

# AAIL100: Introduction to Data Science

Ragesh Jaiswal, IITD

# Hierarchical Clustering

- Hierarchical Clustering Algorithms fall into the following two broad categories:
  - Divisive Algorithms: This is a top-down approach where one starts with all points being in a single cluster and then recursively dividing a cluster into more than one clusters.
  - Agglomerative Algorithms: This is a bottom-up approach where one starts with  $n$  cluster (one each for every data point) and then merge small clusters to construct larger clusters.
- Agglomerative algorithms follow the following general template:
  - Start with  $n$  clusters, each a single data point and then repeatedly merge the two “closest” clusters.

# Hierarchical Clustering

- Agglomerative algorithms follow the following general template:
  - Start with  $n$  clusters, each a single data point and then repeatedly merge the two “closest” clusters.
- Following are the typical notions of closeness of two clusters that are used in practice:
  - Single linkage:  $\min_{i,j} \{D(i,j) | i \in C, j \in C'\}$ .
  - Complete linkage:  $\max_{i,j} \{D(i,j) | i \in C, j \in C'\}$ .
  - Average linkage:
    - $\text{mean}\{D(i,j) | i \in C, j \in C'\}$ .
    - $\|\text{mean}(C) - \text{mean}(C')\|^2$ .
    - $\frac{|C| \cdot |C'|}{|C| + |C'|} \cdot \|\text{mean}(C) - \text{mean}(C')\|^2$  (Ward's Method)

## Theorem

For any  $C, C' \subset \mathbb{R}^d$ , we have

$$\Delta_1(C \cup C') = \Delta_1(C) + \Delta_1(C') + \frac{|C| \cdot |C'|}{|C| + |C'|} \cdot \|\text{mean}(C) - \text{mean}(C')\|^2.$$

Here  $\Delta_1(\cdot)$  denotes the optimal 1-means cost.

End