



Semester : VII

Subject : Big Data Analytics

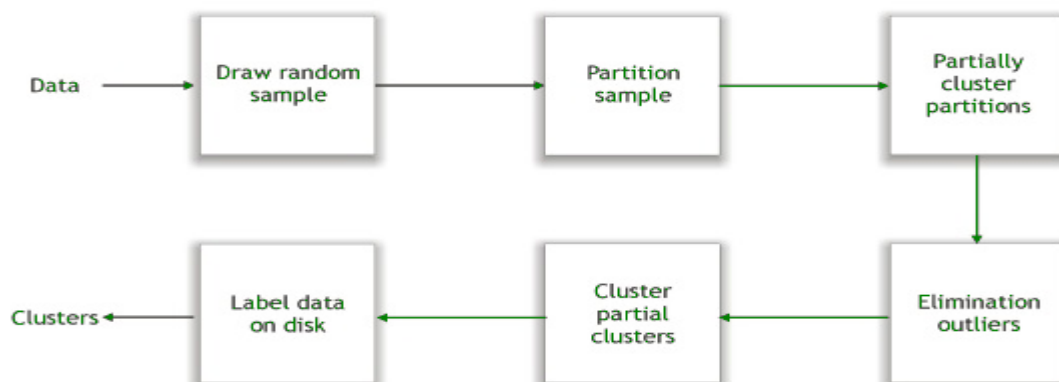
Academic Year: 2024 – 2025

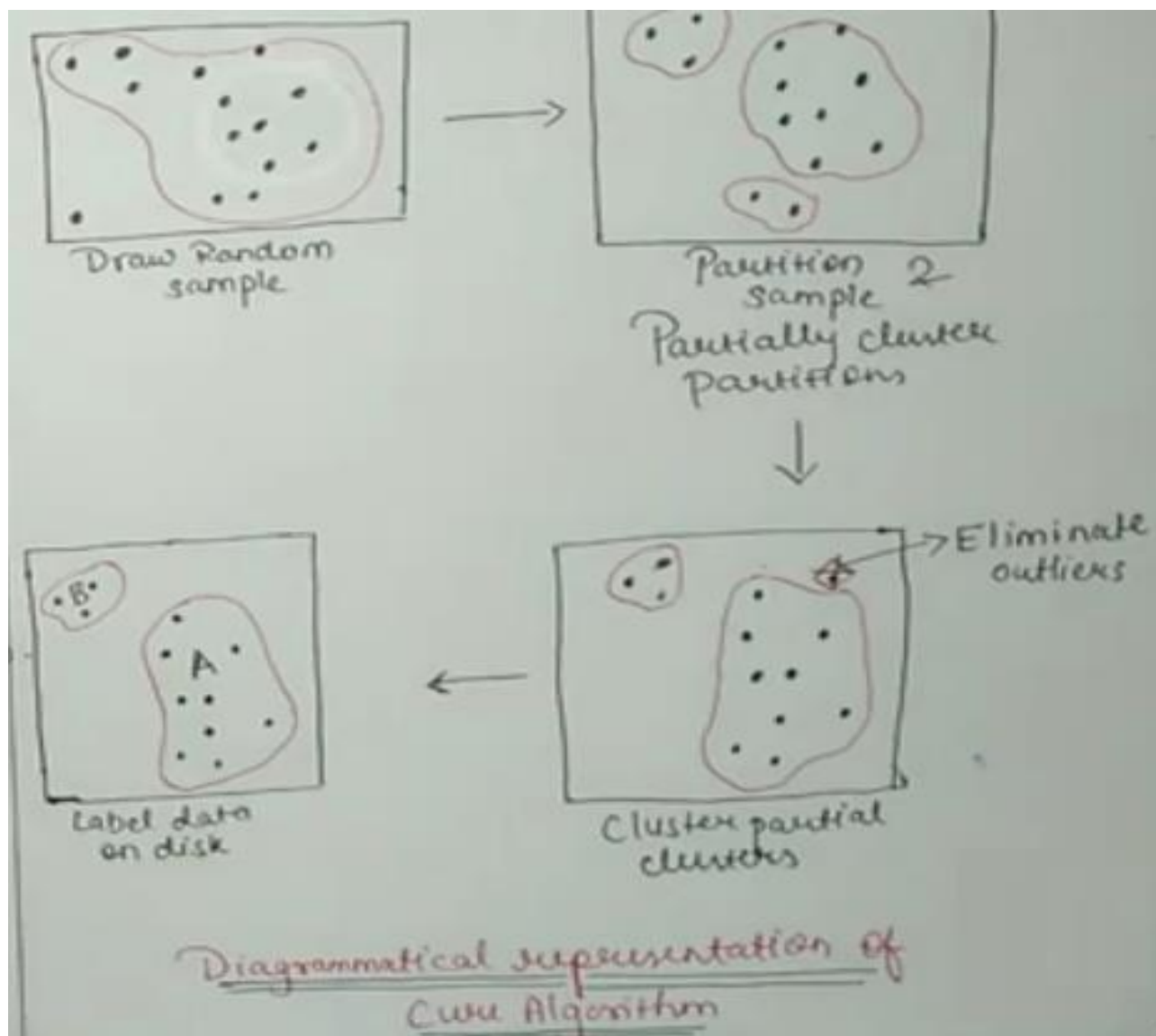
Module 5:

CURE Algorithm

- Clustering is useful for discovering groups & identifying interesting distributions in the underlying data.
- Traditional clustering algorithms either favour clusters with spherical shapes & similar sizes, or are very fragile in the presence of outliers.
- Thus, CURE Algorithm came into picture which is more robust to outliers, and identifies clusters having non-spherical shapes & wide variances in size.
- CURE stands for Clustering Using Representatives.

Architecture of CURE Algorithm





K Means Clustering Algorithm

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in data science. K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if $K=2$, there will be two clusters, and for $K=3$, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The k-means clustering algorithm mainly performs two tasks:



- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Example

Use K-Means Algorithm to Create 3 clusters for given Set of Values:
{22, 9, 12, 15, 10, 27, 35, 18, 36, 11}

Step 1:

Clusters	Objects	Centroids
1	22	22
2	9	9
3	12	12

11 - 9 = 2
12 - 11 = 1

Step 2:

Clusters	Objects	Centroids
1	22, 27, 35, 18, 36	27.6
2	9, 10	9.5
3	12, 15, 11	12.65

Step 3:

Clusters	Objects	Centroids
1	22, 27, 35, 36	30
2	9, 10, 11	10
3	12, 15, 18	15

Step 4:

Clusters	Objects	Centroids
1	27, 35, 36	32.6
2	9, 12, 10, 11	10.5
3	22, 15, 18	18.3

Step 5:

Clusters	Objects	Centroids
1	27, 35, 36	32.6
2	9, 12, 10, 11	10.5
3	22, 15, 18	18.3