CSE 575 Project 2 Report
Aryan Kumar
ASU ID: 1229512728
2/12/2023

Part 1:

- k values: 3, 5
- Initial Points:
    - k = 3:
      [[3.53350737, 0.33198894], [4.95728696, 6.90897984] ,
      [5.14167285, 5.71626939]]

    - k = 5:
      [[5.07631894, 3.30296197], [1.87131855, 3.43365823]
      [5.02471033, 8.23879873], [8.46078528, 2.85204573]
      [7.68097556, 0.83542043]]

- Final clustering centroid:
    - k = 3:
      [[3.23489005, 2.5530322], [4.83091958, 7.29959959]
      [7.23975119, 2.48208269]]

    - k = 5:
      [[3.07812639, 7.08662076], [2.68198633, 2.09461587]
      [7.1655183 , 8.39162215], [5.33907212, 4.46551175]
      [7.55616782, 2.23516796]]
- Loss:
    - k = 3: 1338.1076016520994
    - k = 5: 592.9375729660762

Part 2:

- k values: 4, 6
- Initial Point:
    - k = 4:
      [2.37650624, 8.15241778]

    - k = 6:
      [2.38952606, 7.22195564]

- Final clustering centroid:

k = 4:
[[3.34264769, 6.92602803], [2.85235149, 2.28186483]
[7.17928621, 8.0520791], [6.78374609, 2.85019999]]

k = 6:
[[2.60123296, 6.91610506], [3.21257461, 2.49658087]
[7.75648325, 8.55668928], [7.25262683, 2.40015826]
[5.40252508, 6.73636175], [nan nan]]

- Loss:
  - k = 4: 805.116645747261
  - k = 6: 613.2824392056043

Analysis

By the results I was more successful at implementing strategy 1. For strategy 2 my implementation of the algorithm was still effective for k=4 but failed at k=6 and got NaN values which then caused errors in the loss calculator as well. Based on the cost results it looks like strategy 2 is the better strategy as the cost values were lower for strategy 2 than strategy 1. Just like project 1 it was helpful for my understanding to get past the theoretical knowledge of the algorithms and actually implement them. Along with that I learned about the matplotlib library that has pyplot which can visualize graphs. For my person knowledge I visualized the clusters for all of my calculations and as an example below is the cluster for strategy 1, k=3: