Hierarchical Clustering

Building a **dendrogram** of clusters

Faculty of Mathematics and Computer Science, University of Bucharest
and
Sparktech Software

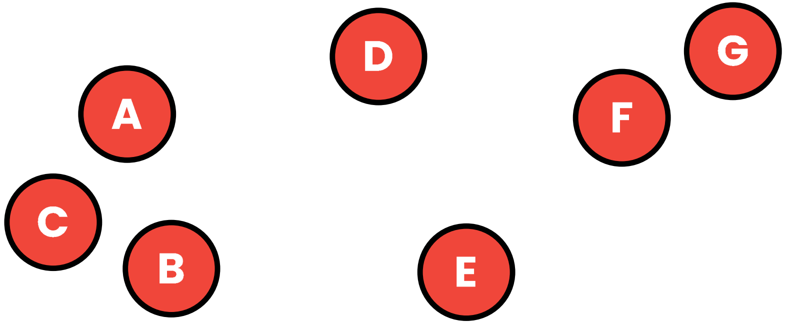
Academic Year 2018/2019, 1st Semester

Hierarchical Clustering

- **Hierarchical Clustering** is a clustering method which seeks to build a hierarchy of clusters.
    - i.e. Each cluster is made up of smaller clusters.

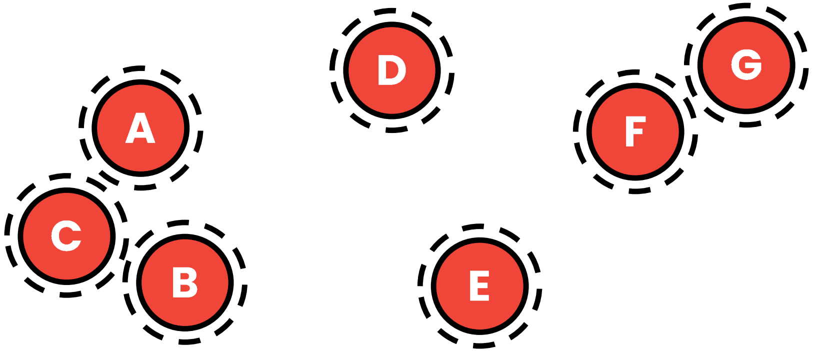
- There are two strategies:
    - **Agglomerative** (or *"bottom-up"*) – each point starts in its own cluster and pairs of clusters are merged until only one cluster remains.
    - **Divisive** (or *"top-down")* – There is a single cluster for a the whole dataset and it is recursively split until each point is in its own cluster. (very rarely used in practice).

Agglomerative Clustering



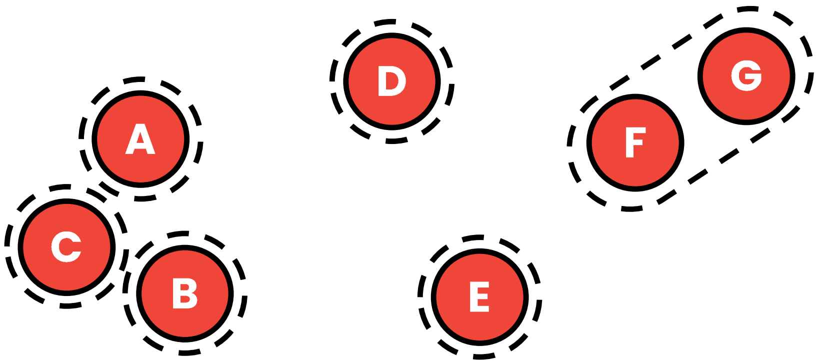
Agglomerative Clustering

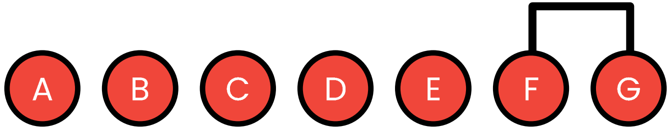
- Each point starts as its own cluster.



Agglomerative Clustering

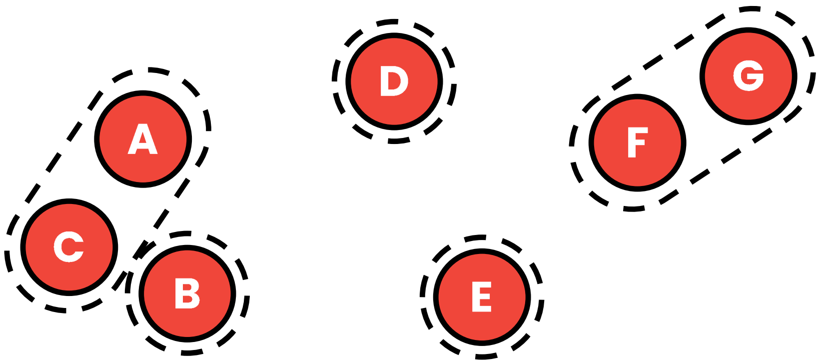
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**

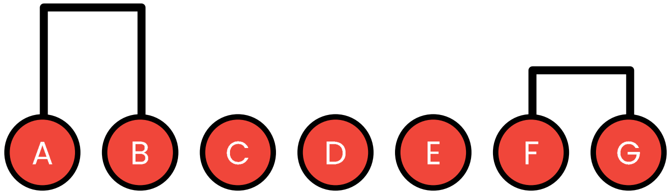




Agglomerative Clustering

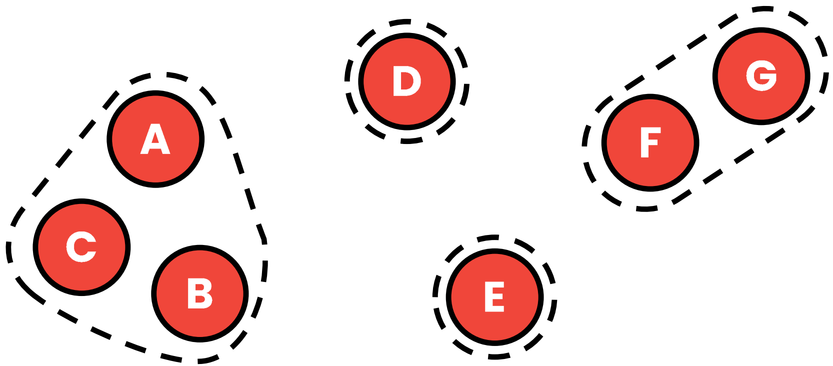
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**

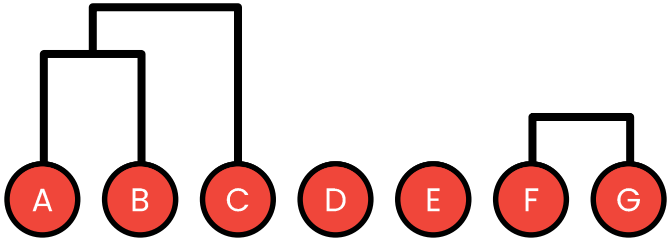




Agglomerative Clustering

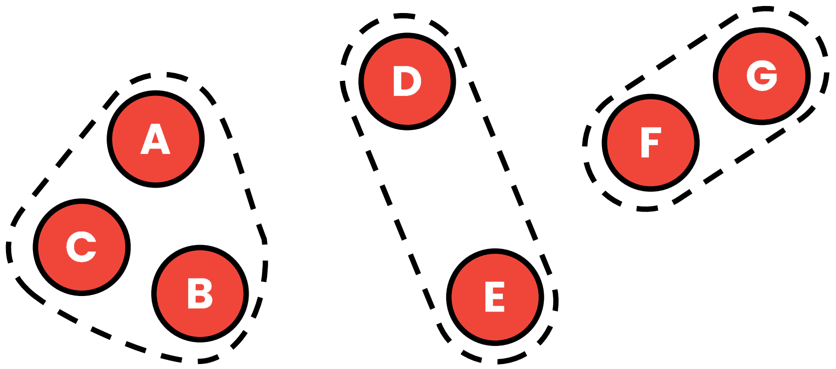
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**

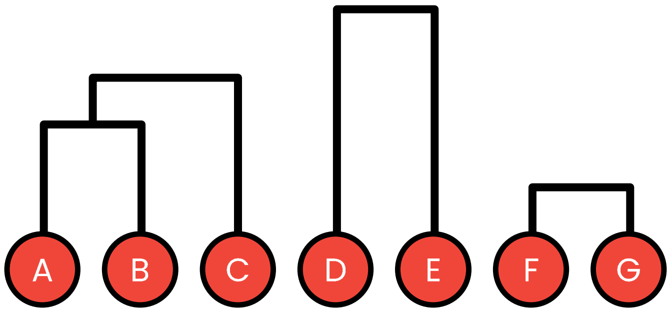




Agglomerative Clustering

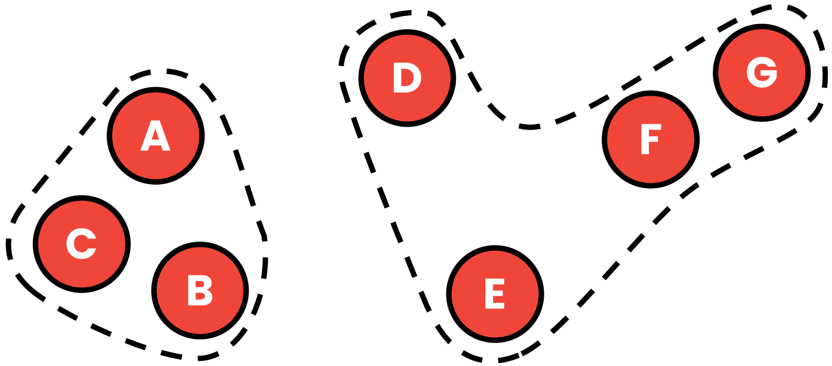
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**

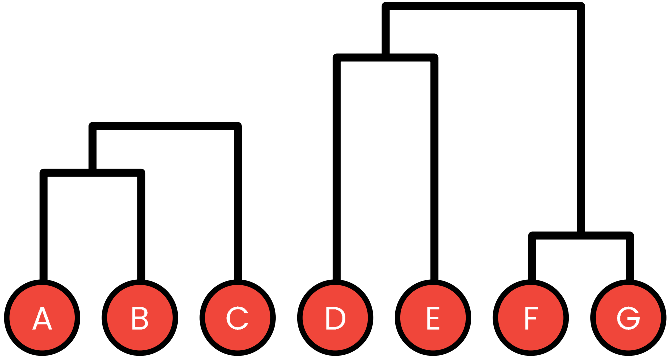




Agglomerative Clustering

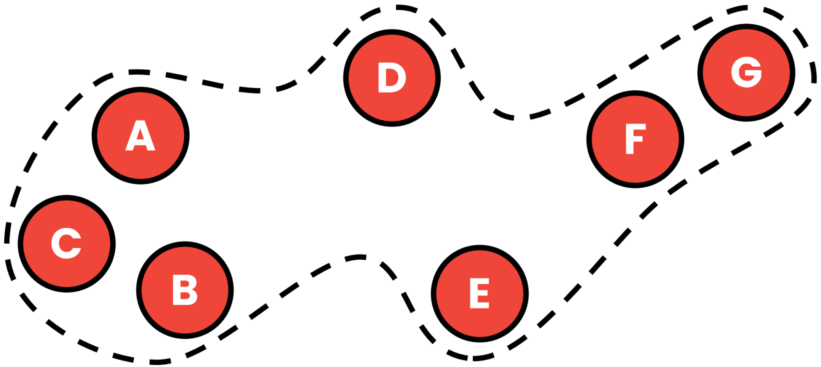
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**

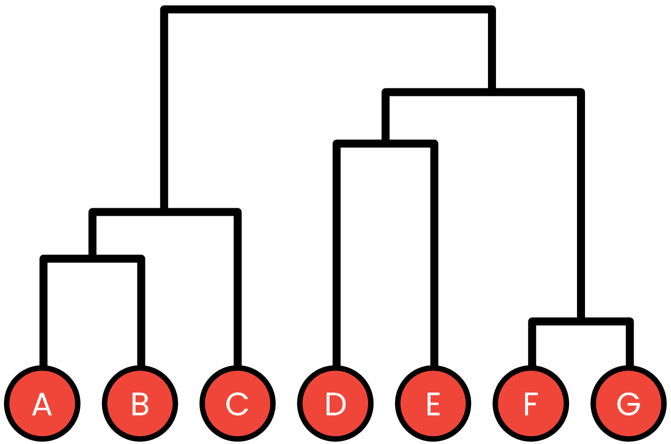




Agglomerative Clustering

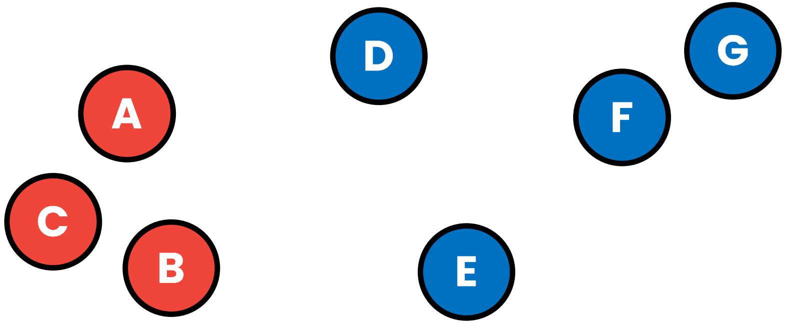
- Each point starts as its own cluster.
- **At every step, the two most similar clusters are merged.**



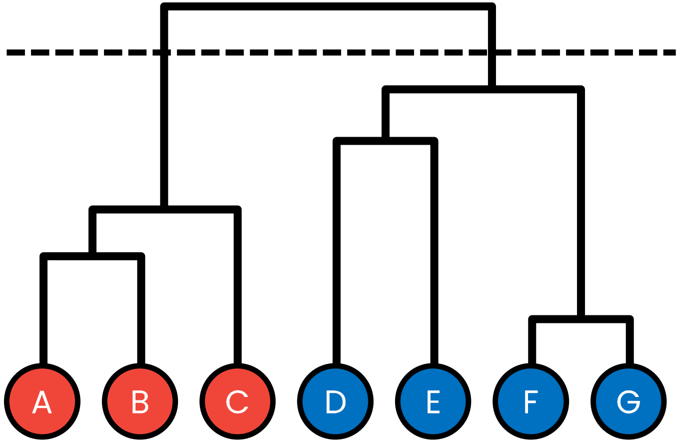


The diagram which shows the tree of clusters is called a **dendrogram.**

Agglomerative Clustering

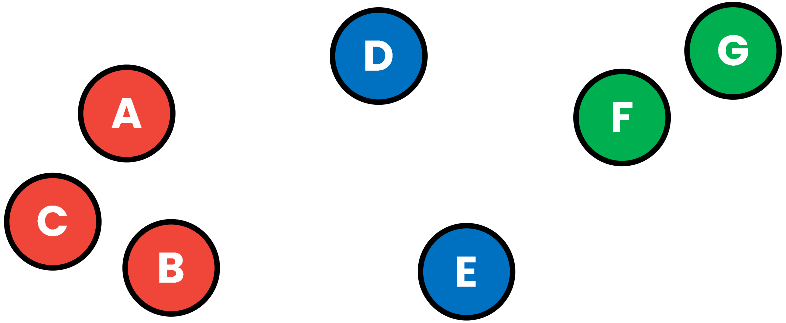


- Each point starts as its own cluster.
- At every step, the two most similar clusters are merged.
- **We can cut at any level to get a clustering.**

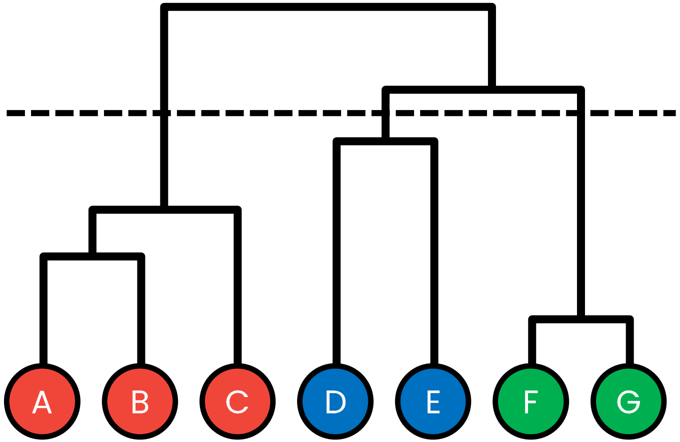


The diagram which shows the tree of clusters is called a **dendrogram.**

Agglomerative Clustering



- Each point starts as its own cluster.
- At every step, the two most similar clusters are merged.
- **We can cut at any level to get a clustering.**



The diagram which shows the tree of clusters is called a **dendrogram.**

Pseudocode

```
def AgglomerativeClustering(X = {x⃗^(1), x⃗^(2), ..., x⃗^(m)})
    C = {C_1 = {x⃗^(1)}, C_2 = {x⃗^(2)}, ..., C_m = {x⃗^(m)}} # each point as its own cluster
    steps = [] # steps to recreate the dendrogram
    while len(C) > 1:
        i*, j* = argmin_{i,j}[dist(C_i, C_j)] # distance between clusters
        C = C − {C_i*, C_j*}
        C = C + {C_i* ∪ C_j*}
        steps.append((i*, j*))
    return steps
```

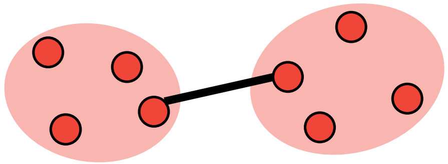
Cluster distance

- We need to define a way to measure distance between clusters (called **linkage criterion**).

Cluster distance

- We need to define a way to measure distance between clusters (called **linkage criterion**).

Single-linkage

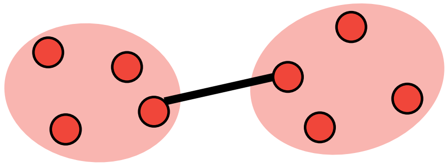


$$\mathrm{dist}(C_1, C_2) = \min_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\mathrm{dist}(\vec{a}, \vec{b})\right]$$

Cluster distance

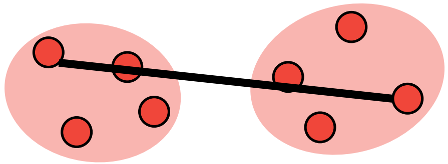
- We need to define a way to measure distance between clusters (called **linkage criterion**).

Single-linkage



$$\mathrm{dist}(C_1, C_2) = \min_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\mathrm{dist}(\vec{a}, \vec{b})\right]$$

Complete-linkage

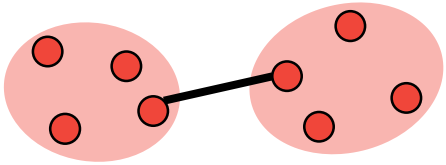


$$\mathrm{dist}(C_1, C_2) = \max_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\mathrm{dist}(\vec{a}, \vec{b})\right]$$

Cluster distance

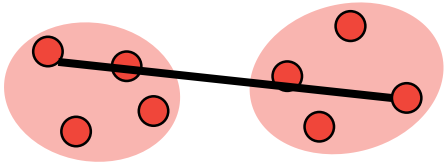
- We need to define a way to measure distance between clusters (called **linkage criterion**).

Single-linkage



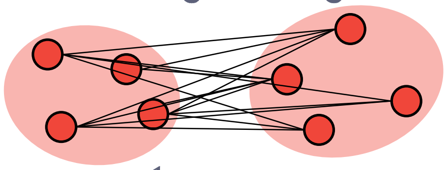
$$\text{dist}(C_1, C_2) = \min_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Complete-linkage



$$\text{dist}(C_1, C_2) = \max_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Average-linkage

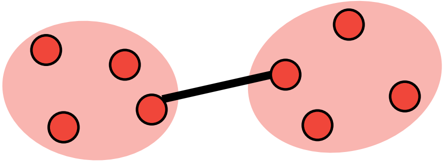


$$\text{dist}(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{\vec{a}\in C_1} \sum_{\vec{b}\in C_2} \text{dist}\left(\vec{a}, \vec{b}\right)$$

Cluster distance

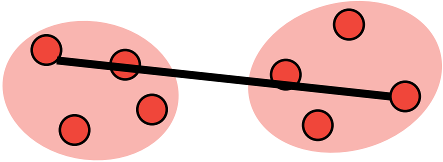
- We need to define a way to measure distance between clusters (called **linkage criterion**).

Single-linkage



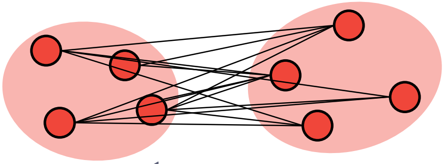
$$\text{dist}(C_1, C_2) = \min_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Complete-linkage



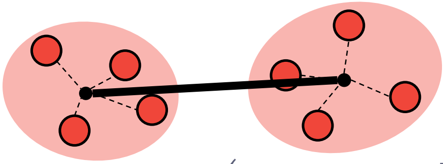
$$\text{dist}(C_1, C_2) = \max_{\vec{a}\in C_1, \vec{b}\in C_2} \left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Average-linkage



$$\text{dist}(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{\vec{a}\in C_1} \sum_{\vec{b}\in C_2} \text{dist}\left(\vec{a}, \vec{b}\right)$$

Centroid-linkage

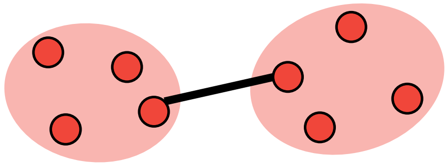


$$\text{dist}(C_1, C_2) = \text{dist}\left(\frac{\sum_{\vec{a}\in C_1} \vec{a}}{|C_1|}, \frac{\sum_{\vec{b}\in C_2} \vec{b}}{|C_2|}\right)$$

Cluster distance

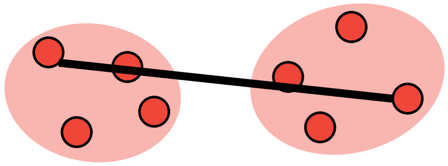
- We need to define a way to measure distance between clusters (called **linkage criterion**).

Single-linkage



$$\text{dist}(C_1, C_2) = \min_{\vec{a}\in C_1, \vec{b}\in C_2}\left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Complete-linkage

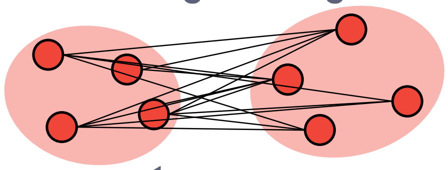


$$\text{dist}(C_1, C_2) = \max_{\vec{a}\in C_1, \vec{b}\in C_2}\left[\text{dist}\left(\vec{a}, \vec{b}\right)\right]$$

Any **distance metric** can be used.

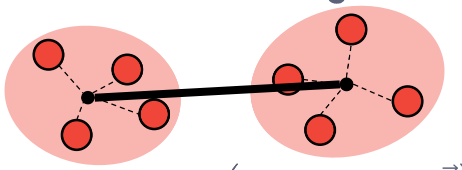
Not all distance metrics can have centroids.

Average-linkage



$$\text{dist}(C_1, C_2) = \frac{1}{|C_1||C_2|}\sum_{\vec{a}\in C_1}\sum_{\vec{b}\in C_2}\text{dist}\left(\vec{a}, \vec{b}\right)$$

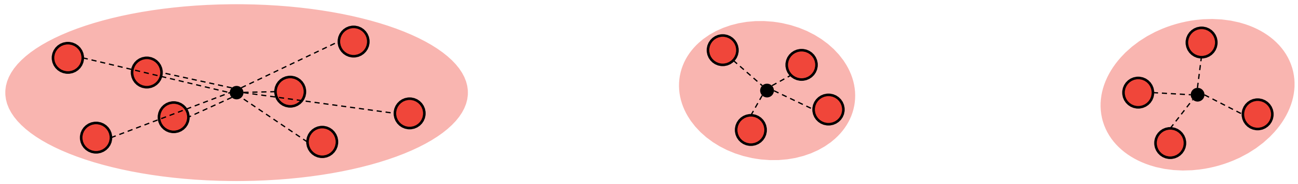
Centroid-linkage



$$\text{dist}(C_1, C_2) = \text{dist}\left(\frac{\sum_{\vec{a}\in C_1}\vec{a}}{|C_1|}, \frac{\sum_{\vec{b}\in C_2}\vec{b}}{|C_2|}\right)$$

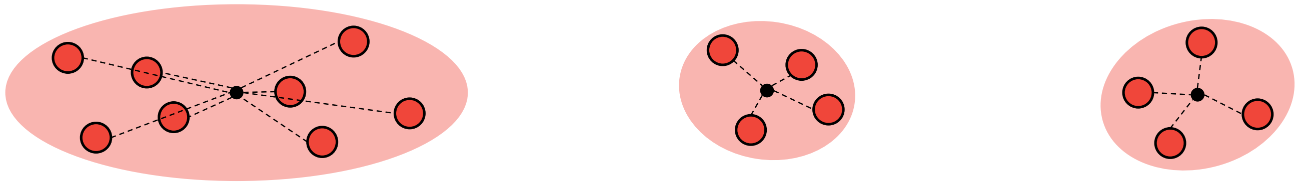
Cluster distance

- **Ward's criterion** defines the distance between clusters as the *increase in variance* due to merging the two clusters.



Cluster distance

- **Ward's criterion** defines the distance between clusters as the *increase in variance* due to merging the two clusters.



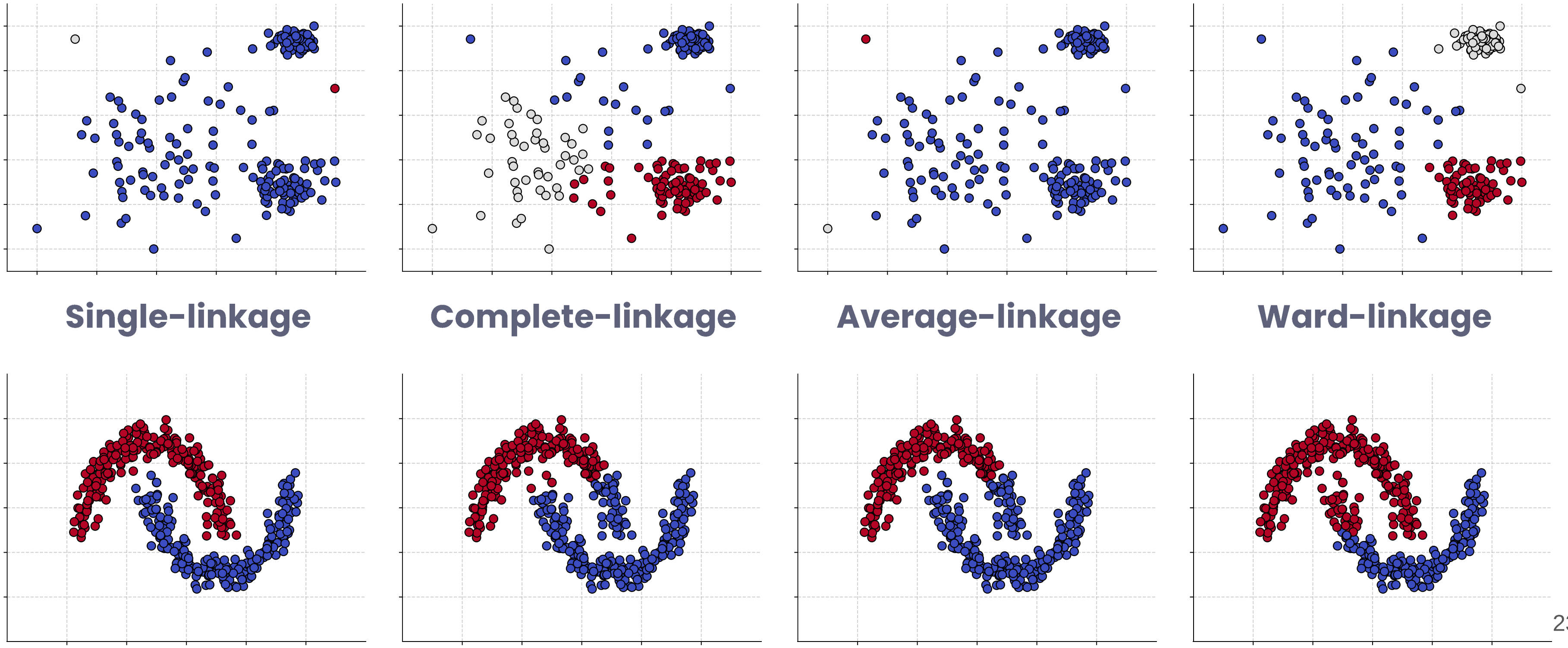
$$\mathrm{dist}(C_1, C_2) = \frac{1}{|C_1 \cup C_2|} \sum_{\vec{x} \in C_1 \cup C_2} \mathrm{dist}\left(\vec{x}, \vec{\mu}_{C_1 \cup C_2}\right) - \frac{1}{|C_1|} \sum_{\vec{a} \in C_1} \mathrm{dist}\left(\vec{a}, \vec{\mu}_{C_1}\right) - \frac{1}{|C_2|} \sum_{\vec{b} \in C_2} \mathrm{dist}\left(\vec{b}, \vec{\mu}_{C_2}\right)$$

$$= \quad Var(C_1 \cup C_2) \quad - \quad Var(C_1) \quad - \quad Var(C_2)$$

Linkage Criteria comparison

- **Single-linkage**
  - Tends to produce long, chain-like clusters.
  - The 2 furthest elements in a cluster might be far apart from each other.
  - Can handle non-convex shapes.
  - Sensitive to noise.
- **Complete-linkage**
  - Tries to keep all points in a cluster close.
  - Tends to produce spherical, compact clusters.
  - Less sensitive to noise than single-linkage
- **Average-linkage**
  - Compromise between single and complete linkages, but produces results closer to complete-linkage
- **Ward-linkage**
  - Similar results to complete-linkage, but it handles clusters with various densities better.

Linkage Criteria comparison



Summary

- **Agglomerative Clustering** is a **hierarchical clustering** method of producing a **dendrogram** of clusters.
- It works by starting with each point in its own cluster and then iteratively selecting the two *closest clusters* and merging them.
- The distance between clusters (called **linkage criterion**) can be defined in multiple ways (**single-linkage, complete-linkage, average-linkage, ward-linkage**).
  - The selected *distance metric* between points and *linkage criterion* have an effect on the *shapes and sizes* of the produces clusters.

Keywords

Hierarchical Clustering

Agglomerative Clustering

Divisive Clustering

Dendrogram

Linkage Criterion

Single-Linkage

Complete-Linkage

Average-Linkage

Ward’s criterion