7.03 Hierarchical Clustering

Hierarchical Clustering

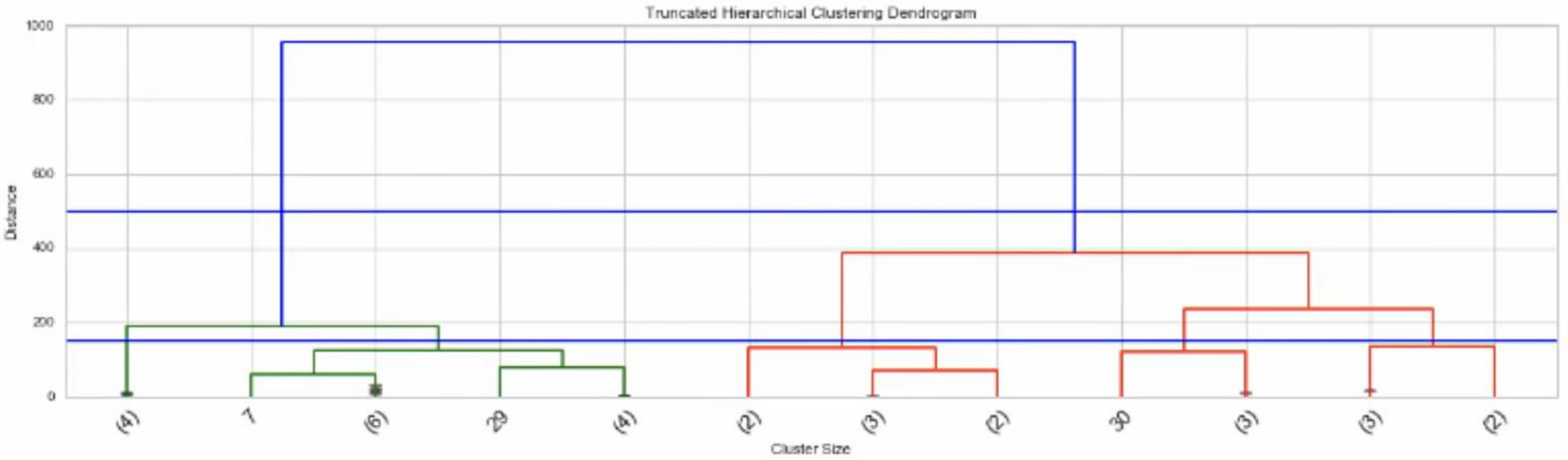
- It predicts subgroups within data by finding the distance between each data point and its nearest neighbours, and then linking the closest neighbours

- It uses the distance metric it calculates to predict subgroups

- Determine number of subgroups in a data set by reviewing the dendrogram visualisation of the clustering results

Hierarchical Clustering

- Dendogram: a tree graph that is useful for visually displaying taxonomies, lineages, and relatedness



Hierarchical Clustering Parameters

- Distance Metric
  - Euclidean
    - “ordinary” straight-line distance between two points
    - Default for calculating distance between two points

- Linkage Parameter
  - Ward
    - Informs our algorithm which method to use to calculate distance between our newly formed clusters
    - Ward seeks to minimize the variance when forming clusters
    - It minimizes the total within-cluster variance.
    - At each step, the pair of clusters with minimum between-cluster distance are merged

Hierarchical Clustering Use Cases

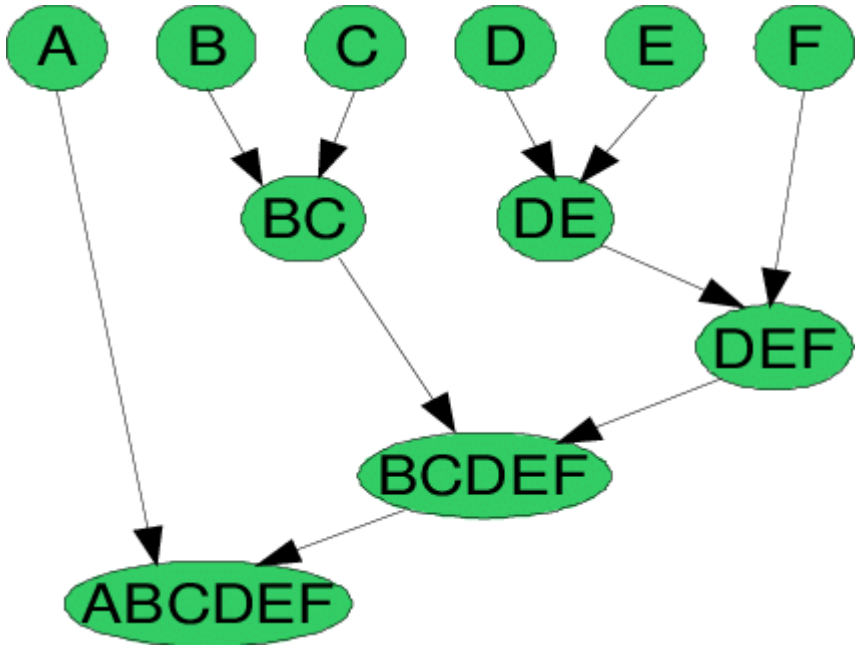
- Hospital Resource Management
- Business Process Management
- Customer Segmentation
- Social Network Analysis

Agglomerative Hierarchical Clustering

- In this technique, initially each data point is considered as an individual cluster
- At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed
- In the next slide, let’s see a pictorial representation of the Agglomerative Hierarchical clustering Technique
  - Let us assume we have six data points {A,B,C,D,E,F}

Agglomerative Hierarchical Clustering

- **Step 1:** In the initial step, we calculate the proximity of individual points and consider all the six data points as individual clusters as shown in the image below.



Agglomerative Hierarchical Clustering

- **Step 2:** In step two, similar clusters are merged together and formed as a single cluster. Let's consider B,C, and D,E are similar clusters that are merged in step two. Now, we're left with four clusters which are A, BC, DE, F.

- **Step 3:** We again calculate the proximity of new clusters and merge the similar clusters to form new clusters A, BC, DEF.

Agglomerative Hierarchical Clustering

- **Step 4:** Calculate the proximity of the new clusters. The clusters DEF and BC are similar and merged together to form a new cluster. We're now left with two clusters A, BCDEF.

- **Step 5:** Finally, all the clusters are merged together and form a single cluster.