Homework Assignment 12 [30 points]

STAT430 Unsupervised Learning - Fall 2021

Due: Friday, November 19 on Compass at 11:59pm CST.

Case Study 1 and 2: See Jupyter notebooks.

Parts 3 and 4: See this pdf.

Problem	Points
Case Study 1	
1.1	1
1.2	0.75
1.3	1
1.4	1.5
1.5	1
1.6	1
1.7	1.5
2.1	2
2.2	0.5
Case Study 2	
1	0.25
2.1	0.5
2.2	0.75
3.1	0.75
3.2	0.25
3.3	1
3.4.1	0.5
3.4.2	0.75
3.5	0.5
3.6	2
Part 3	
3.1	4
3.2	1
Part 4	
4.1	2
4.2	1.5
4.3	2
4.4	2

Part 3: Mini-Batch k-Means Clustering ("By Hand")

Basic Parameters

We would like to cluster the following full dataset below using Mini-Batch k-Means:

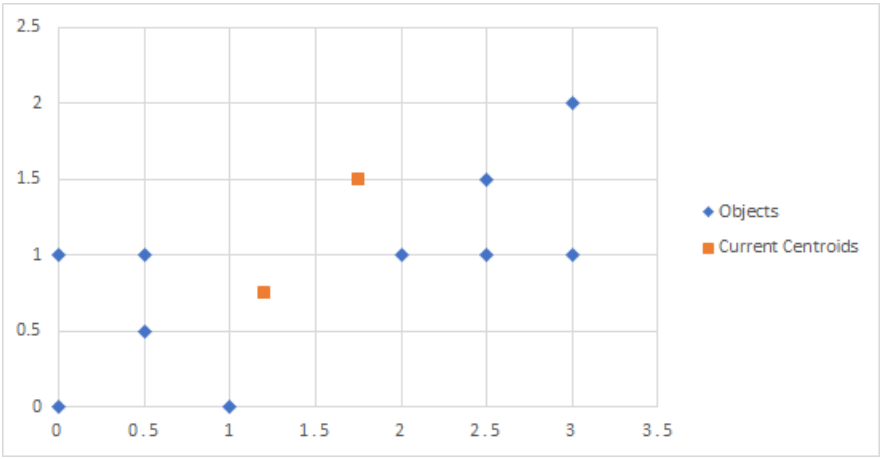
- into k=2 clusters,
- using random batches of size 3.

Centroid Reassignment Subroutine

We would also like to use the version of Mini-Batch k-Means that uses the centroid reassignment subroutine.

- We are using a reassignment ratio of 1/3.
- If this subroutine is activated for a given centroid, then you should "randomly" re-assign that centroid to one of the objects in the current batch and reset the centroid counter $v[\boldsymbol{c_k}]$ for that centroid to be 0.
- *For the purpose of this assignment, if you have to "randomly" re-assign a centroid, then you should assign it to first object in the current random batch (ie. the object with the lowest index).*

	Dataset	
	x	y
Object 1	0	0
Object 2	0	1
Object 3	0.5	0.5
Object 4	0.5	1
Object 5	1	0
Object 6	2	1
Object 7	2.5	1
Object 8	2.5	1.5
Object 9	3	1
Object 10	3	2



We have already run several iterations (ie. random batches) of the algorithm. The current values of the following objects in the algorithm are given below.

a) **Centroids:**
   - $\boldsymbol{c_1} = (1.2, 0.75)$
   - $\boldsymbol{c_2} = (1.75, 1.5)$

b) **Centroid Counters:**
   - $v[\boldsymbol{c_1}] = 70$
   - $v[\boldsymbol{c_2}] = 20$

#3.1. Suppose we have now just begun the *last* iteration of the algorithm. Our final random batch is comprised of objects 1, 2, and 10. Complete the rest of this iteration and give:

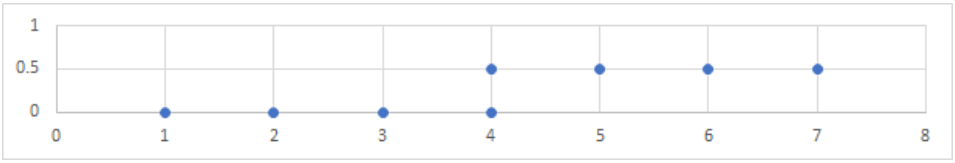
a) the final location of the two centroids $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ and
b) the final centroid counter values $v[\boldsymbol{c_1}]$ and $v[\boldsymbol{c_2}]$**.**

#3.2. Now that we have completed the final iteration of the Mini-Batch k-Means algorithm, we would like to read in ALL of the objects from the dataset one final time and give each of these objects a final cluster label using the method that we described in class. Give the final cluster labels for each of the 10 objects below.

Part 4: Spectral Clustering ("By Hand")

We would like to cluster the following dataset below using spectral clustering methods.

	Dataset	
	x	y
Object 1	1	0
Object 2	2	0
Object 3	3	0
Object 4	4	0
Object 5	4	0.5
Object 6	5	0.5
Object 7	6	0.5
Object 8	7	0.5



First, we would like to cluster this dataset, by using an affinity matrix generated using the 0/1 KNN approach.

#4.1. First, create an affinity matrix from this numerical dataset using the 0/1 KNN approach with k=3. After you have created this affinity matrix, force it to be symmetric using the method that we described in class. This *symmetric* affinity matrix will be the one that we use for the rest of part 2.1.

#4.2. Represent this symmetric affinity matrix from 2.1.1 as a weighted complete graph, where only the edges with non-zero edge weights are visualized. Next to each of the non-zero edge weights write down the corresponding edge weight.

#4.3. Given the graph in 4.2, out of all possible bi-partitions of the 8 objects that you could make (ie. all possible clusterings with 2 clusters), what bi-partition will create the minimum normalized cut? What is this minimum normalized cut value?

#4.4. Calculate the normalized Laplacian of the affinity matrix from 4.2.