Simon Lizarraba CU
Simon13.

Homework Assignment 3 – [30 points]

STAT437 Unsupervised Learning – Fall 2023

Due: Friday, September 15 on Canvas

Reference the attached Jupyter notebook for the case study 1 and 2 questions.

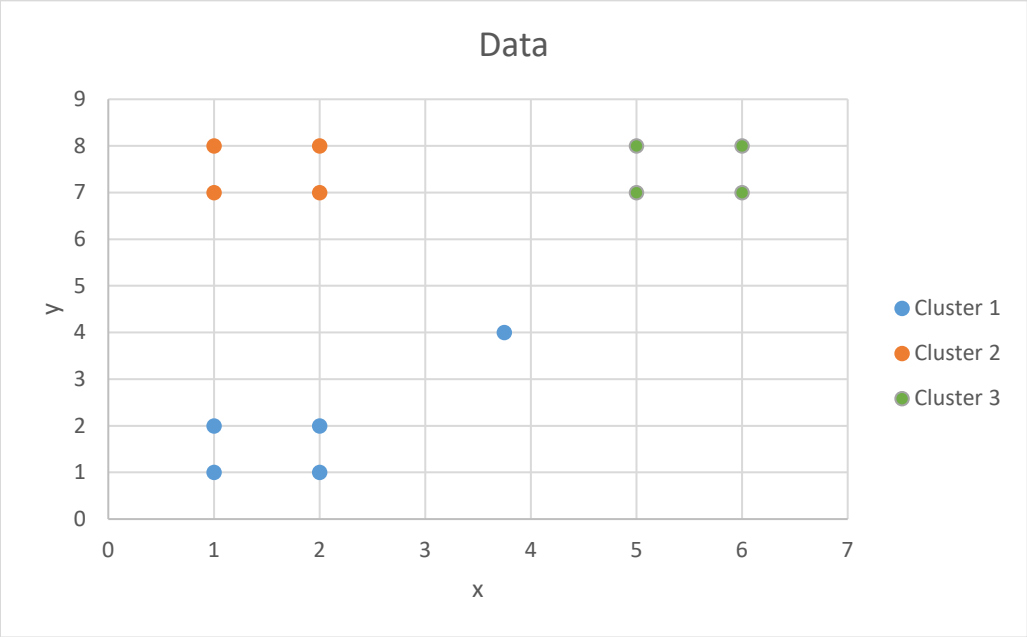
Pdf Questions	Points
1	2.5
2	2.5
3	2
Case Study 1 Questions	**Points**
1.1	0.25
1.2	0.25
1.3	0.5
1.4	1
1.5	0.75
1.6	1
2.1	0.75
2.2	0.25
2.3	0.25
2.4	0.5
2.5	1
2.6	0.5
Case Study 2 Questions	**Points**
1.1	0.25
1.2	0.25
1.3	1
1.4	0.75
2.1	0.5
2.2	0.25
2.3	0.5
3	0.5
4.1	0.75
4.2	0.5
5	0.75
6.1	0.75
6.2	0.75
6.3	0.75
6.4	0.75
7.1	1.5
7.2	0.75
8	1
9.1	0.5
9.2	0.5
9.3	0.5
9.4	0.75
10.1	0.75
10.2	0.75

Question #1:

Calculate the silhouette score of object 5 using the information below. Then interpret what this silhouette score says about object 5 with respect to this clustering.

	Data			
		x	y	Distance Object 5 (3.75, 4) is away from this object.
Cluster 1	Object 1	1	1	4.07
	Object 2	2	2	2.66
	Object 3	1	2	3.40
	Object 4	2	1	3.47
	Object 5	3.75	4	--
Cluster 2	Object 6	1	7	4.07
	Object 7	1	8	4.85
	Object 8	2	7	3.47
	Object 9	2	8	4.37
Cluster 3	Object 10	5	7	3.25
	Object 11	5	8	4.19
	Object 12	6	7	3.75
	Object 13	6	8	4.59

The silhouette score is equal to 0.138. This means that the object is relatively close in a similar way to other clusters. This can be confirmed by the plot, where it's possible to see the location of the point in close proximity to the other clusters



COHESION METRIC

$a_i = \frac{1}{|c_k|-1} \sum dist(x_i, x_j)$

$a = \frac{1}{4}(13.6)$

$a = 3.4$

SEPARATION METRIC

$b = \min_{k \neq k} \frac{1}{|c_k|} \sum dist(x_i, x_j)$

To cluster 2: $\frac{1}{4}(16.76) = 4.19$

To cluster 3: $\frac{1}{4}(15.78) = 3.945$

$b = \min(4.19, 3.945) = 3.945$

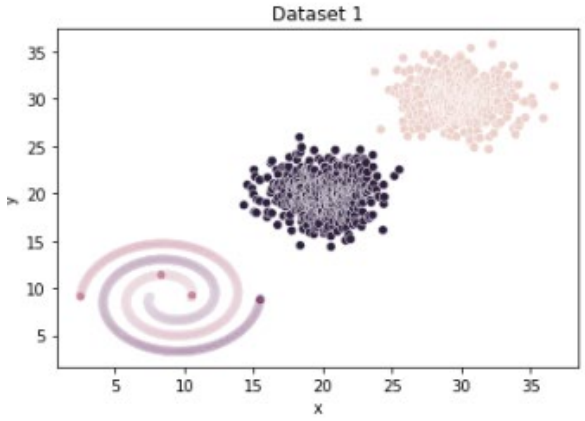
$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}$

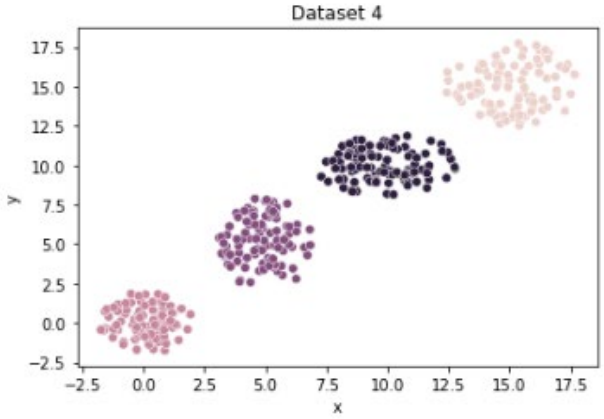
$= \frac{3.945 - 3.400}{3.945}$

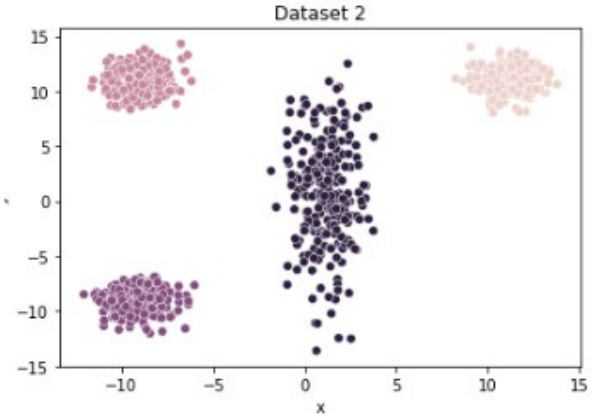
$S_i = 0.138$

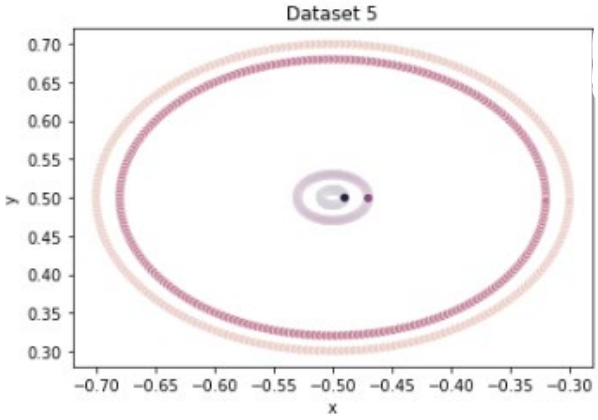
Question #2:

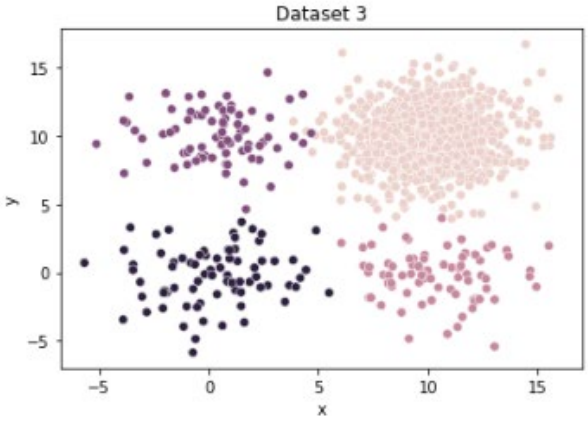
Match one of the 5 datasets and clusterings (1-5) to the one of the five silhouette plots (A-E) that were created from one of these datasets and clusterings.



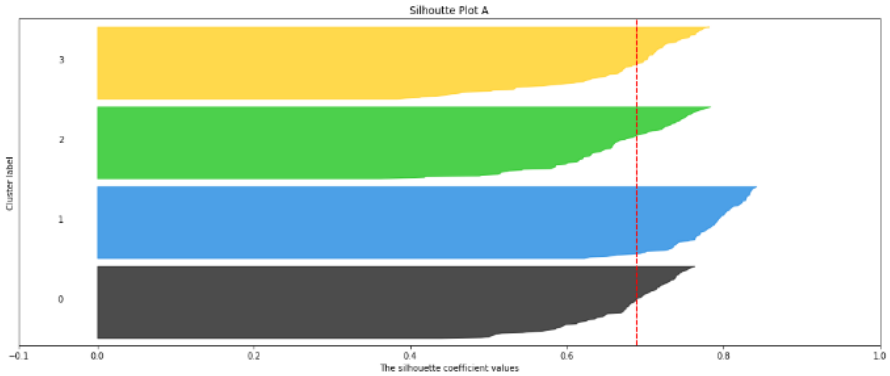






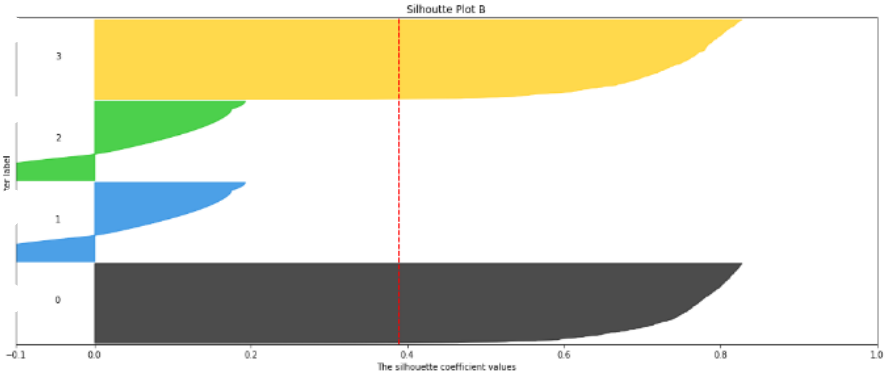


A.



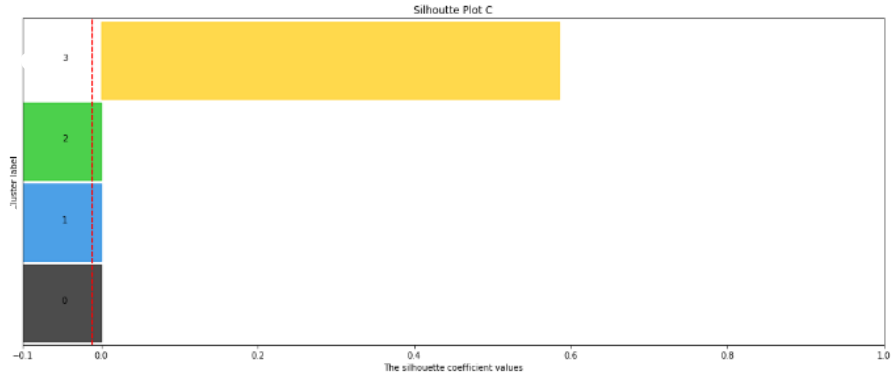
DATASET 4.

B.



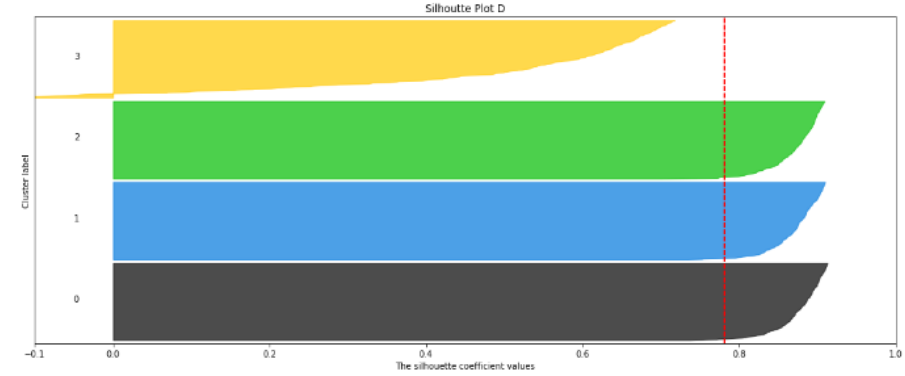
DATASET 1.

C.



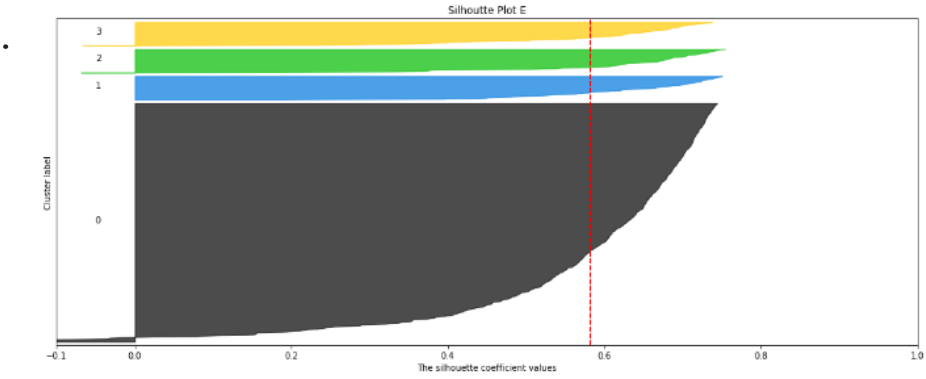
DATASET 5

D.



DATASET 2

E.



DATASET 3

Question #3:

Suppose we cluster a dataset comprised of three objects into the following clustering with k=2 clusters shown below (represented using cluster labels 0 and 1). Call this **clustering 1.** Find another clustering of this dataset of 3 objects **(call it clustering 2)** in which the **rand index** of clustering 1 and clustering 2 is **1/3**.

	Clustering 1
Object 1	0
Object 2	0
Object 3	1

clustering 2

0
1
0

$$\text{Rand} = \frac{f_{00} + f_{11}}{f_{00} + f_{01} + f_{10} + f_{11}} = \frac{1+0}{1+1+1+0} = \frac{1}{3}$$

$f_{00} = (2,3)$. Apart in both = 1

$f_{01} = (1,3)$ together 2, apart 1 = 1

$f_{10} = (1,2)$. together 1, apart 2 = 1

$f_{11} = 0$ No pair is in the same cluster