

Homework Assignment 1 – [30 points]

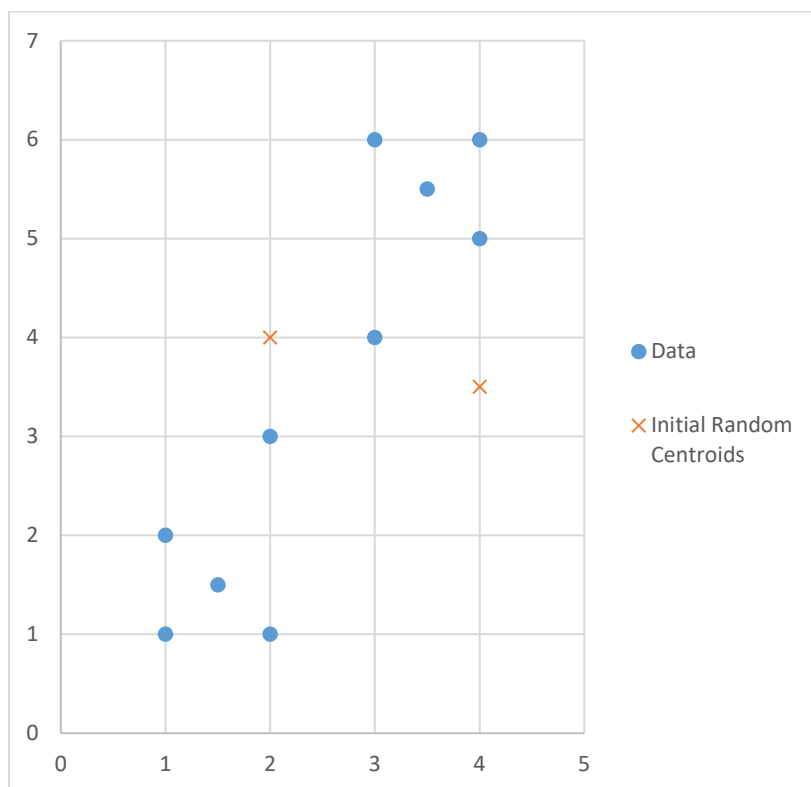
STAT437 Unsupervised Learning – Spring 2025

Due: Friday, January 31 11:59pm CST on Canvas

Problem	Points
1	3.5
2	4
3.a	0.75
3.b	0.75
3.c	1.5
3.d	1.5
4.1	0.5
4.2	0.5
4.3	1
4.4	1
4.5	1.5
4.6	1
4.7	1
4.8	1
4.9	1.5
4.10	2
5.a	1
5.b	1
6	2.5
7	2.5

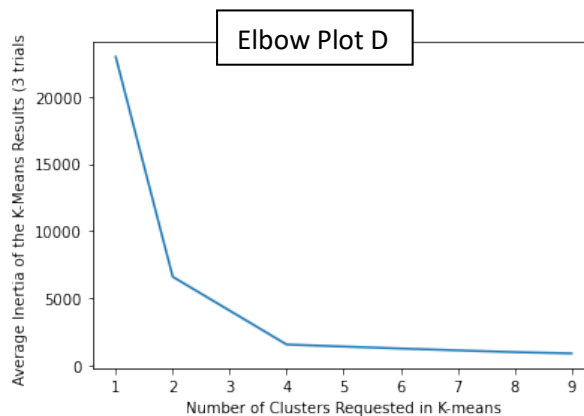
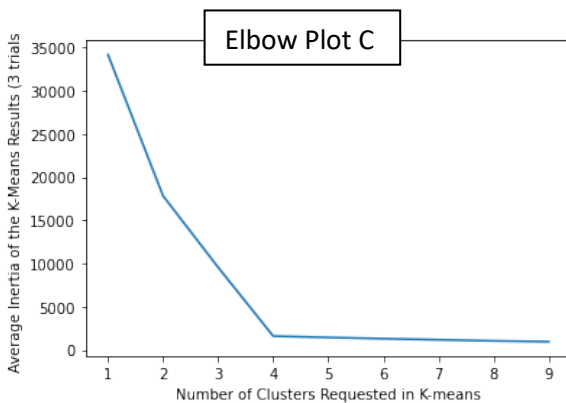
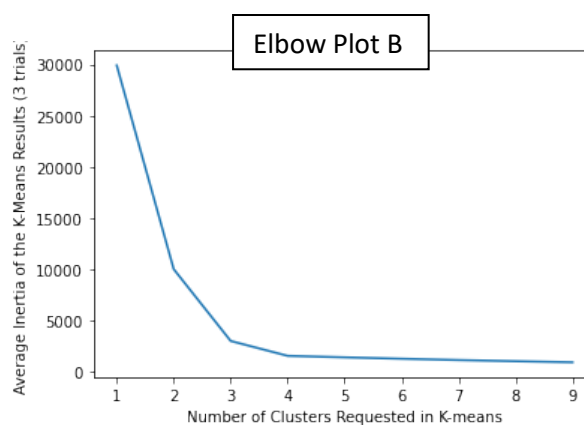
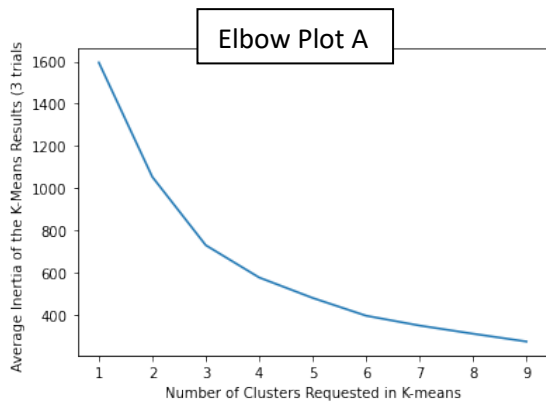
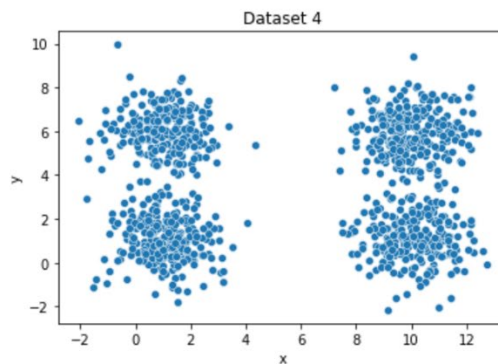
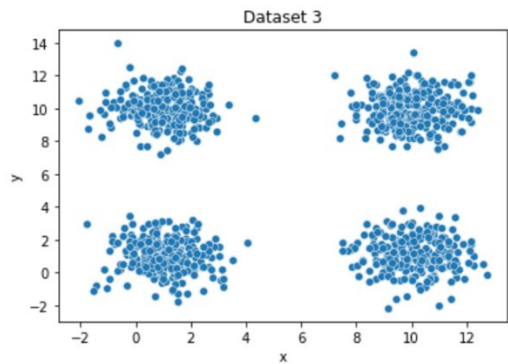
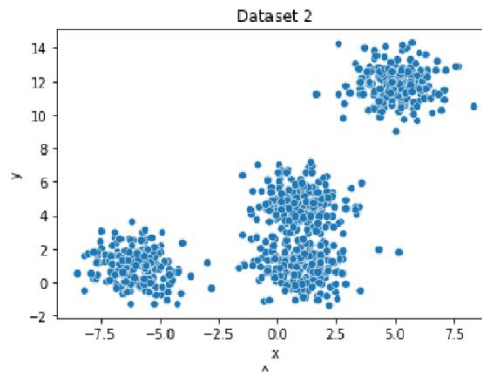
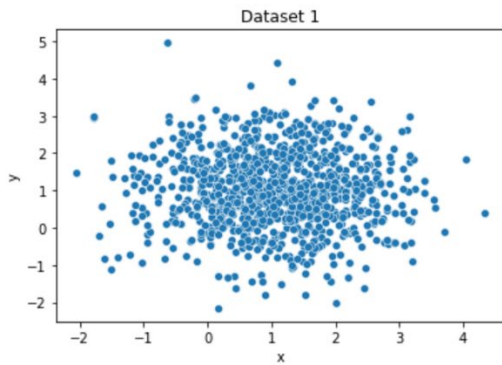
Question #1: Plotted and shown below is a two-dimensional dataset with 10 objects. Also plotted below are two centroids that have been randomly initialized to be (2,4) and (4,3.5). What will be the NEXT position of the two centroids in the first step of the k-means algorithm? Show your work.

	Data		Additional Information	
	x	y	Squared Distance to Initial Random Centroid 1 (2,4)	Squared Distance to Initial Random Centroid 2 (4,3.5)
Object 1	1	1	10.00	15.25
Object 2	2	3	1.00	4.25
Object 3	1	2	5.00	11.25
Object 5	1.5	1.5	6.50	10.25
Object 7	3	6	5.00	7.25
Object 4	2	1	9.00	10.25
Object 6	3	4	1.00	1.25
Object 8	4	5	5.00	2.25
Object 9	4	6	8.00	6.25
Object 10	3.5	5.5	4.50	4.25

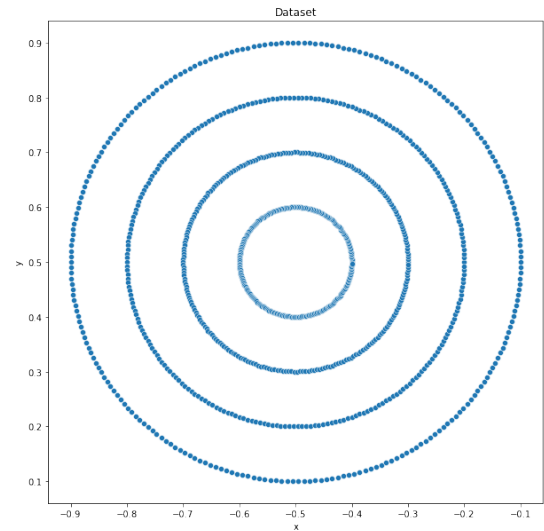


Question #2: Match the dataset to the k-means elbow plot that was created from this dataset.

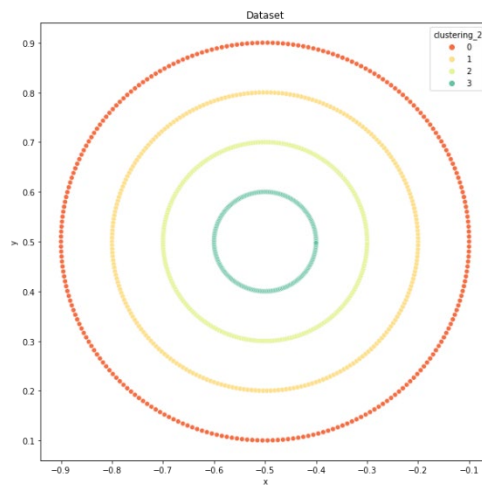
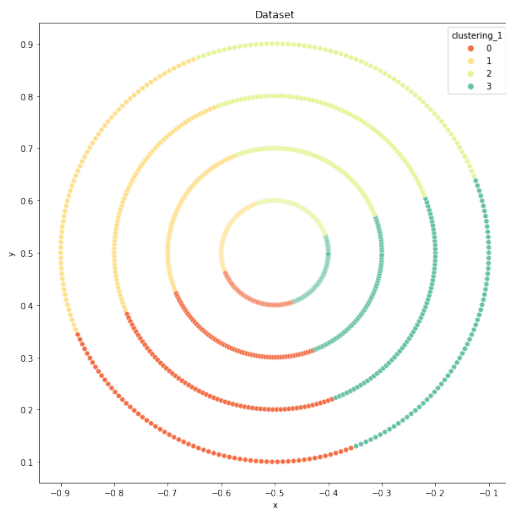
Hint: To answer this question you will need to approximate what each of the k-means elbow plots will look like for each of the 4 datasets below. That is, for a given dataset and a given cluster number k ($k=1,2,3,4,5$) think about what the inertia of that “optimal” k-means clustering would be.



Question #3: The data displayed below shows four “clusters”. We can see that in a dataset such as this, an “inherent” cluster is one of the four concentric circles of points.



Displayed below are two clusterings of the same dataset (ie. Clustering 1 and Clustering 2). Each clustering has 4 clusters, which are color-coded.



- For Clustering 1, approximate where the centroids of the four clusters would be.
- For Clustering 2, approximate where the centroids of the four clusters would be.
- Which of the two clusterings (ie. clustering 1 or clustering 2) do you think would have smaller inertia? Explain your answer.

- d. Do you think the k-means clustering algorithm will work well for this dataset? Why or why not?

Question #4

1. Edit the Jupyter notebook (.ipynb) file to complete/answer questions 4.1-4.10.
2. Submit your completed Jupyter notebook (.ipynb) file as well as any other files you used to answer Questions 1-4 to Canvas.

Question #5 Conceptual Questions

- a. **True or False:** The k-Means clustering algorithm (searching for k clusters) is guaranteed to find the clustering with the absolute lowest possible inertia for a given dataset.

- b. **True or False:** If a dataset has two “inherent” clusters where one cluster has way more observations in it than the other, then this means that the k-means clustering algorithm will not be able to successfully identify these two clusters.

Question #6

Complete the “Getting to Know you Survey” on Canvas quizzes.

Question #7: Video

Pretend you are a TA for this class. In a 3-4 minute video explain to a “student” in the class the following.

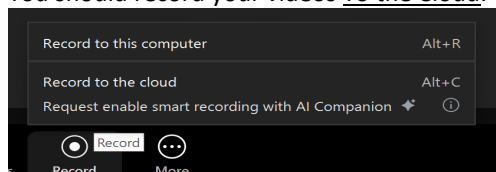
1. How you got your answer for #1.
2. How you got your answer for #2.

IMPORTANT Video Element of ALL Homework Assignments:

- In order to receive points for each video submission, you need to do **ALL** of the following.
 - Have your camera on.
 - Show your FULL screen in Zoom (not just a particular application).
 - We should be able to hear the audio. Make sure to turn your mic on.
 - You should give a good faith attempt to answer the prompt.
 - Your video meet the minimum time requirement.
 - It should not sound like you are just reading off a script.
 - It's ok if your video recording is not the most eloquent. What's important is that you are putting together YOUR authentic thoughts on your particular understanding of the assignment and the lecture content.

How to Submit Videos:

- You should record your videos in your UIUC Zoom client.
- You should record your videos To the Cloud.



- You can find your recording link at <https://illinois.zoom.us/recording/>.
- Click on the corresponding video and Copy shareable link to paste the link to the video prompt in the corresponding Jupyter notebook.