HW 6: Data manipulation and visualization

Let M be a dataset with *m* rows (objects) and *n* columns (attributes) where the last column represents the target attribute (e.g., with values 1 and 2 that indicate classes 1 and 2). All values of X_1-X_n in M are normalized to [0,1]. See Table 1 as an example.
Notation: $a_1=(a_{11},a_{12},a_{13},...a_{1n})$, $a_i=(a_{i1},a_{i2},a_{i3},...a_{in})$.

object	X_1	X_2	X_3	X_4	Target Class
a_1	0.3	0.2	0.1	0.5	1
a_2	0.2	0.4	0.3	0.6	1
a_3	0.4	0.2	0.0	0.3	1
a_4	0.7	0.5	0.7	0.9	1
a_5	0.9	0.7	0.6	0.7	1
a_6	0.1	0.1	0.1	0.1	2
a_7	0.1	0.2	0.1	0.2	2
a_8	0.0	0.0	0.2	0.2	2
a_9	0.2	0.1	0.4	0.4	2
a_{10}	0.3	0.4	0.5	0.6	2

1. Find all a_i in Table 1 that have property (1)

   $|a_{11}-a_{i1}|< T \;\&\; |a_{12}-a_{i2}|<T \;\&\; |a_{13}-a_{i3}|<T \;\&|a_{14}-a_{i4}|<T$ (1)

where T=0.25, i.e, you find a set M(1,T) of all objects in M such that they differ from a_1 no more than T in each coordinate.

2. Draw set M(1,T) in Collocated Paired Coordinates (CPC) with objects of class 1 shown in one color and objects of class 2 shown in another color, i.e., you are drawing objects of class 1 that are close to a_1 and objects of class 2 that are also close to a_1.

3. Draw set M(1,T) in Shifted Paired Coordinates (SPC) with objects of class 1 shown in one color and objects of class 2 shown in another color in a way that object a_1 is represented as a single 2-D point, i.e., you are drawing objects of class 1 that are close to a_1 and objects of class 2 that are also close to a_1.

4. Create a set M(1,T,1) as a subset of M(1,T). The set M(1,T,1) includes only objects of class 1. Create a set M(1,T,2) as a subset of M(1,T). The set M(1,T,2) includes only objects of class 2.

5. Create a set N= M(2)\ M(1,T,2), where M(2) is a set of all objects from class 2 in M, i.e., N(1,T,2)=M(2)\ M(1,T,2) is s set of all objects from class 2 without objects that are close to a_1.

6. Draw sets M(1,T,1) and N(1,T,2) in Collocated Paired Coordinates (CPC) with objects of class 1 shown in one color and objects of class 2 shown in another color, i.e., you are drawing objects of class 1 that are close to a_1 and objects of class 2 that are far away from a_1.

7. Draw set M(1,T,1) in Shifted Paired Coordinates (SPC) with objects of class 1 shown in one color and objects of class 2 shown in another color in a way that object a_1 is represented as a single 2-D point, i.e., you are drawing objects of class 1 that are close to a_1 and objects of class 2 that are far away from a_1.

8. Generalize 1-8 to be able to run on a table with any m up to 100 rows and any n up to 10 and with abilities to use instead of a_1 and a_i as a base point.

9. Run 1-8 for all other a_i instead of a_1.

10. Run 1-8 on three datasets with m=100 and n=10 of two classes with a1 as a first objects of the first class. In your experiments you can change a threshold of similarity T. Write a report with analysis of obtained visualizations. In 2 and 3 you may get very similar graphs for objects from two classes. In this case you may try to find subtle “micro” features that discriminate objects of two classes.