HW 2 REPORT – Wang Zijia 1002885

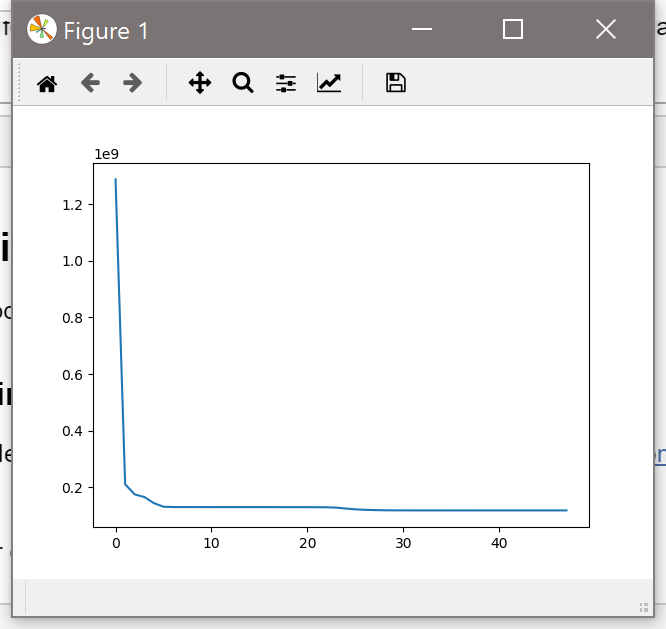
1. K-means
2. There are 6 clusters in the end.
3. The final centroids of each cluster:

[(241.21841411478403, 238.61387142567432, 233.8511458122085), (194.41149364755447, 136.3287472845764, 90.93252583766704), (136.2655562958028, 61.08973065575331, 10.103854573141716), (157.28881438861958, 97.59135595523944, 51.43351606034522),

(78.9274371361328, 37.10828688219993, 13.070704818530906), (25.978002322570312, 23.235754234686386, 23.605990629630124)]

1. The number of pixels associated to each cluster respectively:

[4930, 15190, 52535, 22075, 40365, 74917]



1. Visualize the result by replacing each pixel with the centroid to which it is closet and displaying the resulting image.

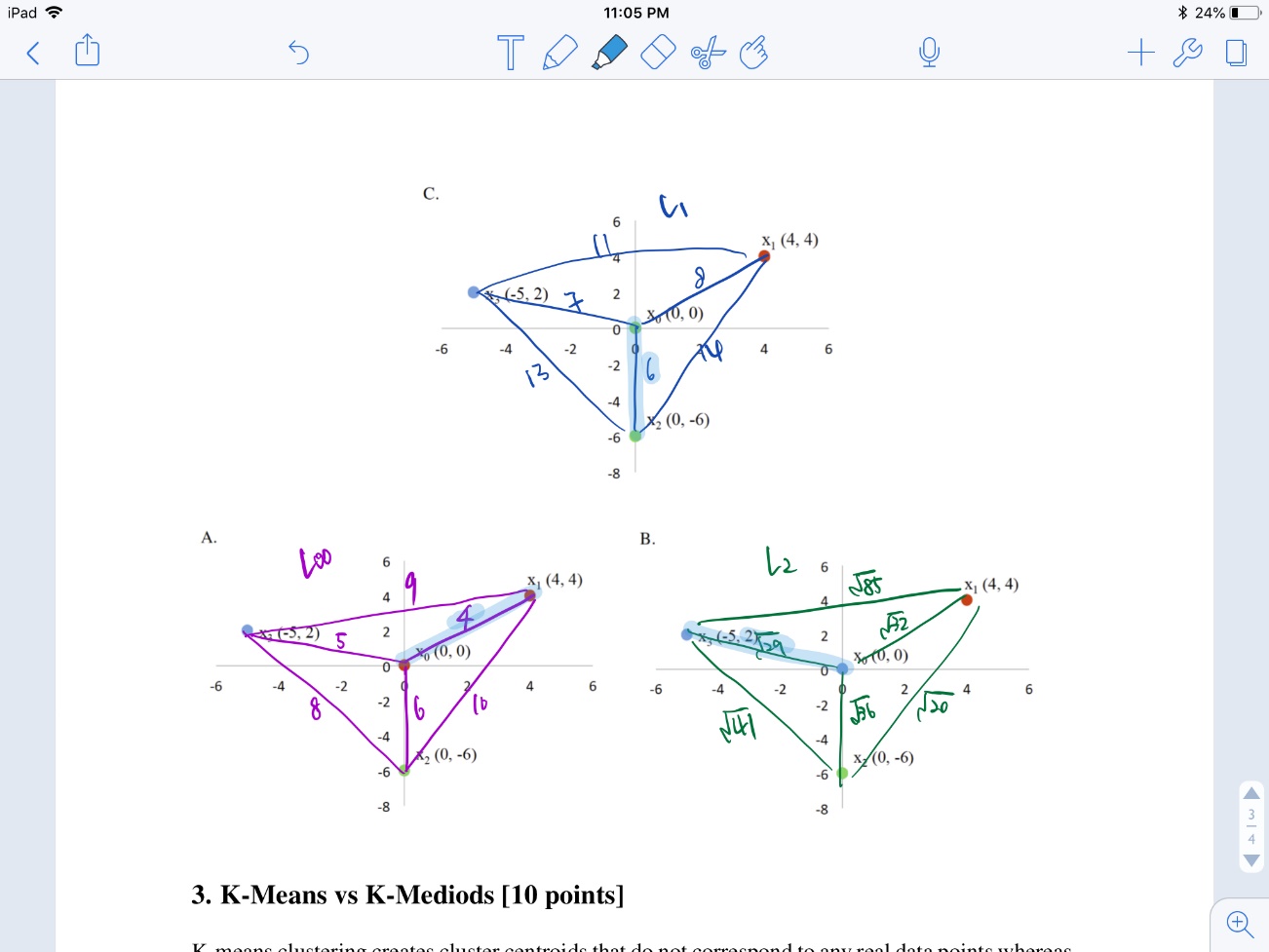


1. K-Medoids

Figure A uses the L∞ distance.

Figure B uses the Euclidean distance.

Figure C uses the Manhattan distance.



1. K-Means vs K-Medoids

* Advantages of K-Medoids:
  + Less sensitive to outliers and noise
  + A centroid is located inside a cluster as a representative point of the cluster
  + Applicable to clustering the news topic
* Disadvantages of K-Medoids:
  + Not as simple as K-means to implement
  + Higher running time (time complexity)
  + Not efficient for large data sets