Alfredo Reina Corona

FINAL GRADE: 95%

COMMENTS:

“task 5: part b is wrong -5 points ”

Task 1:

Task 3:

k=1 ; classification accuracy= 0.9743

k=3 ; classification accuracy= 0.9751

k=5 ; classification accuracy= 0.9760

No. If the loop is allowed to run for even 1 more iteration, the means of the blue cluster will be near the center, while the mean of the red cluster will be at the singular red point. When recomputing which point is closest to each cluster, the points directly adjacent to the red dot will also be re-classified and changed to red. This cycle will continue until a the blue dots are too far from the red median to change.

Task 4:

a) No. If the clusters are assigned randomly then there is some degree of variability. The solution produced is the local optimum, so if a large enough set of points is presented, it may cause the final result to be different even if the points stay exactly the same.

b) Yes. Assuming a tiebreaker is never needed and point locations never change, each cluster will start as a single point. There points will gradually merge and since there is no variability within the mergers the process will stay the the same, step by step, from start to finish.

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, 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]

b) [2], [4], [7], [11], [16], [22], [29], [37]

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

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

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

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

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

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

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