

Tao Lipeng z5048267

Part 1 implementation:

I divided each piece of data into P chunks, and clustered each piece of data using scipy's l1 distance formula's kmeans algorithm. Unlike traditional kmeans, the distance is calculated using l1 distance instead of l2 distance. When using points, use the median of all points instead of the average of all points, and then store the center point of each cluster as a codebook, and store the cluster label to which each data belongs as codes.

Part 2 implementation

Firstly, I divide each piece of data into P blocks and calculate the distance of the center point of each cluster from the data calculation of each block and the pq algorithm. And then I take the sum of the distances of all the corresponding blocks and rearrange from small to large. And then take the final result from the point in the cluster with the smallest distance until the result is full of T .