CoreSet for Minimum Enclosing Ball

Anmol Anand
Department of Computer Science
Texas A&M University
aanand@tamu.edu

1 Introduction

A core-set for the Minimum Enclosing Ball (MEB) problem is a reduced set of points that retains the essential information needed to find the smallest ball containing all the original data points. It significantly reduces computational complexity, allowing for faster solutions on large datasets while maintaining accuracy.

Given a set of points in 2-dimensional space (offline or in a streaming setting), and a constant $\epsilon > 0$, the experiments in this project aim to find the $(1+\epsilon)$-coreset of the given set of points. In these experiments, we draw the following observations:

- First, visualize the original data points and the computed coreset data points in the 2-dimensional space for different values of ϵ.
- Second, how the accuracy of finding MEB varies with the size of the coreset and consequently the value of ϵ.

2 Code

The code can be found in this Github repo github.com/anmol-anand/coreset-for-meb. The instructions to run the code are mentioned in the README.

Note: The experiment is run for an offline set of points and an online streaming list of points. For each of these experiments, 500 data-points are randomly sampled in 2-dimensional space, and $(1+\epsilon)$-coreset is computed for these data-points for several values of ϵ. Accuracy is compared for different values of ϵ.

3 Sample Generation and Error Computation

For both the online and offline setting, 500 random samples are generated, and the $(1+\epsilon)$-coreset is computed for several values of ϵ. To compare the error in solving the MEB using the original set vs the coreset, we compare the length of the diameters of the two sets. (Diameter being the largest distance between a pair of points in a set):

$$\text{Error } = \left(\frac{diameter(\text{original set})}{2} - \frac{diameter((1+\epsilon)\text{-coreset})}{2} \right)^2 \tag{1}$$

4 Analysing the results

For both, the online and offline settings, we compute coresets for various values of epsilon and observe that the points in the coresets roughly lie on the boundary of the convex hull of the original set of points, as observed in Figure 1. where the points in blue are in the original set and the highlighted points in red are also in the coreset.

Furthermore, we observe that as the value of epsilon increases, the size of the coreset decreases, and the error in finding the Minimum Enclosing Ball (Eqn. 1) increases. This trend holds true for both the offline setting (Figure 2) and the online setting (Figure 3).

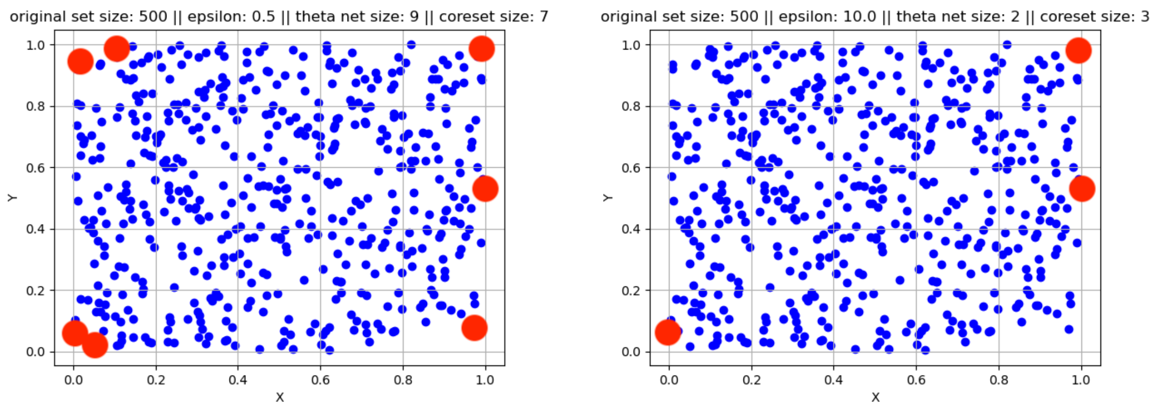

Figure 1: Offline coreset computation on an original set of size = 500

(a) epsilon = 0.5 gives a coreset of size = 7 —— (b) epsilon= 10.0 gives a coreset of size = 3

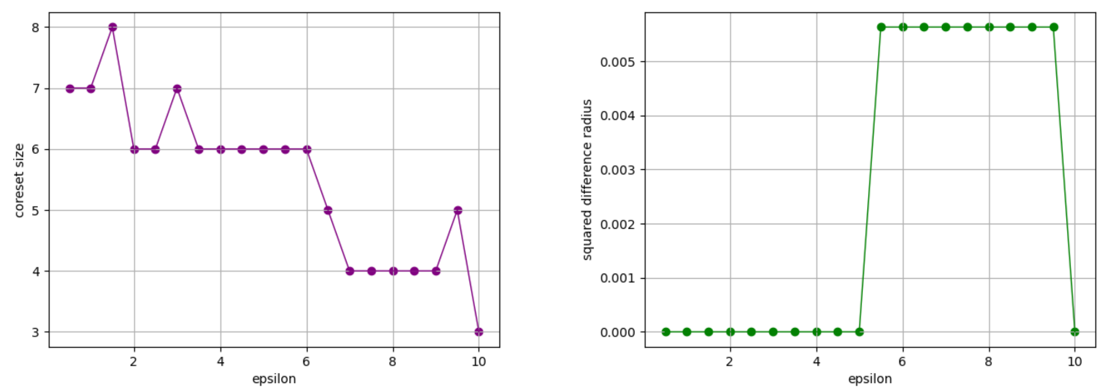

Figure 2: For the offline setting, we can see that as the value of epsilon increases, the coreset size decreases, and the error increases.

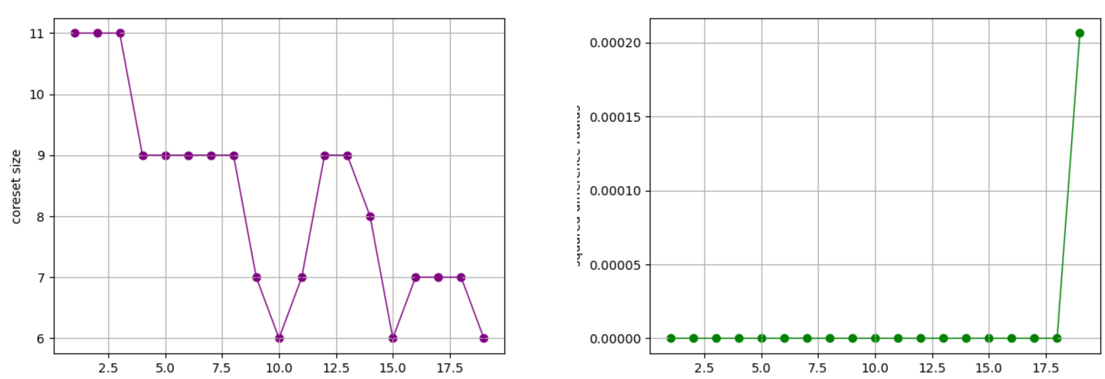

Figure 3: For the online setting, we can see that as the value of epsilon increases, the coreset size decreases, and the error increases.