* Clustering in ML *

↳ Unsupervised Learning method.

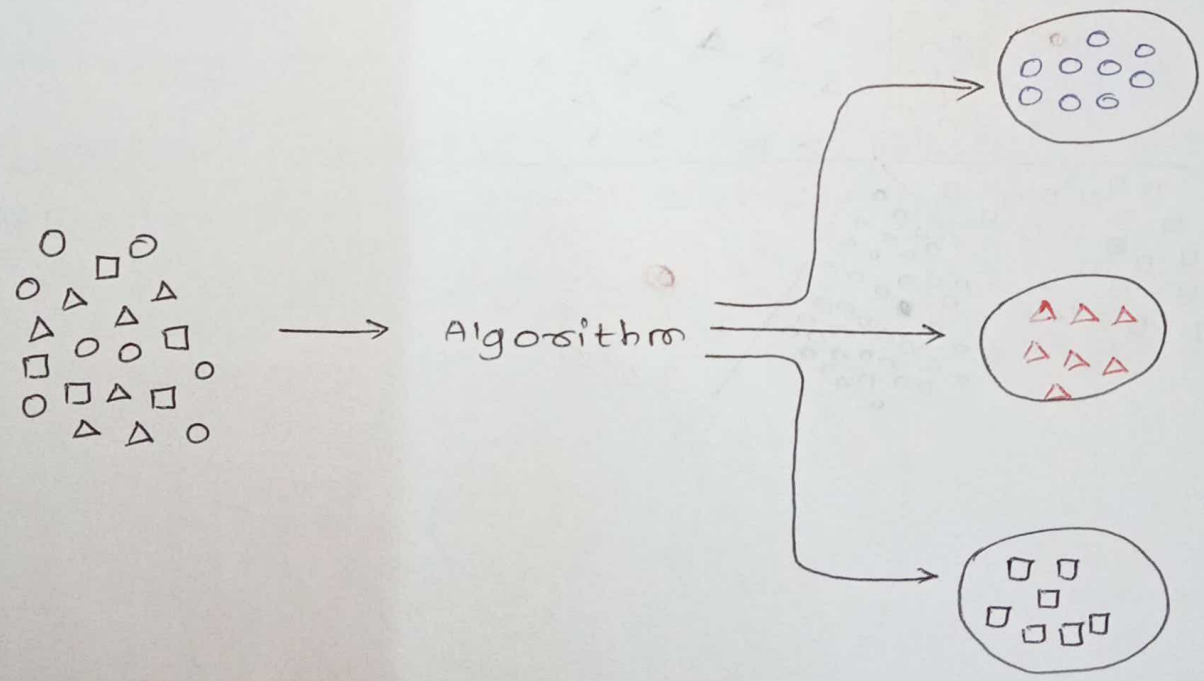
↳ No supervision & groups unlabelled data.

↳ Way of clustering data points into different clusters, consisting similar data points.
↳ Group objects with possible similarities.

↳ Each cluster/group divided with Cluster-ID. Use this id to process large & complex datasets.

↳ Commonly use for statistical data analysis.

↳ Clustering same like classification but,
Classification: Labelled dataset
Clustering: Unlabelled dataset.



⊙ Types of Clustering methods:-

Clustering broadly divided into:

- Hard Clustering - Datapoints belong only one group.
- Soft Clustering - Can belong to another group also.

1) Partitioning Clustering:

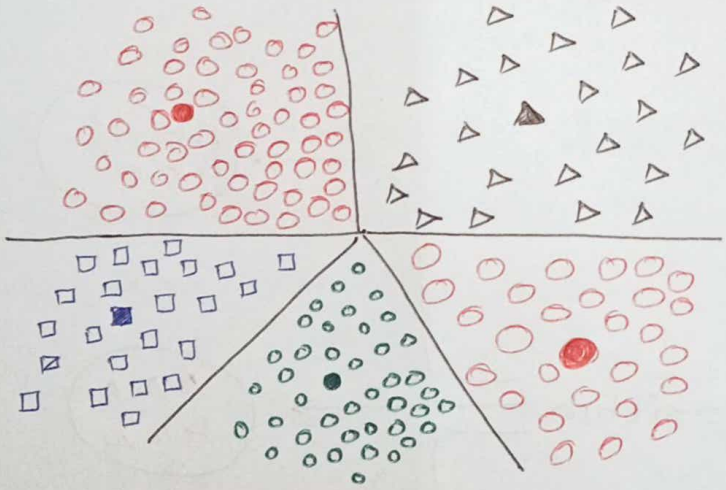
↳ Divide in non-hierarchical groups.

↳ Also k/a centroid-based method.

↳ Example - K-means clustering.

↳ Divide in set of k groups, k is num. of predefined groups.

↳ Cluster centroid - Min dist from all points.



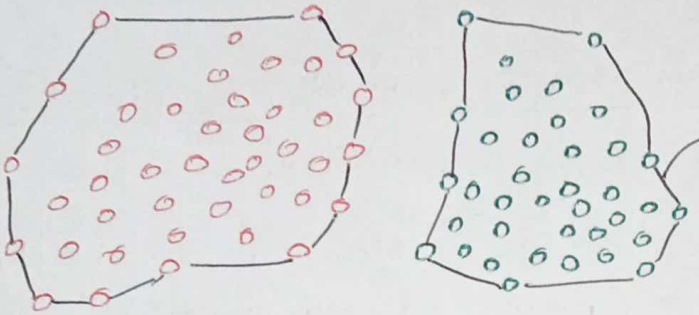
2) Density-Based Clustering —

↳ connect highly-dense areas into clusters & form shape till connect dense region.

↳ Find diff. clusters & connect highly-dense areas.

↳ Dense areas divided by sparser areas.

↳ complex for varying densities & high dimensions.



↳ Form arbitrarily shaped distributions

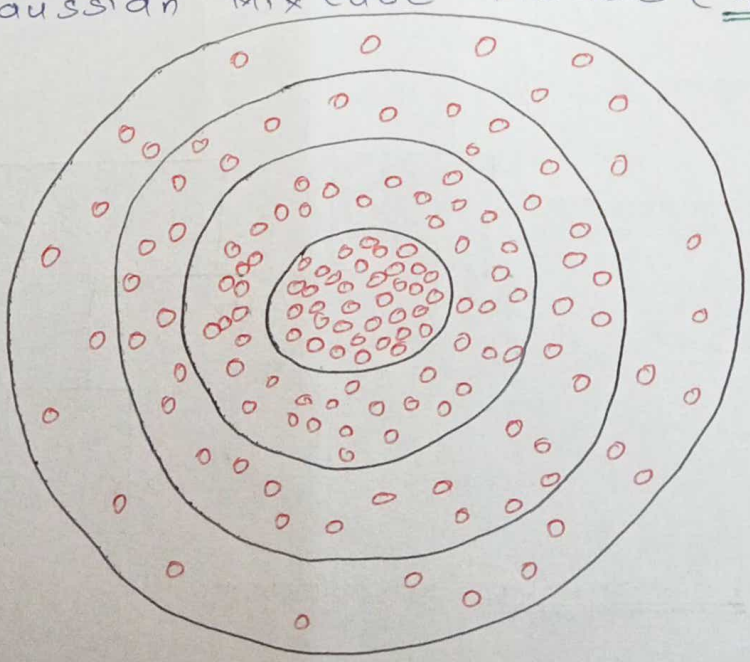
↳ Example - DBSCAN (Density-Based Spatial clustering of Applications with Noise).

3) Distribution Model-Based Clustering —

↳ Divide by probability of how dataset belong to particular distribution.

↳ Grouping done by Gaussian Distribution.

↳ Example - Expectation-Maximization clustering Algo uses Gaussian Mixture Models (GMM).



4) Fuzzy Clustering –

↳ Soft method, belong to more than 1 group/cluster.

↳ Dataset has membership coefficients, depend on degree of membership to be in cluster.

↳ Example – Fuzzy c-means Algo (Fuzzy k-means Algo).

5) Hierarchical Clustering —

↳ Form clusters in tree-shape based on hierarchy.

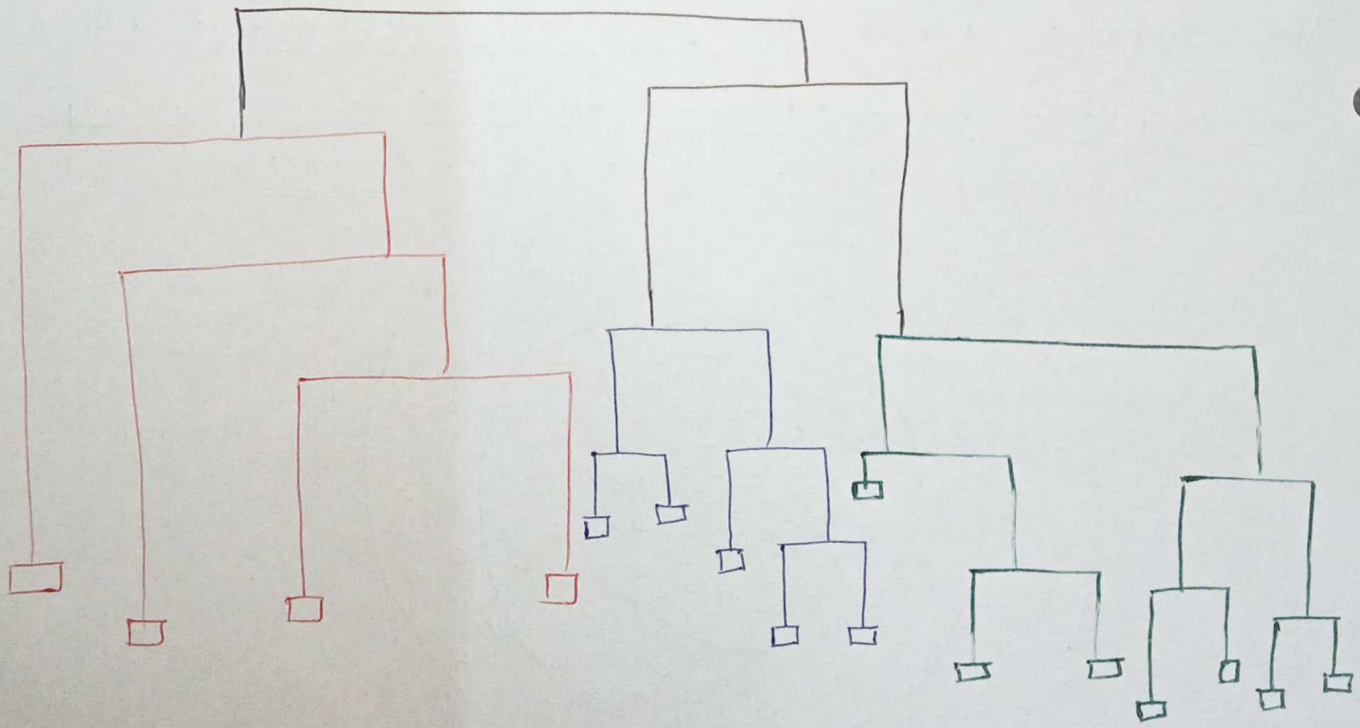
↳ New clusters formed using previous.

↳ Can use alternative for partitioning method, no required to define k (num. of clusters).

↳ Dataset divided into tree-like clusters, called dendrogram.

↳ Num of clusters selected by tree pruning.

↳ Example – Agglomerative Hierarchical Algo.



- Agglomerative: bottom-up approach
- Divisive: top-down approach