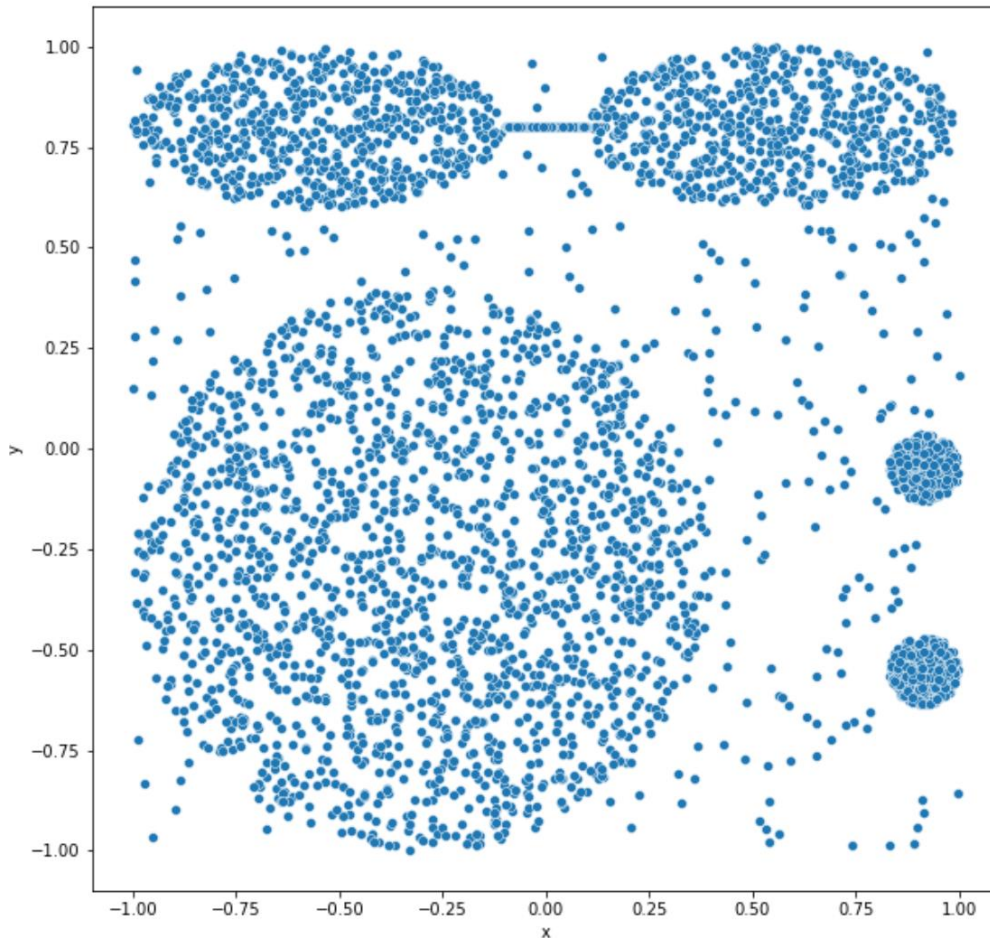# Homework Assignment 2

STAT430 Unsupervised Learning - Fall 2021

Due: Friday, September 10 on Compass

Question #1: [1.5 pt]

What type of cluster definition would be most useful in describing the five clusters represented in the plot shown below?

* Clusters are well separated
* There are two density based clusters
* Dense clusters on bottom right confirm to prototype based cluster.

Question #2: [3 pt]

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.

		x	y	Distance Object 5 is away from this object.
	Data			
Cluster 1	Object 1	1	1	4.53
	Object 2	2	2	3.54
	Object 3	1	2	3.54
	Object 4	2	1	4.53
	Object 5	1.5	5.5	--
Cluster 2	Object 6	1	7	1.58
	Object 7	1	8	2.55
	Object 8	2	7	1.58
	Object 9	2	8	2.55
Cluster 3	Object 10	5	7	3.81
	Object 11	5	8	4.30
	Object 12	6	7	4.74
	Object 13	6	8	5.15

```
a_5 = (4.53+3.54+3.54+4.53)/4
    = 4.035

b_5_orng = (1.58+2.55+1.58+2.55)/4
         = 2.065

b_5_grn = (3.81+4.30+4.74+5.15)/4
        = 4.5

Hence, b_5 = 2.065

Silhouette_5 = (2.065-4.035)/4.035
             = -0.48823
```

```
* Silhouette score of object 5 is
-0.48823. The object 5 is actually
closer to cluster_2 (orange) than
cluster_1 (blue).
```
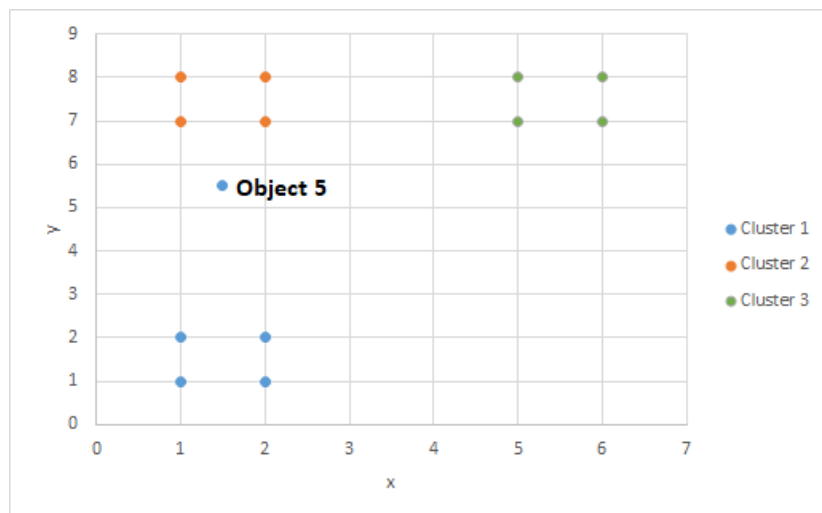
Question #3: [2 pt]

Suppose we have a dataset comprised of three objects (*object 1, object 2, and object 3*). Each object has two dimensions (ie. x and y dimensions). Suppose we come up with a clustering that assigns *objects 1 and 2* to cluster 1 and *object 3* to cluster 2 (see below). Come up with an example of (x,y) values for *objects 1, 2, and 3* below such that *object 2* has a silhouette score of exactly 0. *(Ie: fill in the blanks below).*

Clustering:

Cluster 1:

      Object 1: (____x1 = 1____, ____y1 = 1____)

      Object 2: (____x2 = 2____, ____y2 = 2____)

Cluster 2:

      Object 3: (____x3 = 3____, ____y3 = 1____)

```
b_2_1 = 1.414
b_2_3 = 1.414

Hence b_2 = 1.414

Now, a_2 = 1.414

Hence Silhouette score of object 2 is zero. (a_2 = b_2)
```
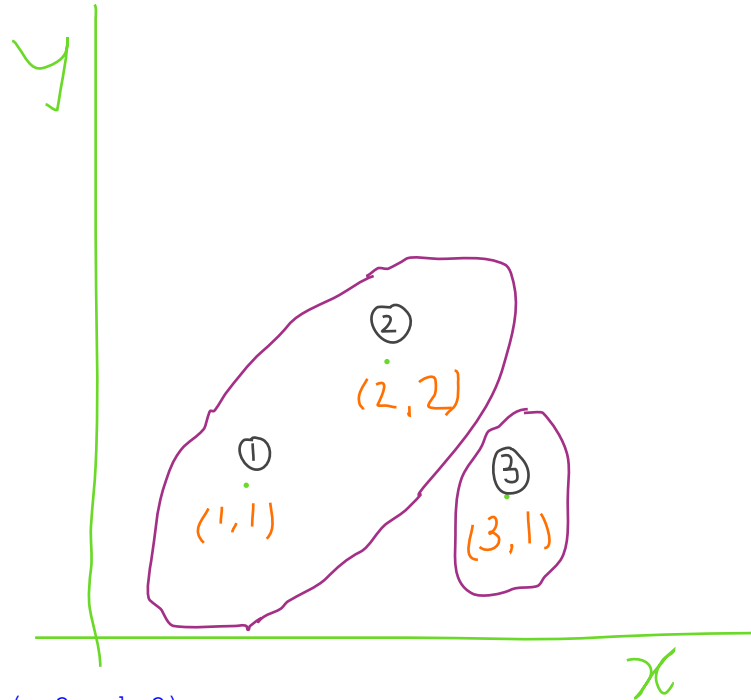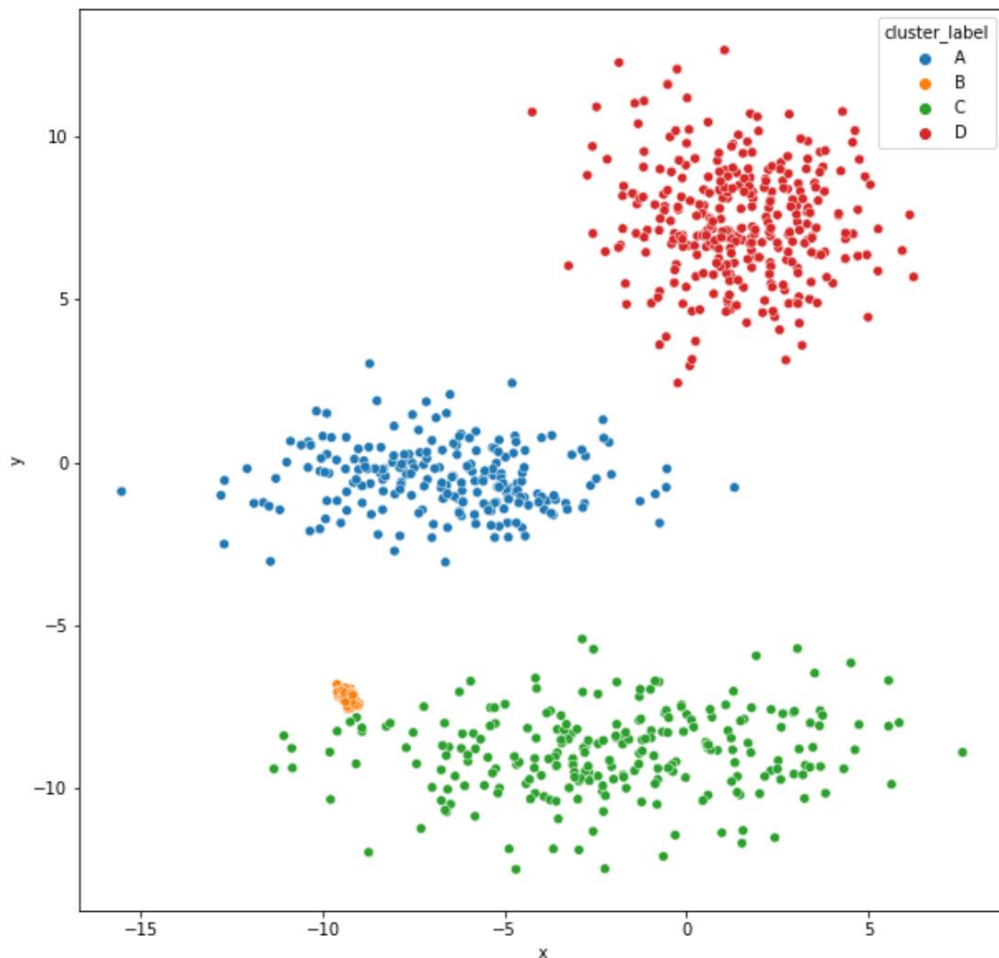
Question 4: [4 pt]

The silhouette plot below was created from the following clustering and dataset shown in the scatterplot below. Match the cluster labels (A-D) shown in the scatterplot to the corresponding cluster labels (0-3) in the silhouette plot. Explanations are not required, but may help with partial credit if you get something wrong.

C-1
> Some of the objects are very close to cluster B than C giving negative Silhouette for some objects
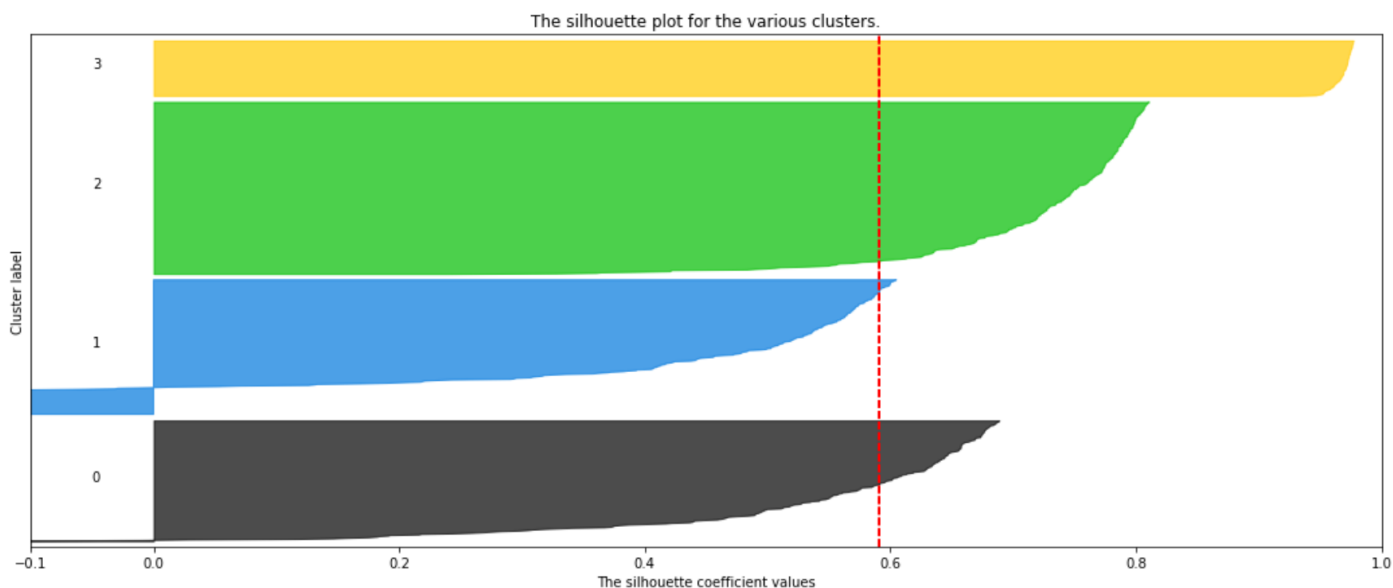
B-3
> All objects are tightly packed and closer to its centroid, giving Silhouette of ~1 for all objects

D-2
Has very similar looking plot with A. There are fewer objects away from its centroid and closer to other cluster compared to cluster A.

A-0
> Objects are scattered and away from centroid

The silhouette plot for the various clusters.

Question #5:

1. Download the Assignment_02.zip file from Compass.
2. Edit the Jupyter notebook (.ipynb) file to complete/answer questions for question 5.
3. Submit your completed Jupyter notebook (.ipynb) file as well as any other files you used to answer Question 1-4 to compass.