

Homework Assignment 5 [30 pts]

STAT430 Unsupervised Learning – Spring 2022

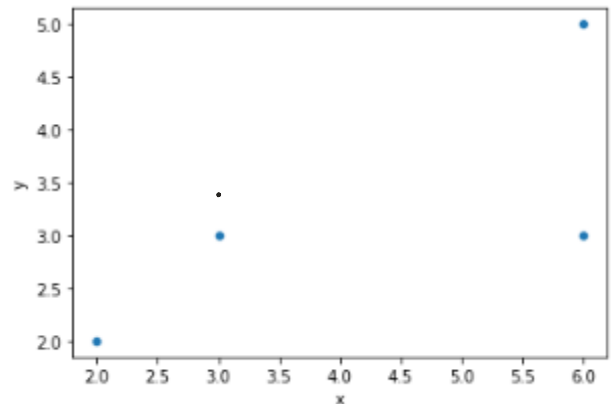
Due: Friday, February 24 on Canvas at 11:59pm CST.

Problem	Points
1	0.25
2	1.25
3.1	1
3.2	1.5
3.3	1.5
4.1	1.5
4.2	1
5.1	1
5.2	1
6.1	1
6.2	1
7.1	1
7.2	1
8.1	1
8.2	1
9.1	1
9.2	1
10.1	1.5
10.2	1.5
11	6
12	3

Questions 1-10: See the Jupyter notebook

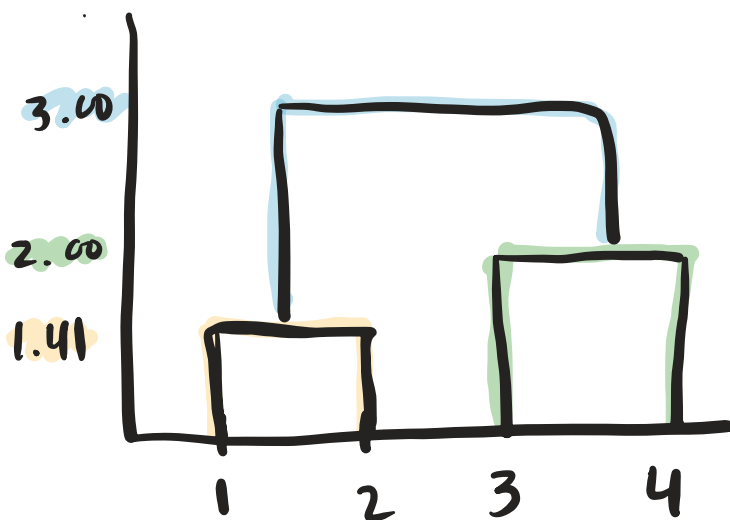
Question 11: For the dataset below, sketch three dendrograms: a.) one found with agglomerative hierarchical clustering with single linkage proximity, b.) one found with agglomerative hierarchical clustering with complete linkage proximity, and c.) one found with agglomerative hierarchical clustering with average linkage proximity. Show your work and make sure your dendrograms are drawn to scale and labeled appropriately.

	Data	
	x	y
Object 1	2	2
Object 2	3	3
Object 3	6	6
Object 4	6	5



	Distance Matrix			
	Object 1	Object 2	Object 3	Object 4
Object 1	0	1.41	4.12	5
Object 2	1.41	0	3	3.61
Object 3	4.12	3	0	2
Object 4	5	3.61	2	0

Single Linkage.



[Smallest Distance]
UPDATED MATRIX

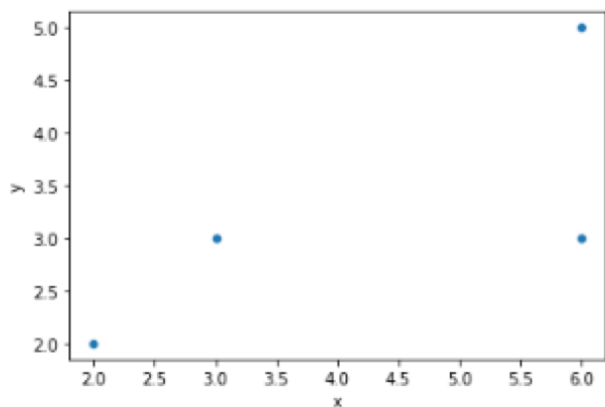
a.

	1, 2	3	4
1, 2	0	3	3.61
3	3	0	2
4	3.61	2	0

b.

	1, 2	3, 4
1, 2	0	3
3, 4	3	0

Complete linkage



	Distance Matrix			
	Object 1	Object 2	Object 3	Object 4
Object 1	0	1.41	4.12	5
Object 2	1.41	0	3	3.61
Object 3	4.12	3	0	2
Object 4	5	3.61	2	0

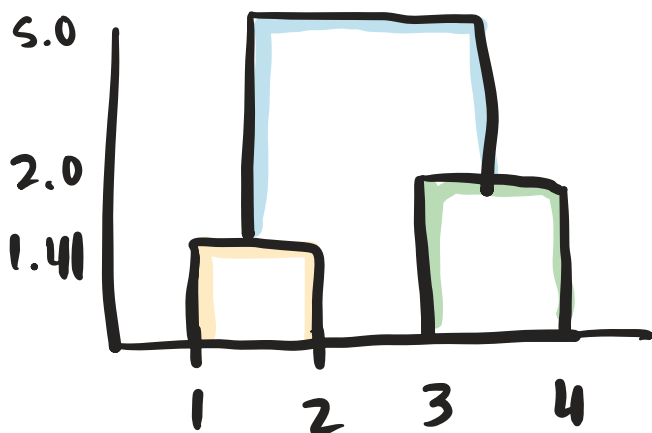
EMax Distance]
UPDATED MATRIX

a.

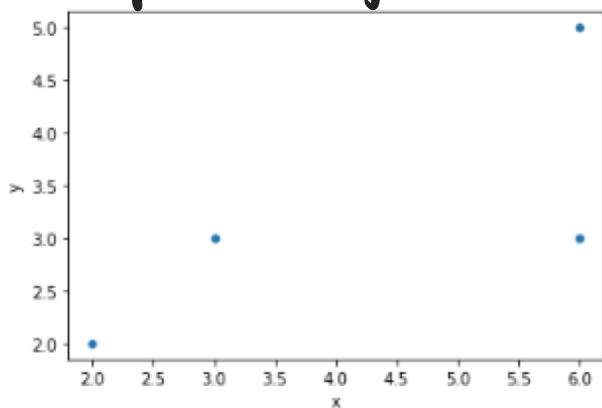
1,2	1,2	3	4
3	0	4.12	5
4	4.12	0	2
	5	2	0

b

1,2	1,2	3,4
3	0	5.0
3,4	5.0	0



Average linkage



	Distance Matrix			
	Object 1	Object 2	Object 3	Object 4
Object 1	0	1.41	4.12	5
Object 2	1.41	0	3	3.61
Object 3	4.12	3	0	2
Object 4	5	3.61	2	0

UPDATED MATRIX

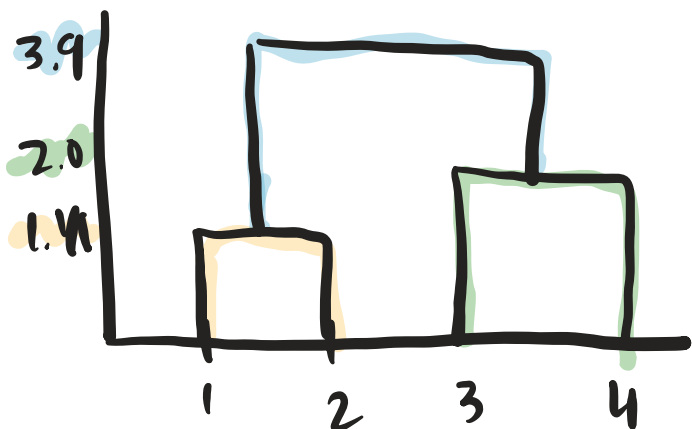
a.

1,2	1,2	3	4
3	0	3.56	4.3
4	3.56	0	2
	4.3	2	0

b

1,2	1,2	3,4
3	0	3.93
3,4	3.93	0

$$\begin{aligned} &1,2 - 3 \\ &(4.12 + 3) / 2 = 3.56 \\ &1,2 - 4 \\ &(5 + 3.61) / 2 = 4.31 \\ &1,2 - 3,4 \\ &(3.56 + 4.3) / 2 = 3.93 \end{aligned}$$

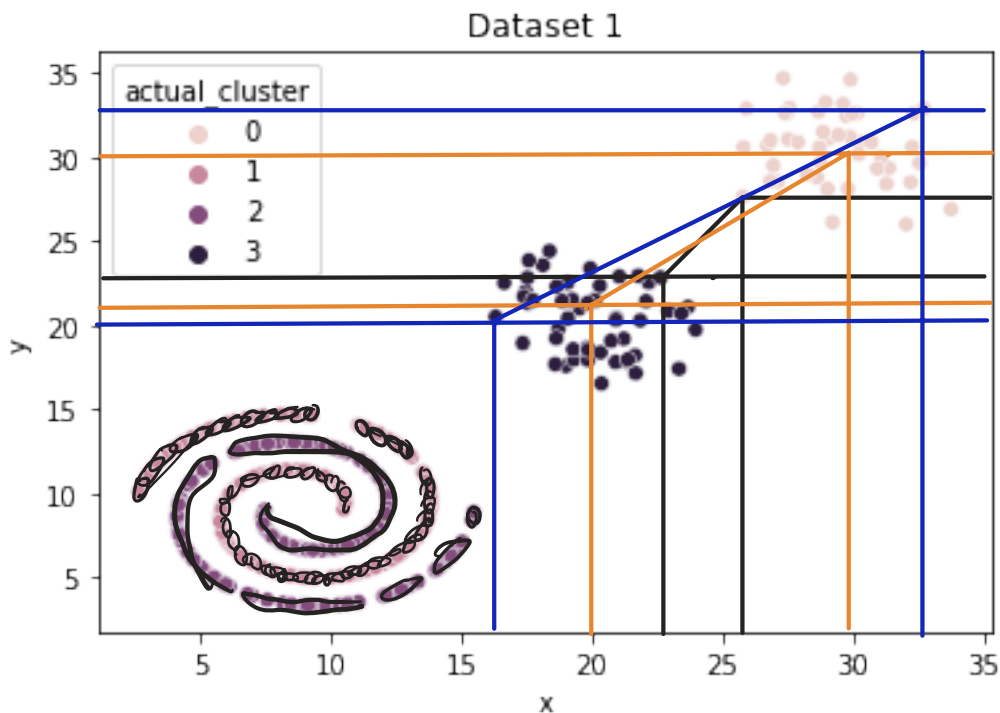


Question 12: We have clustered the dataset below using hierarchical agglomerative clustering using each of the following linkage functions:

- Single linkage
- Complete linkage
- Average linkage
- Ward's linkage

Each of the dendrograms for these linkage functions are also given below. State which dendrogram corresponds to each linkage function.

Hint: This dataset also has preassigned class labels {1,2,3,4}. Rather than label each unique object id in the dendrograms, we instead give the pre-assigned class label of the object instead (Zoom in to see it).



$$10^2 + 10^2$$

$$100 + 100$$

$$12$$

$$\sqrt{200}$$

$$20^2 + 15^2$$

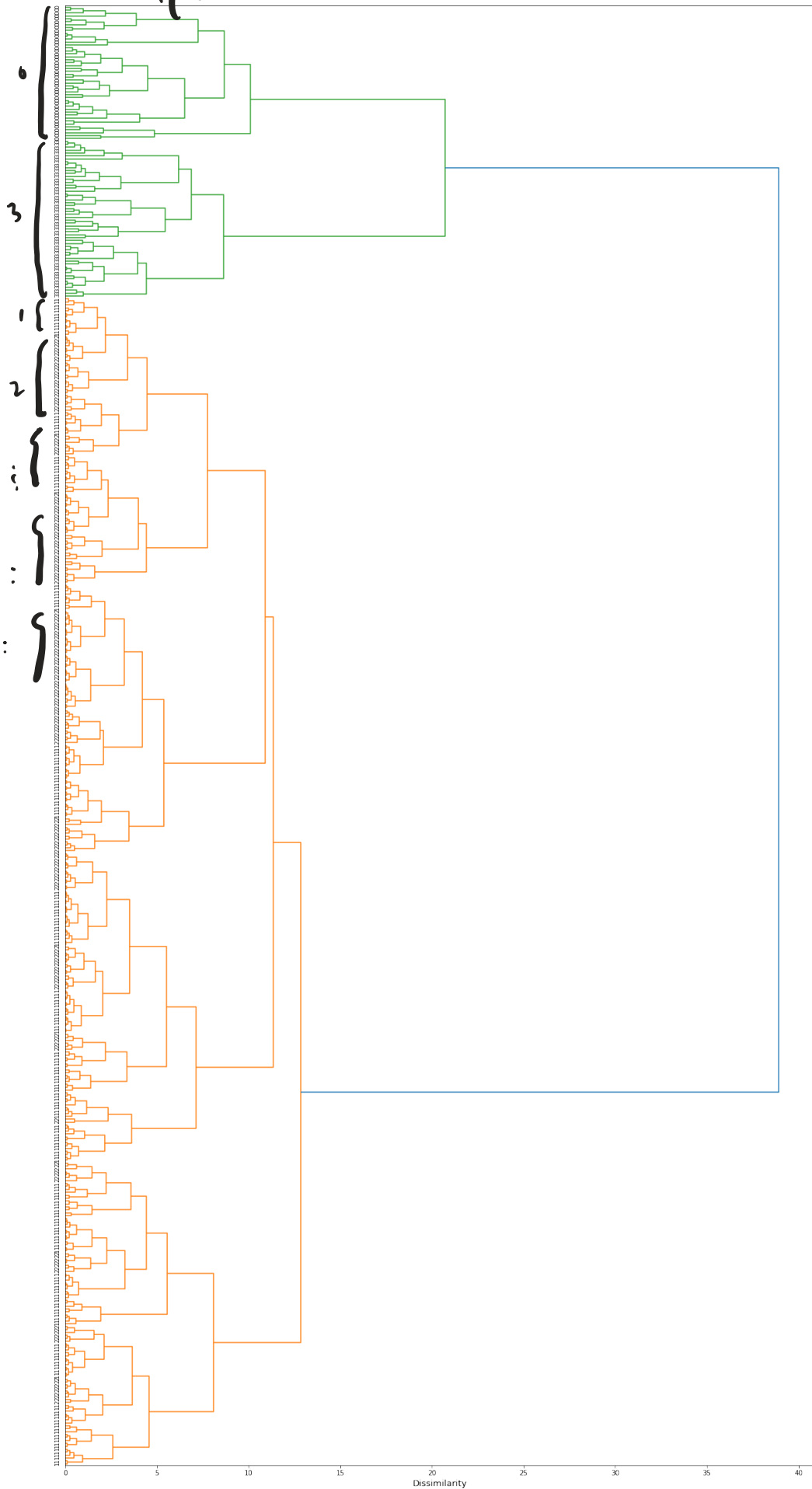
$$625$$

$$\sqrt{625} \approx 25$$

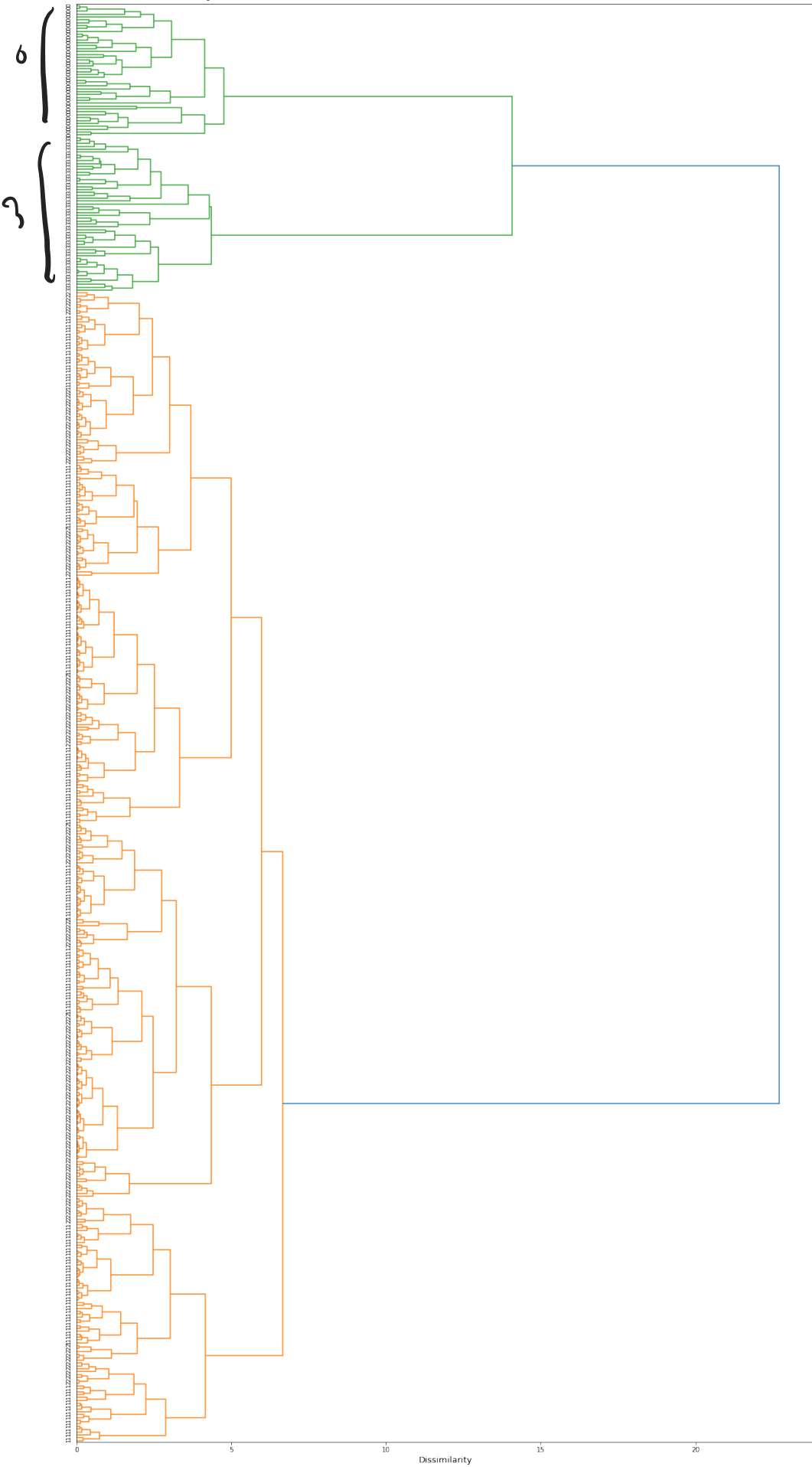
$$2^2 + 2^2$$

$$\sqrt{8}$$

Complete



Average



Single

