Hierarchical Clustering

Sunday, May 21, 2023 11:45 AM

(Unsupervised Machine learning)

a) Dimensionality Reduction (PCA) = Principle Component Analysis

b) K Means (k-Hyper)

c) Hierarchical clustering.

Hierarchical Clustering

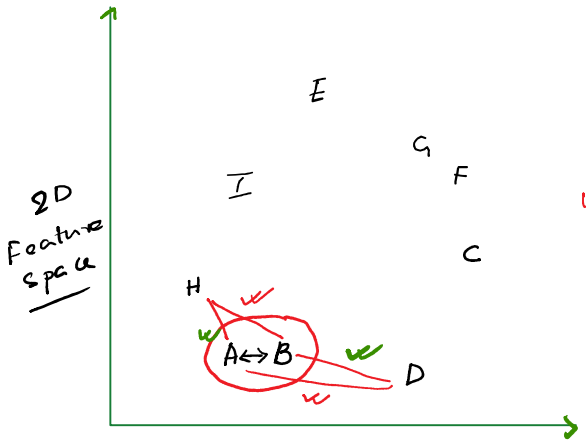
a) it does not need to know K in advance

b) Use similarity or distance Metric

c) Use Dendogram

d) Deterministic (Reproducible)

e) Greedy (Local Solution)



Linkage

a) Complete (max) — From H → Clust(AB), From D → Clust(AB) ~ A

b) Single (min) — From H → Clust(AB) ~ A

c) Avg

d) Centroid

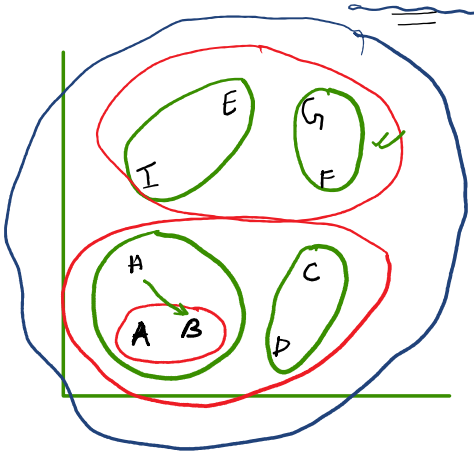
e) Ward

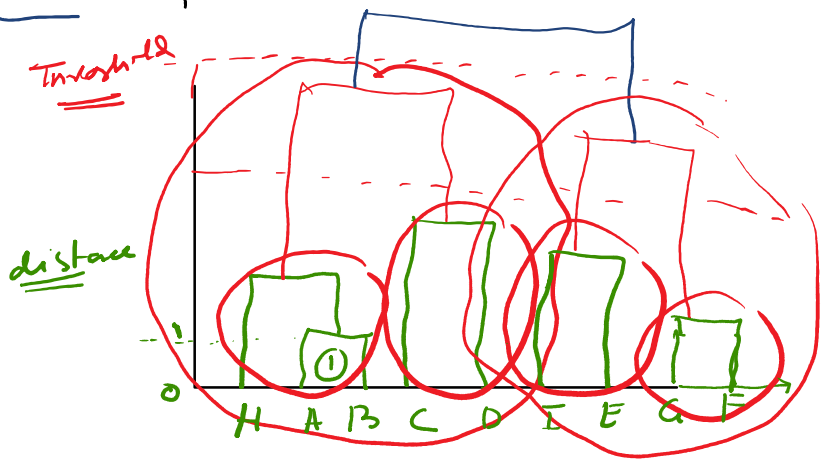
Some points who minimize the variance of the clusts

max min 2

A↔B

Draw complete linkage of Hirarchical Cluster.





Linkage selection dendrogram will look like

Based on linkage selection dendogram will look diffe

* Threshold → Hyperparameter

Conclusion

× Choice of similarity metric is very Imp

like Euclidian distance

↳ Should not use for "Sparse" Metric

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

dot product Better result

Imp :- Feature Scaling is very imp Standardisation