

Advances in Machine Learning

Problem State : Implementation of various clustering :
K-Means, & Hierarchical

Input : Breast Cancer dataset (or any other dataset)

Output : Clustering of k-means & Hierarchical

Conclusion : Thus, we implemented different clustering algorithms

Frequently Asked Questions :

Question 1 : Different graph based and density clustering techniques?

* Density based clustering : It identifies groups/clustering in the data, based on the idea that a cluster in a data space is a contiguous region of high point density

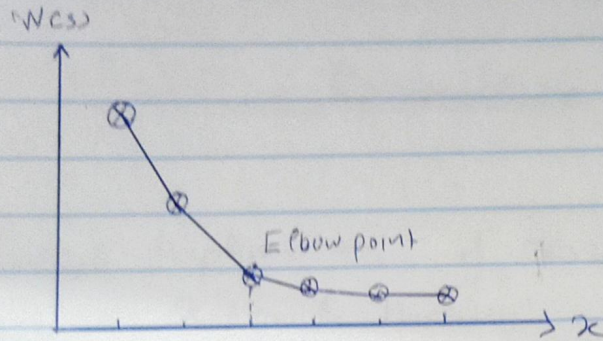
Example: K-means ; Hierarchical cluster

★ Graph based clustering: It uses a proximity graph (Cite Laplacian matrix and congruency matrix). Each edge between nodes has a weight which is the proximity between the two points.

Example: Spectral clustering

Question 2: Explain the method to decide no of clusters K in distance based methods?

⇒ Consider K means clustering. With every centroid, we can calculate the within cluster sum of squares or WCSS. As an increase in K , the WCSS value decreases. However, the rate of decrease also goes down and after a point, it is negligible. We can plot this across multiple K s to find the appropriate one.



The elbow point, or the point after which the decrease rate gets low is considered to be an appropriate number of clusters.

Question 3

Advantages of K-means :

- Guarantees convergence.
- Can even start the position of centroids
- Easily adapts to new examples
- Generalize to clusters of different shapes.