Q6 A] (ii)

CURE Algorithm

- CURE stands for Clustering using Representative Algorithm
- It is an efficient data clustering algorithm for large databases
- CURE algorithm works better in spherical as well as non-spherical clusters.
- CURE uses random sampling and partitioning to speed up clustering.

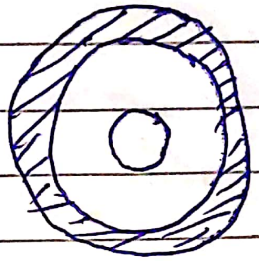
Data
↓
Make Random Sample
↓
Make Partitioning of Sample
↓
Partially Cluster Partitions
↓
Eliminate Outliers
↓
Cluster Partial Clusters
↓
Label data in Disk

CURE OVERVIEW

AMEY THAKUR BDA 13-10-2021 (12)

BE B 50 IAT-II Amey

- The CURE algorithm is divided into two phases.
  1. Initialization in CURE
  2. Completion of the CURE algorithm.



Two clusters, one surrounding the other

- Figure illustrates two clusters.
- The inner cluster is an ordinary circle, while the second is a ring around the circle.
- The arrangement is not completely pathological

- It is a large scale clustering algorithm which uses a combination of partition based and hierarchical algorithms.
- This algorithm assumes Euclidian space and identifies cluster of any space.
- CURE starts by treating each data point as a single cluster and then recursively merges two existing clusters into one until we have only k clusters.
- CURE uses two data structures to compute minimum distance between representative points.
- Two stages of CURE implementation
  1. Initialization in CURE
  2. Completion of CURE Algorithm.

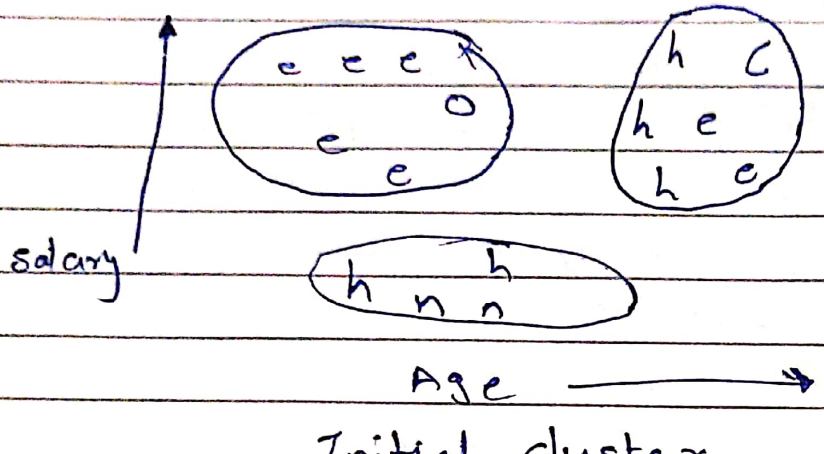
1. Initialization in CURE
   - Take a small sample data and cluster it in main memory using hierarchical clustering algorithm.
   - Select a small set of points from each cluster to be representative points. These points should be chosen to be as far from one another as possible.

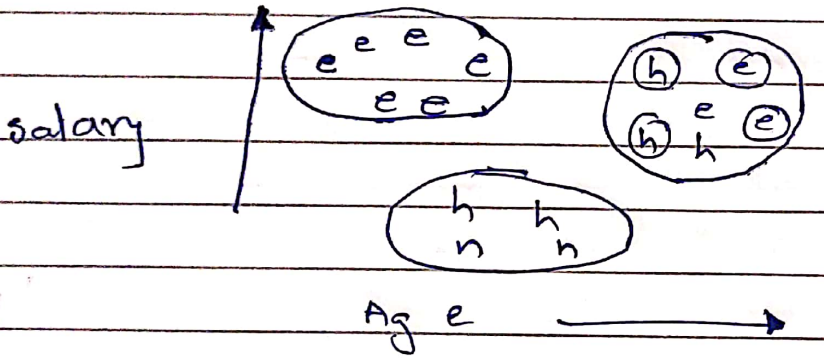
2. Completion of CURE algorithm
   - Once, initialization is complete, we have to cluster the remaining points and output the final clusters.

AMEY THAKUR BDA 13-10-2021 (4)

BE B 50 IAT-II Amey

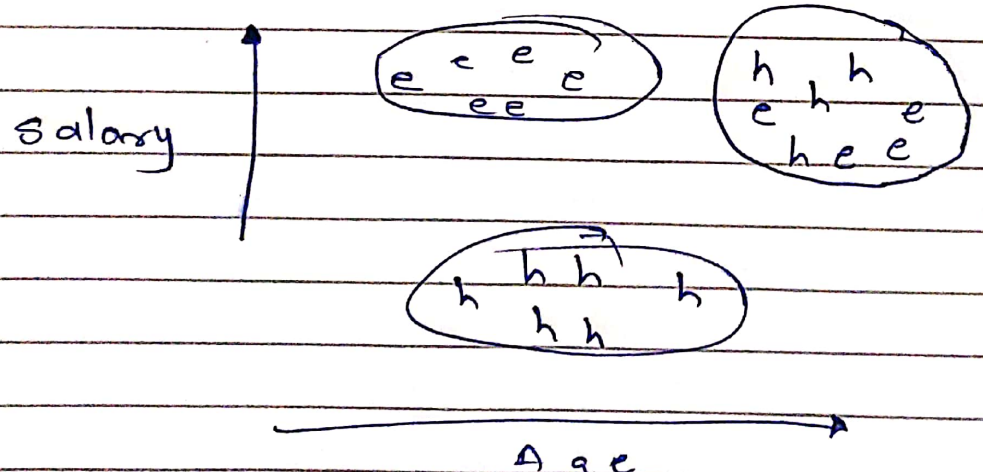
Figure illustrating CURE





Pick 4 remote Points for each cluster

Pick disperse point

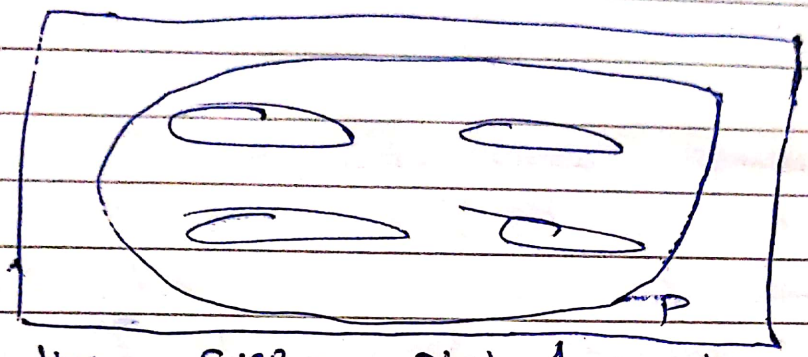


moves point 20 % toward centroid

shrink point 20 %

AMEY THAKUR BDA 13-10-2021 (5)

BE B 50 IAT - II Amey



completing CURE - pick 4 remote points for each cluster

AMEY THAKUR BDA 13-10-2021 (6)

BE B 50 IAT-II Amey

Q6 B] (ii)

Applications of Recommendation system

(1) Product Recommendations:

- Perhaps the most important use of recommendation system is at on-line retailers.
- We have noted how amazon or similar online vendors strive to present each returning user with some suggestions of products that they might like to buy.
- These suggestions are not random, but are based on the purchasing decisions made by similar customers or on other techniques.

(2) Movie Recommendation:

- Net Flix offers its customers recommendations of movies they might like.
- These recommendations are based on ratings provided by users.

AMEY THAKUR BD1 13-10-2021 (7)

BE B 50 TAT-II Amey

(3) News Articles:

- News services have attempted to identify articles of interest to readers, based on the articles that they have read in the past.
- The similarity might be based on the similarity of important words in the documents, or on the articles that are read by people with similar reading ~~taste~~.
- The same principles apply to recommending blogs and among the millions of blogs available, videos on YouTube, or other sites where content is provided regularly.