In this experiment, we aimed to investigate the performance of the CURE algorithm for refining the clusters obtained through agglomerative clustering. The experiment was performed on the MNIST dataset, which consists of images of handwritten digits from 0 to 9. The dataset was preprocessed by normalizing the pixel intensities to a range of [0, 1].

We first defined a list of cluster sizes to try, which included 10, 100, 200, and 500. For each cluster size, we ran the algorithm 10 times to obtain a more reliable estimate of the performance.

In each iteration of the algorithm, we first performed agglomerative clustering to obtain the initial clusters. We then applied the CURE algorithm to refine the clusters. In the CURE algorithm, we identified the representative point for each cluster by computing the centroid of the cluster and selecting the point that was farthest from the centroid. We then removed this point from the data and re-applied agglomerative clustering to obtain the final set of clusters. We calculated the entropy for each cluster and the total entropy for all clusters.

Our results showed that the total entropy decreased as the cluster size increased.

We also observed that the number of runs had a significant impact on the results. In some cases, the total entropy varied widely between different runs. This suggests that the performance of the algorithm is highly dependent on the initial randomization of the data.

In conclusion, our results indicate that the CURE algorithm is an effective method for refining clusters obtained through agglomerative clustering, and that larger cluster sizes lead to more refined clusters. However, the results are sensitive to the number of runs and the initial randomization of the data, suggesting that caution should be exercised in interpreting the results.