Nabin Chapagain
1001551151

Task 2

No, given clustering can not be the final result of K-means algorithm if we are using Euclidean Distance in our algorithm.

If we start from the position where given image is at, the mean of all the blue dots would be somewhere in the center of the "square" made of blue dots and the center of red dot(s) would be right where the original dot is at. Now, computing euclidean distance for each points, mean of red cluster (which is red dot itself in the first iteration) would be much closer to upper half of the "square" while mean of blue cluster would be closer to the lower half so, upper half would turn into red cluster and lower half would remain blue cluster. Now, the same process goes again until it doesn't repeat anymore but the main thing is, No. above given clustering cannot be final result of K-means algorithm.

Task 3

While using EM algorithm, we fit a mixture of Gaussians to the same data. Since EM algorithm is a greedy algorithm, the result greatly depends on the initialization values and hence even with same K, we might not get the same result all the time.
Perfect example to illustrate this is given in the lecture where for two gaussians with N1: mean=4 and stdev =2 and N2: mean = 7 and stdev=1 gives different mixture when weights change from w1=0.5, w2 =0.5 to w1 = 0.3, w2 =0.7. It shows different initialization factor changes the results.

For Agglomerative Clustering algorithm, for a specific dataset, d_{min} is always going to be same for the clusters and no matter how many times we rerun the algorithm, we are going to end up with same result.
This also can be perfectly seen from the example in the slides. No matter how many times we run the code, d_{min} between A, B, C, D, E, F and G is going to be same and the path taken by the algorithm is going to remain same throughout the process. This shows, if we use same distance (d_{min}), agglomerative clustering will give us same results.

Task 4

for 2, 4, 7, 11, 16, 22, 29, 37; difference between adjacent numbers is (4-2) 2, (7-4) 3, (11-7) 4, (16-11) 5, (22-16) 6, (29-22) 7, (37-29) 8.

Part a) Using d_{min} :

2 4 7 11 16 22 29 37 (each of them are individual clusters)

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

Nabin Chapagain
1001551151

2 4 7 11 16 22 29 37

2 4 7 11 16 22 29 37

2 4 7 11 16 22 29 37

Part b) Using d_{max}

2 4 7 11 16 22 29 37 (each of them are individual clusters)

2 37 4 7 11 16 22 29

2 37 29 4 7 11 16 22

2 37 29 4 7 22 11 16

2 37 29 4 7 22 11 16

2 37 11 16 29 4 7 22

2 37 11 16 29 4 7 22

2 37 11 16 29 4 7 22