CS4103 Geometric Foundations of Data Analysis

Each homework should be submitted as a single .pdf document with an accompanying .py file to both Graham Ellis and Emil Sköldberg. The .pdf document should provide your answers, the methods used to obtain your answers, and an appendix listing any Python code used. The .py file should be a machine readable version of the appendix code.

The homework will be graded according to a scheme in which *content* is weighted at 70% and *presentation* is weighted at 30%.

1 Fourth Homework

Please try to submit this by 16 April as two files:
MA500_Fourth_Homework_firstname_familyname.pdf
MA500_Fourth_Homework_firstname_familyname.py

1. Use the TDAmapper package for **R** to produce a graphical representation of the data set for the Miller-Reaven diabetes study. This data comes with the TDAmapper package. Briefly compare your graphical representation with the representation from the paper

   > *MILLER R. J.: Discussion - projection pursuit. Ann. Statist. 13 , 2 (1985), 510–513. With discussion*

   which is recalled in the paper *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.*

2. Use Mapper, and other techniques if necessary, to investigate the following synthetic $n \times n$ distance matrices. Show any graphical representations that you use.

   data1.txt

   data2.txt

   data2.txt

   data4.txt

   data5.txt

   data6.txt

   Match the distance matrices to the following descriptions of data sets used to produce them. You might need to consider Cayley-Menger determinants.

   (a) Points selected from near the x-, y- or z-axes in $\mathbb{R}^3$.

   (b) Points selected from near the x- or y-axes in $\mathbb{R}^3$.

   (c) Points selected from a 2-d digital image of a starfish with 6 limbs.

   (d) Points selected from a 2-d digital image of a starfish with 5 limbs.

   (e) Points selected from a torus $\mathbb{S}^1 \times \mathbb{S}^1$ embedded in $\mathbb{R}^3$.

   (f) Points selected from an annulus in the plane.

2 Fifth Homework

Please try to submit this by 7 May as two files:
MA500_Fifth_Homework_firstname_familyname.pdf
MA500_Fifth_Homework_firstname_familyname.py

1. Download this csv file which contains data on water levels at Galway Port.
2. For time t let $x(t) = (h_0, h_2, h_4) \in \mathbb{R}^3$ denote the vector consisting of the height h_0 of the water at time t, the height h_2 of the water two hours after time t, and the height h_4 of the water four hours after time t. Choose a 2-week period, and create a set S consisting of the vectors $x(t)$ for 200 random times t in the chosen 2-week period.
3. Use the R-TDA package to compute the mod-2 persistence barcodes (or persistence diagrams if that is easier), in degrees 0 and 1, for the filtered simplicial complex constructed from S using the Euclidean metric.
4. Give a brief interpretation of your barcodes/persistence diagrams.