

## MA500/CS4102 Geometric Foundations of Data Analysis I

Homework solutions should be submitted as a single .pdf document with an accompanying .py file.

The .pdf document should contain a main part in which you present your answers to the questions, and in which you provide a description of the mathematical methods used to obtain your answers. This main part should contain no Python code. an appendix listing any Python code used. The .txt file should be a machine readable version of the appendix code which, when run, reproduces your answers.

The homework will be graded according to a scheme in which content (=correctness of your answers, choice of methods, python code) is weighted at 70% and presentation (=manner in which you present your answers, methods and code) is weighted at 30%.

### 3 Third Homework

Please submit as two files:

MA500 Third Homework firstname familyname.pdf

MA500 Third Homework firstname familyname.py

1. Implement an algorithm that applies single-linkage hierarchical clustering to an  $n \times n$  matrix of distances (or dissimilarities) and returns the corresponding barcode.
2. Create a sample  $S$  of  $n$  points in  $R^2$  that are clearly partitioned into several distinct 'clusters'. Plot the points  $S$ .
3. For the Euclidean metric, and then the taxicab metric, construct the two  $n \times n$  distance matrices for your set  $S$  of points.
4. Apply your implementation to the two matrices in (3) and display the resulting barcodes.