Nghia Lam

1001699317

Task 1

Training and testing on the pendigits dataset, with k=1.

- classification accuracy=0.9743

Training and testing on the pendigits dataset, with k=3.

- classification accuracy=0.9750

Training and testing on the pendigits dataset, with k=5.

- classification accuracy=0.9763

Task 3

Sence the clustering is based on Euclidean distance between two dots these sets of clusters could not be the result of k-means algorithm as if it was based on Euclidean distance then the result would be separated horizontally. Thus, each cluster would be half and half based on the way each point is shown in the picture, and because it is not this means that this clustering cannot be the final result of the k-means algorithm.

Task 4

a) EM or Expectation-Maximization can give different results based on where it starts and how the data is categorized. EM can produce different results when running multiple times with the same K as it is sensitive to its initial positions and conditions. Depending on where it start could create different cluster variation with in the map.

b) If we have clusters where the distances between all points are the same, then you could have different results as to when and where it will combine the clusters together. However, because this clustering never needs to break a tie in this case it is deterministic. The answer is yes.

Task 5

a)

2	4	7	11	16	22	29	37
2,4	7	11	16	22	29	37	
2,4,7	11	16	22	29	37		
2,4,7,11		16	22	29	37		
2,47,11,16		22	29	37			
2,47,11,16,22		29	37				

2,47,11,16,22,29 37

2,47,11,16,22,29,37

b) 2 4 7 11 16 22 29 37

2,4 7 11 16 22 29 37

2,4 7,11 16 22 29 37

2,4 7,11 16,22 29 37

2,4 7,11 16,22 29,37

2,4,7,11 16,22 29,37

2,4,7,11,16,22 29,37

2,4,7,11,16,22,29,37