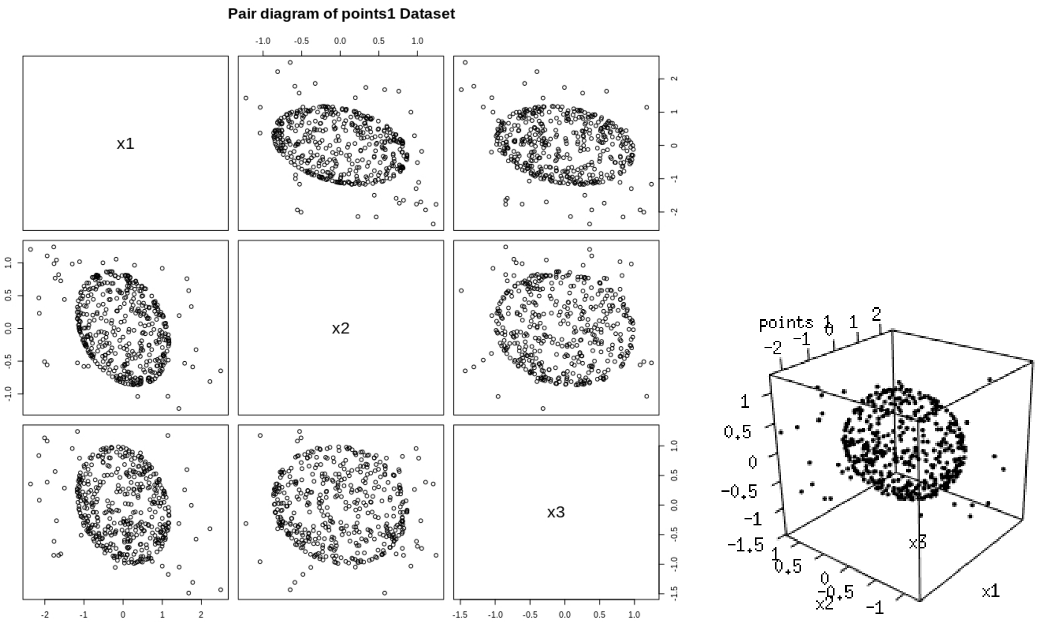
The datasets we are going to study are just three bunches of points on the space. This points are represented on cartessian coordinates on a csv files by columns, *X1, X2* and *X3*.

First, in order to have an idea of each of the datasets I have made it a simple R code that could bring some clues on its shape and could be useful to know in advanced to have an intuition about what their persistence diagrams should be like. (See the Appendix)

First, some summarys on the datasets have been done

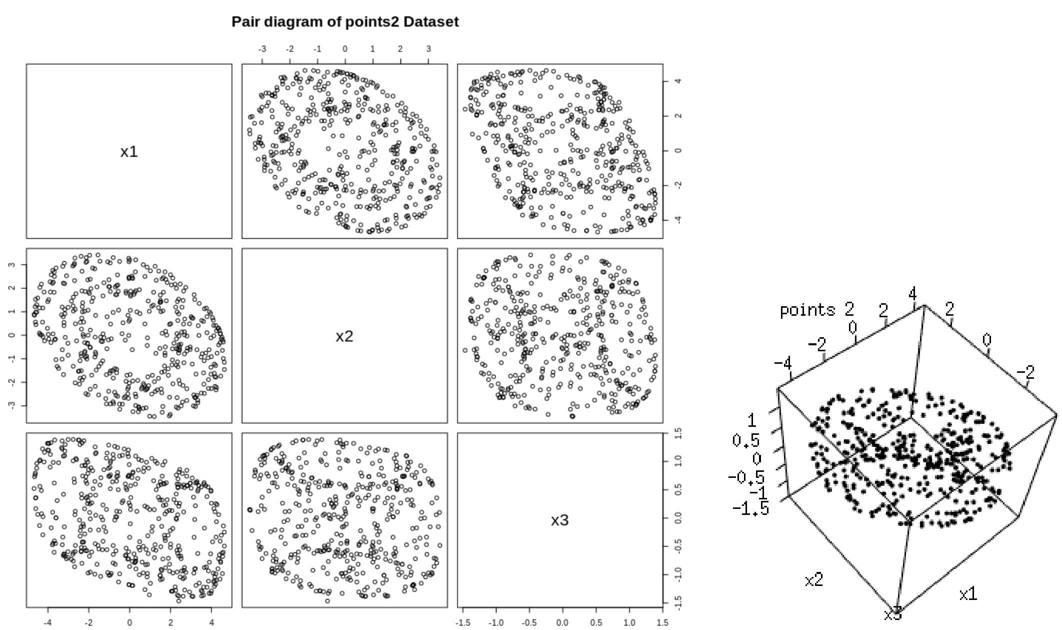
```
> summary(data1)
       x1                 x2                  x3
 Min.   :-2.36374   Min.   :-1.222399   Min.   :-1.48675
 1st Qu.:-0.65768   1st Qu.:-0.451946   1st Qu.:-0.50571
 Median : 0.05588   Median :-0.031955   Median :-0.09062
 Mean   :-0.01243   Mean   :-0.003166   Mean   :-0.05599
 3rd Qu.: 0.59328   3rd Qu.: 0.442404   3rd Qu.: 0.39622
 Max.   : 2.49243   Max.   : 1.242441   Max.   : 1.24336
> summary(data2)
       x1                x2                  x3
 Min.   :-4.6773   Min.   :-3.45492   Min.   :-1.466455
 1st Qu.:-2.0608   1st Qu.:-1.53789   1st Qu.:-0.673516
 Median : 0.1211   Median :-0.01564   Median : 0.037610
 Mean   : 0.1652   Mean   :-0.01442   Mean   :-0.007514
 3rd Qu.: 2.3534   3rd Qu.: 1.46430   3rd Qu.: 0.645450
 Max.   : 4.6326   Max.   : 3.41390   Max.   : 1.388798
> summary(data3)
       x1                 x2                 x3
 Min.   :-2.55408   Min.   :-1.23625   Min.   :-1.546380
 1st Qu.:-0.86303   1st Qu.:-0.62039   1st Qu.:-0.096799
 Median :-0.07918   Median :-0.07707   Median : 0.011875
 Mean   :-0.04447   Mean   :-0.01162   Mean   :-0.001196
 3rd Qu.: 0.78682   3rd Qu.: 0.64134   3rd Qu.: 0.106685
 Max.   : 2.35220   Max.   : 1.17474   Max.   : 1.441551
```

This result give us an idea on where the datasets are located on the space. By looking some information we can have also a little intuition about the noise but we require another analysis to be able to detect it propperly and also have an intuition about the topological shape. For this we have ploted the pair diagrams and also the 3D plots as it can be seen on the following figures.



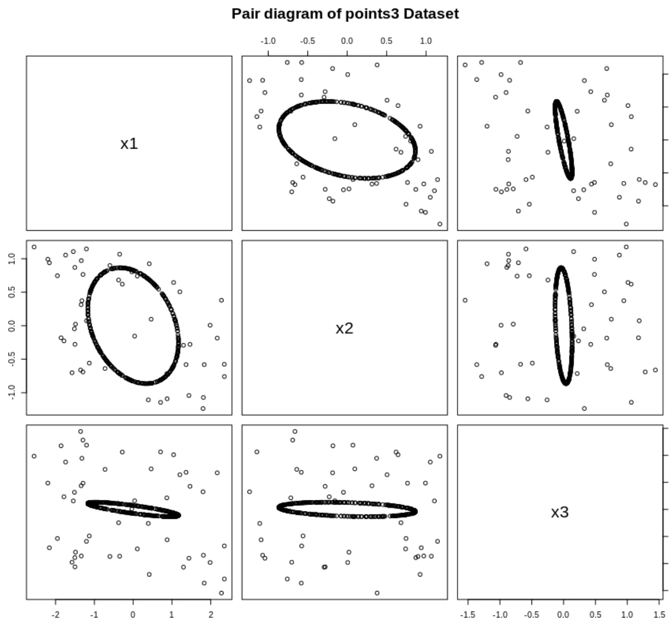
(a) (b)

Figure 1: Pair diagram and 3D plot of the points1 dataset

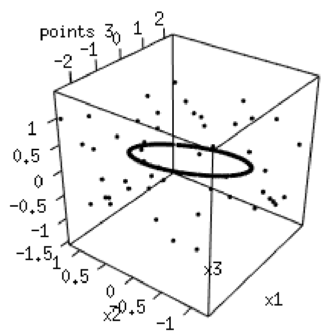


(a) (b)

Figure 2: Pair diagram and 3D plot of the points2 dataset



(a)



(b)

Figure 3: Pair diagram and 3D plot of the points3 dataset

On the figures 1, 2 and 3 we are able to visualize space like representation of the given datasets. Since our brain is kind of a pattern searching machine, we are already able to interpret these images and conclude that, the points1 dataset could be something similar of a noisy sphere*. The points3 dataset could be something similar of a noisy circle. When it comes the turn of the points2 dataset, the decision is not so obvious since it is very noisy but one could see, however, that a very noisy torus shape can be seen with a bite of imagination.

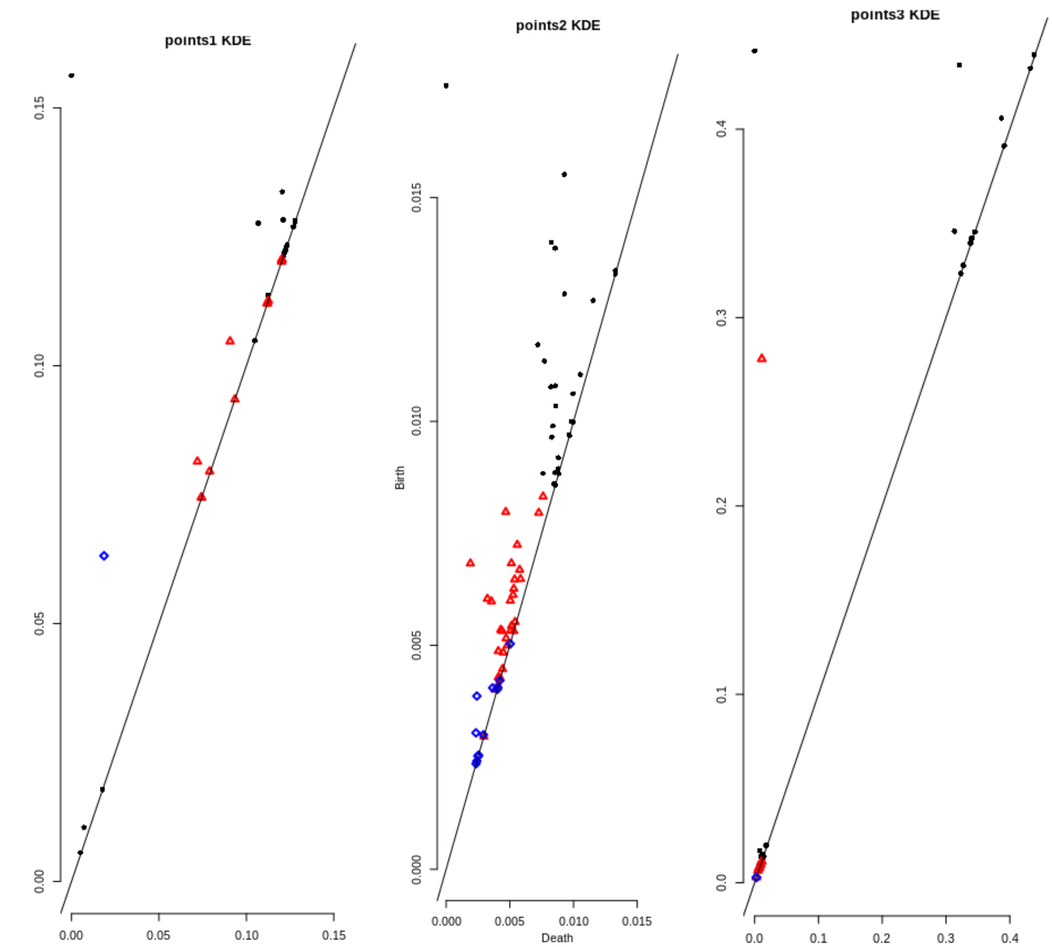
By looking the borders of the distinguishable shapes that appear on the points1 and points2 datasets, we can also suspect that the shapes are *empty* or, in other words, the torus and the sphere would be 2-d shapes.

All the previous analysis has been very qualitative and poor since our brains could, sometimes, trick us, and we have not measured anything. To really study the topological shapes behind the datasets, we will perform on the next section the persistence diagram of each one.

2 Persistence diagrams

On this section we are going to perform persistence diagrams on each of the datasets that we have already described. We have intended two strategies, to plot the KDE diagrams and to plot the Vietoris-Rips diagram. Since the second strategy is more time consuming and also the results were not clear when it came to determine the topological shape, I have finally done the analysis by looking on the KDE diagrams.

*In fact, depending on the axis representation, it would be more precise to call it ellipsoid. However, we are dealing with topological shapes so by sphere (or a shape x) we truly mean a shape topologically identical of a sphere (or x)



(a) (b) (c)

Figure 4: Persistent diagrams

On the figure 4 we can see all the KDE diagrams that have been ploted, one for each dataset. On the figure 4a, corresponing to the diagram for the points1 dataset, we can easily see a blue rhomboid that appears to be far from the diagonal respect from the other points. This fact, means that we have a shape with a 2 dimensional hole. This result confirms the suspicions that raised us the previous section, since we have postulated that it was a 2D sphere like shape. I have used a distance of 0.05 to determine which points are close enough to be neighbours.

On the figure 4c, corresponding to the diagram for the points3 dataset, we can easily see a red triangle that appears to be far from the diagonal respect from the other points. This fact means that we have a shape with a 1 dimensional hole. This result confirms the suspicions that raised us the previous section, since we have postulated that it was a circle like shape. I have used, again, a maximum distance of 0.05.

On the figure 4b, corresponding to the diagram for the points2 dataset, the results are not so obvious to interpret. This was also expected from the fact that it has been difficult to observe the shape even with our own eyes on the previous section. However we have postulated that it could be a very noisy torus. If that was the case we should obtain two 1 dimensional holes and one 2 dimensional hole (and the noise). We can see that two triangles appear to be slightly far form the diagonal respecting the other triangles, so that could prove the two one dimensional holes. The same happens with a blue rhomboid that seems to be lightly far from the diagonal respect to the other rhomboids meaning one two dimensional hole could be there. However, even in this case, the noise is comparable with the points we expect to obtain so a further analysis must be performed in order to clean the data and determine whether we are, or not, in front of a 2d torus.

3 Results

points1 $\rightarrow$ Sphere, points2 $\rightarrow$ Very noisy torus (?), points3 $\rightarrow$ Circle

Appendix

3.1 Dataset basic analysis code

Here you can see the basic code that has been done on the first place in order to have an intuition of the datasets to be dealed with.

```
#Import dataset

data1 = read.csv("/home/genis/Escritorio/tda/points1.csv")
data2 = read.csv("/home/genis/Escritorio/tda/points2.csv")
data3 = read.csv("/home/genis/Escritorio/tda/points3.csv")

# Summary

summary(data1)
summary(data2)
summary(data3)


# Visualize the dataset

pairs(data1, main = "Pair diagram of points1 Dataset", pch = 21)
pairs(data2, main = "Pair diagram of points2 Dataset", pch = 21)
pairs(data3, main = "Pair diagram of points3 Dataset", pch = 21)

# 3D Plot

library(rgl)

plot3d(data1, xlab="x1", ylab="x2", zlab = "x3", main="points 1")
plot3d(data2, xlab="x1", ylab="x2", zlab = "x3", main="points 2")
plot3d(data3, xlab="x1", ylab="x2", zlab = "x3", main="points 3")
```

3.2 Code to make the persistence diagrams of each of the datasets

Here you can see the code (that is, in fact, the continuation of the previous part) that let me plot the persistence diagrams.

```
# PERSISTENCE
library("TDA")

##Points1

Xlim <- c(-2.5,2.5)
Ylim <- c(-1.5, 1.5)
Zlim <- c(-1.5,1.5)

by <- 0.05

# Persitent homology on points1 with KDE function and VR

DiagKDE <- gridDiag(X=data1,
                    FUN = kde, h=0.3, sublevel=FALSE,
                    lim= cbind(Xlim, Ylim, Zlim), by=by,
                    library="Dionysus",printProgress = TRUE)
plot(x = DiagKDE[["diagram"]],main="points1 KDE")
#DiagVR <- ripsDiag(X=data1,
#                    maxdimension = 1, maxscale = 5, dist = 'euclidean',
#                    library="GUDHI",printProgress = FALSE)
#plot(x = DiagVR[["diagram"]],main="Vietoris-Rips Diagram")

#Points2

```

```r
Xlim <- c(-4.7,4.7)
Ylim <- c(-3.5, 3.5)
Zlim <- c(-1.5,1.5)

by <- 0.059

# Persitent homology on points2 with KDE function and VR

DiagKDE <- gridDiag(X=data2,
                        FUN = kde, h=0.45, sublevel=FALSE,
                        lim= cbind(Xlim, Ylim, Zlim), by=by,
                        library="Dionysus", printProgress = TRUE)
plot(x = DiagKDE[["diagram"]],main="points2 KDE")
#DiagVR <- ripsDiag(X=data2,
#                       maxdimension = 1, maxscale = 5, dist = 'euclidean',
#                       library="GUDHI",printProgress = FALSE)
#plot(x = DiagVR[["diagram"]],main="Vietoris-Rips Diagram")

#Points3

Xlim <- c(-2.6,2.4)
Ylim <- c(-1.25, 1.25)
Zlim <- c(-1.6,1.6)

by <- 0.05

# Persitent homology on points3 with KDE function and VR

DiagKDE <- gridDiag(X=data3,
                        FUN = kde, h=0.25, sublevel=FALSE,
                        lim= cbind(Xlim, Ylim, Zlim), by=by,
                        library="Dionysus",printProgress = TRUE)
plot(x = DiagKDE[["diagram"]],main="points3 KDE")
#DiagVR <- ripsDiag(X=data3,
#                       maxdimension = 1, maxscale = 5, dist = 'euclidean',
#                       library="GUDHI",printProgress = FALSE)
#plot(x = DiagVR[["diagram"]],main="Vietoris-Rips Diagram")
```