Foundations of Machine Learning AI2000 and AI5000

FoML-26
Hierarchical Clustering

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

So far in FoML

- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions
    b. Bias-Variance Decomposition
    c. Decision Theory - three broad classification strategies
    d. Neural Networks
- Unsupervised learning
    a. K-Means clustering

For today

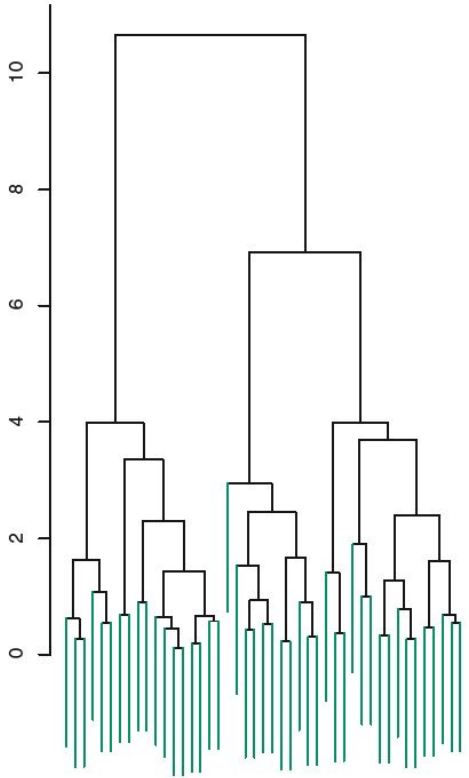
- Clustering
  - Hierarchical Clustering

Some of the contents are taken from - Intro to Statistical Learning

Hierarchical clustering

- Unlike K-Means, no need to specify the ‘K’
- Results in a tree-like representation of the data - Dendrogram

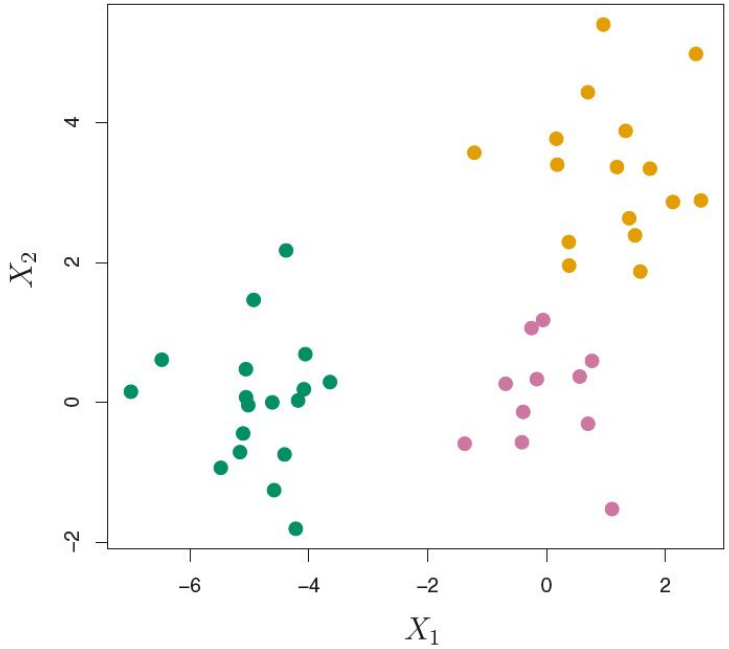
Hierarchical clustering



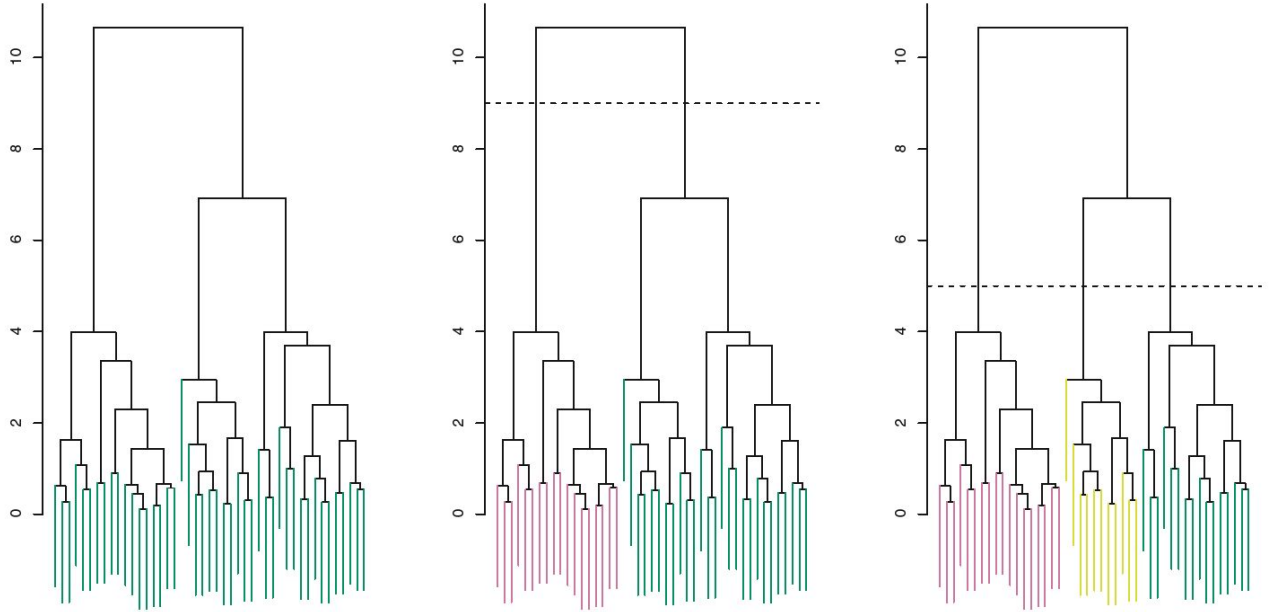
Bottom-up or Agglomerative clustering

- Most common of Hierarchical Clustering
- Builds the Dendrogram - starting from the leaves combining the clusters up to the trunk

Hierarchical clustering



Hierarchical clustering



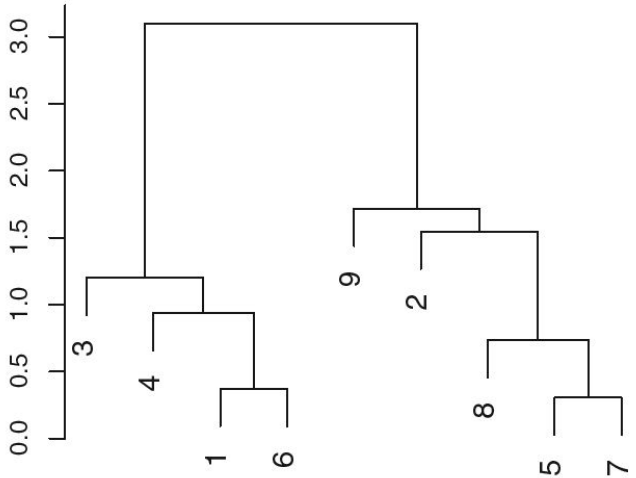
Hierarchical clustering

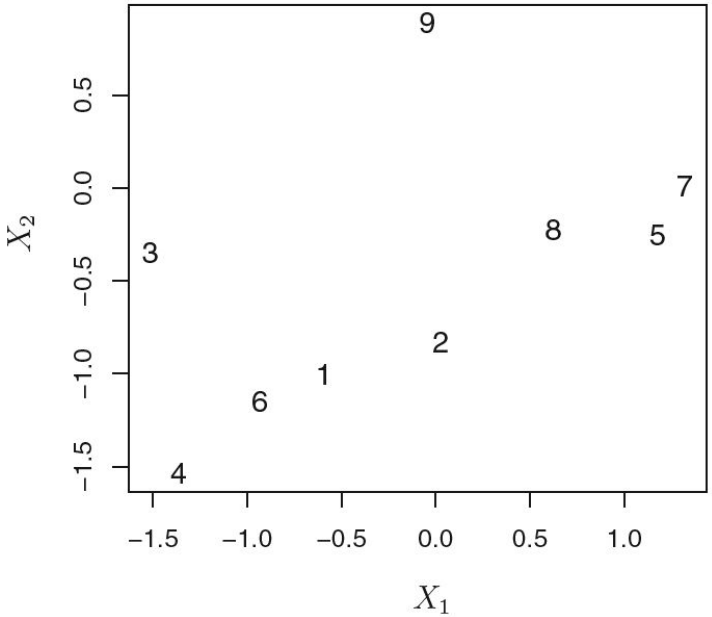
- Each leaf represents one of the observations
- As we move up, some leaves begin to fuse into branches
    - Observations that are similar/close to each other

Hierarchical clustering

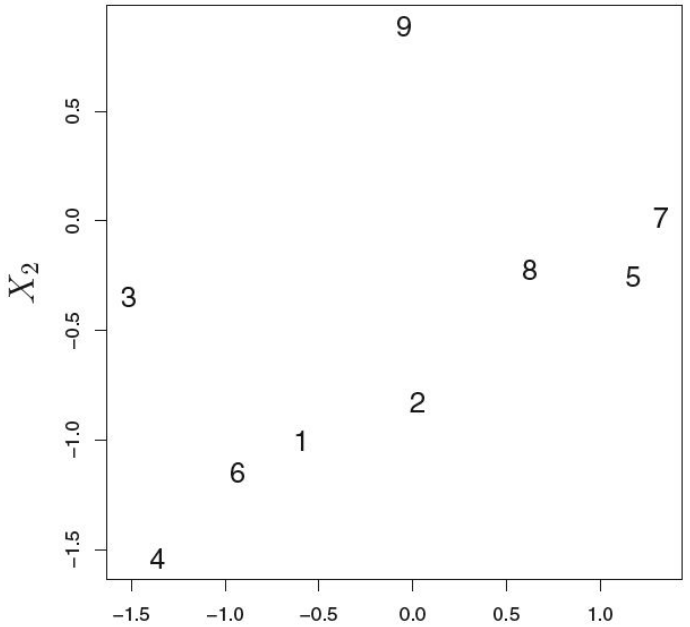
- The earlier the fusion occurs - the similar the groups are to each other
- Observations that fuse later are less similar

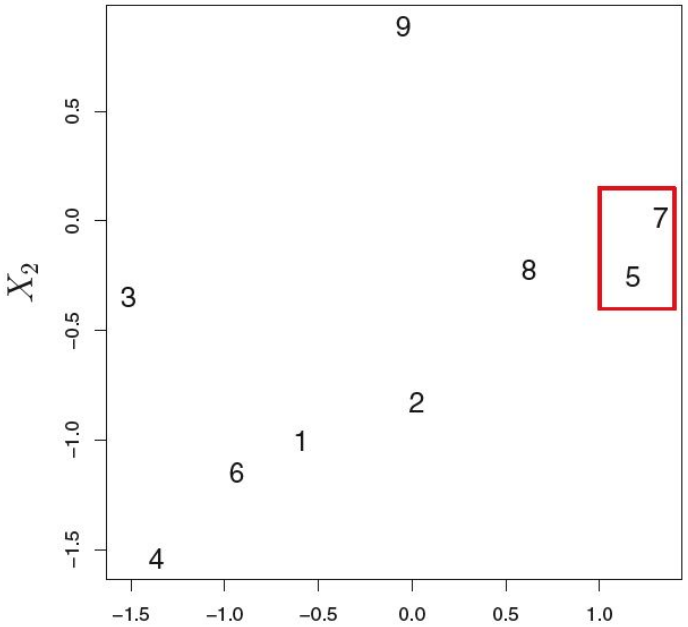
Hierarchical clustering



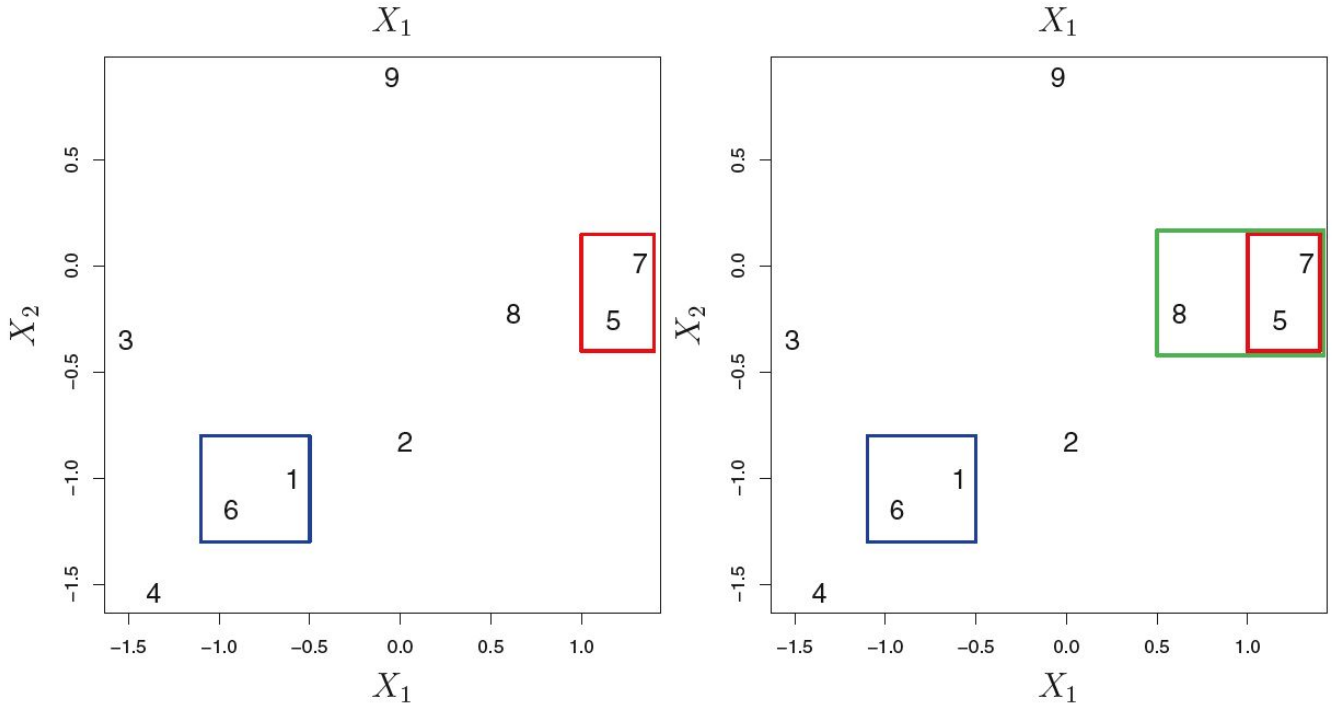


Hierarchical clustering





Hierarchical clustering



Hierarchical clustering Algorithm

1. Begin with n observations and a measure (such as Euclidean distance) of all the $\binom{n}{2} = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.

2. For $i = n, n-1, \ldots, 2$:

   (a) Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.

   (b) Compute the new pairwise inter-cluster dissimilarities among the $i-1$ remaining clusters.

Hierarchical clustering

- Concept of dissimilarity b/w a pair of groups of observations?
- Linkage
  - Complete - maximal similarity
  - Single - minimal
  - Average - average of pairwise similarities
  - Centroid - similarity of cluster centers

Next

- GMM for clustering

Rough work