

[15] 1.) Create the single linkage dendrogram for the data set with distance matrix

$$\begin{matrix} & a & b & c & d & e \\ b & 7 & & & & \\ c & 2 & 5 & & & \\ d & 10 & 3 & 8 & & \\ e & 6 & 1 & 4 & 4 & \end{matrix}$$

List the clusters at each merging height:

The clusters at time 0 are $\{a\}, \{b\}, \{c\}, \{d\}, \{e\}$

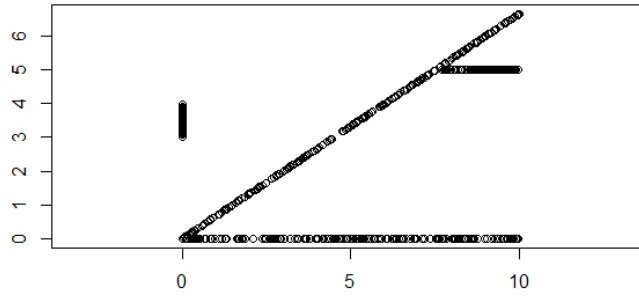
The clusters at merging height = _____ are _____

The clusters at merging height = _____ are _____

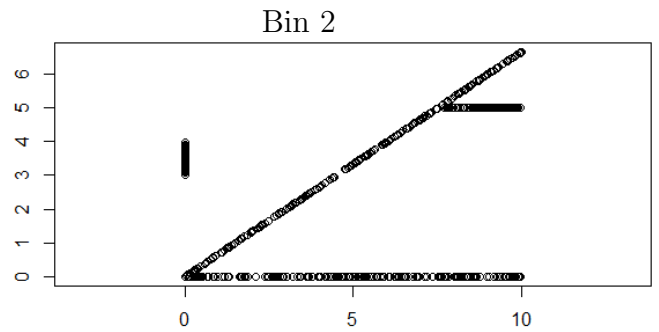
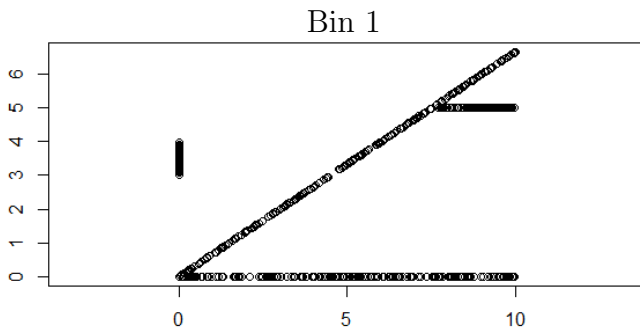
The clusters at merging height = _____ are _____

The clusters at merging height = _____ are _____

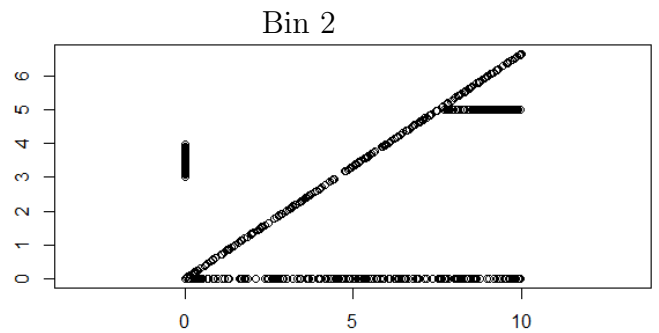
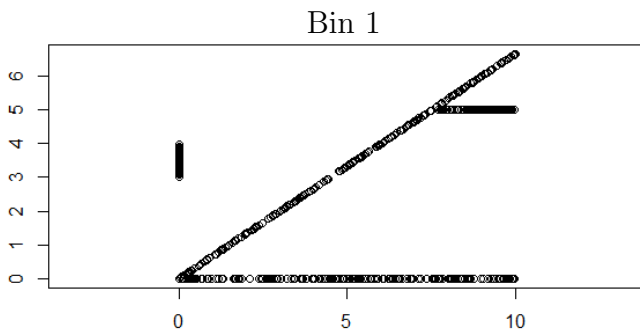
[20] 2.) Suppose TDA mapper is used to analyze the following data set where filter function = projection to the x-axis and 2 bins are created using 2 intervals with 50% overlap,



a.) Draw one rectangle in each of the figures below to indicate a bin:



b.) Circle the clusters in Bin 1 in the figure below on the left, and circle the clusters in Bin 2 in the figure below on the right:



c.) Based on your answers above, draw the output of TDA mapper:

[6] 3.) For the following matrix, A , find its eigenvalues and corresponding eigenvectors. For each eigenvalue, find both its algebraic multiplicity and the geometric multiplicity.

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

An eigenvalue of A is _____. It has algebraic multiplicity = _____ and geometric multiplicity = _____.

An eigenvector corresponding to this eigenvalue is _____.

A 2nd eigenvalue of A is _____ with algebraic multiplicity = _____ and geometric multiplicity = _____.

An eigenvector corresponding to this eigenvalue is _____.

[4] 4.) Calculate the following, given the following distance matrix:

$$\begin{matrix} & a & b & c & d \\ b & 7 & & & \\ c & 2 & 5 & & \\ d & 10 & 3 & 8 & \end{matrix}$$

$\delta_1(b) =$ _____ $\delta_2(b) =$ _____ $\delta_3(b) =$ _____ $\delta_4(b) =$ _____

Recall $\delta_k(b)$ refers to the knn distance.

Multiple Choice: **Circle the best answer.**

- [5] 5.) The key idea(s) of topology that make extracting patterns via TDA mapper possible are
- i.) Coordinate free: There is no dependence on the coordinate system chosen. Thus one can compare data derived from different platforms.
 - ii.) Invariant under small deformations and thus less sensitive to noise.
 - iii.) The output of TDA mapper gives compressed representations of the shape of the data (depending on filter function used).
 - iv.) All of the above.
 - v.) None of the above.
-

- [5] 6.) One **benefit** to the variety of parameters one can choose in TDA mapper is that
- i.) You can try a variety of parameters until you get the output that you want.
 - ii.) You can probe the data set from a variety of different perspectives.
 - iii.) You can look for green jelly beans.
 - iv.) You may get a false positive.
-

- [5] 7.) If two vertices are close in the graph created by TDA mapper, then the data points represented by these vertices are close in the original data set.
- i.) True
 - ii.) False
-

- [5] 8.) An edge of the graph created by TDA mapper can represent the intersection of two clusters from the same bin.
- i.) True
 - ii.) False
-

- [5] 9.) In the video *Applications of TDA to the Understanding of Disease and Drug Discovery*, Pek Lum, Ayasdi, in the diabetes graph, what does it mean when a node is colored red?
- i.) This node represents a patient that must have diabetes.
 - ii.) This node represents a cluster of patients who all must have type II diabetes.
 - iii.) This node represents a cluster of patients where most of them have type II diabetes.
 - iv.) This node represents a cluster of patients where 50% have type II diabetes.
 - v.) None of the above.
-

[5] 10.) The output of python mapper applied to a particular data set is given below. How many data points are in this data set?



- i.) 1 ii.) 2 iii.) 3 iv.) 4 v.) 5
- vi.) There is insufficient information to determine the number of data points in this data set.
-

[5] 11.) Suppose you want to analyze flight departure delays at a small airport over a period of 365 days. Five airlines fly out of this airport. If 100 flights depart from this airport every day and if each coordinate of an element \mathbf{x} in your data set represents the delay time in minutes for the i th departure of the day, then your data set lives in \mathbb{R}^n where $n =$

- i.) 1 ii.) 5 iii.) 12 iv.) 20 v.) 60 vi.) 100 vii.) 365
-

[5] 12.) Suppose my data set consists of the following points:

$$(1, 3), (1.1, 3), (0.9, 3), (4, 3), (7, 3), (0.1, 7), (4, 7), (5, 7)$$

If my filter function is projection to the y -axis and I use 2 intervals with 50% overlap to determine my overlapping bins, then the output of TDA mapper will contain k edges, where $k =$

- i.) 0 ii.) 1 iii.) 2 iv.) 3 v.) 4 vi.) 5 vii.) 6 viii.) 7 ix.) 8
- x.) $0 \leq k \leq 8$ xi.) $1 \leq k \leq 8$ xii.) $0 \leq k \leq \frac{(8)(7)}{2}$ xiii.) $1 \leq k \leq \frac{(8)(7)}{2}$
-

[5] 13.) Suppose A is a **symmetric** matrix whose characteristic polynomial is $(r - 1)^2(r + 3)$. What can you say about the rank of A ?

- i.) 0, 1, 2 ii.) 0, 1 iii.) 0, 2 iv.) 1, 2
- v.) 1, 2, 3 vi.) 1, 3 vii.) 2, 3 viii.) 1, 2
- ix.) 0 x.) 1 xi.) 2 xi.) 3
-

[5] 14.) Suppose M is a **symmetric** matrix whose characteristic polynomial is $r(r - 4)^3(r + 4)^2$. What can you say about the geometric multiplicity of the eigenvalue 4?

- i.) 0, 1, 2 ii.) 0, 1 iii.) 0, 2 iv.) 1, 2
- v.) 1, 2, 3 vi.) 1, 3 vii.) 2, 3 viii.) 1, 2
- ix.) 0 x.) 1 xi.) 2 xi.) 3
-

[5] 15.) Suppose A is a 9x9 matrix, and suppose 0 is an eigenvalue of A with algebraic multiplicity = geometric multiplicity = 4. Also assume that 3 is an eigenvalue of A with algebraic multiplicity = geometric multiplicity = 5. What is the rank of A ?

- i.) 1 ii.) 2 iii.) 3 iv.) 4
- v.) 5 vi.) 6 vii.) 7 viii.) 8