

ML-6

siddharth bhagvagar

November 2024

1 Task-1

when $k = 1$ (pendigits)

`classification accuracy=0.9743`

when $k = 3$ (pendigits)

`classification accuracy=0.9750`

when $k = 5$ (pendigits)

`classification accuracy=0.9763`

2 Task-2

please refer `k_means.py`, no instructions were given for the output.

3 Task-3

For the image, this cannot be the final version of the k-means algorithm as the blue adjacent to red is closer to the red one than the mean(centroid) of the blues, so the distance between the red and the blue ones adjacent to red will be less. Hence, the algorithm will continue.

4 Task-4

4.1 Part-a

No, EM can produce different results when run multiple times with the same K because it starts with random initialization of parameters, which leads to the variation of results.

for example (1,4,6,9) and $k=2$

first run: - cluster 1 \rightarrow 2 and cluster \rightarrow 8 which results in (1,4) and (6,9)

second run: - cluster 1 \rightarrow 3 and cluster \rightarrow 5 which results in (1) and (4,6,9)

4.2 Part-b

Yes, agglomerative clustering with d_{min} will always produce the same results when applied to the same dataset, assuming no ties need to be broken because there is no random implementation part. The initial state is fixed and each step has an outcome until all become one cluster, leading to the same result every time.

5 Task-5

5.1 Part-a

Using d_{min} :

{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}

5.2 Part-b

Using d_{max} :

{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}