Project report

zid: z5192421

Part1:

Firstly I divide the input datafile into p parts . Then calculate the distance from each point to each center. (here, we need to use L1 distance to directly calculate the difference value rather than square root of the square deviation). Then find out the points contained in each center, and then calculate the new center points according to these points using K-means(the median is needed here instead of the average). Finally loop max_iter times. After get final centroids, calculate a Code File according to these new centroids as the output.

Part2:

Divide the queries into p parts, and then compare the separated queries with codebooks to find the distance from each query to each center with a qc_dist(p,k). Instead of using multi_index(failed to pass attests using multi_index), I calculated the distance from each points to the query according to Code and qc_dist. Q_dist (N, 1) and sorted it. I selected the point with the largest value and repeated it T times to get the largest value of the first N points (N>=T).