Hierarchical Clustering

TEAM YOLO
Zeqing Jin
Xianlin Shao
Yifei Zhang
Zilan Zhang

Berkeley
UNIVERSITY OF CALIFORNIA

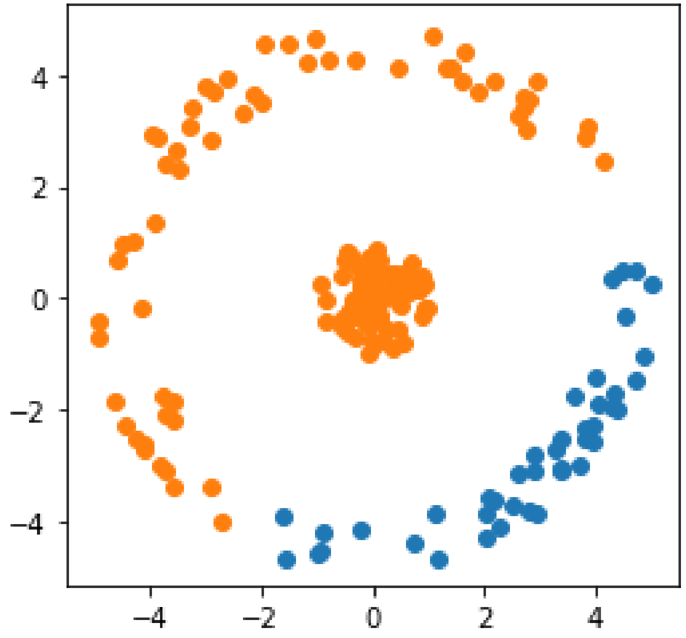
Introduction

- Hierarchical clustering is a general family of clustering algorithms that build clusters by merging or splitting them successively. [2]

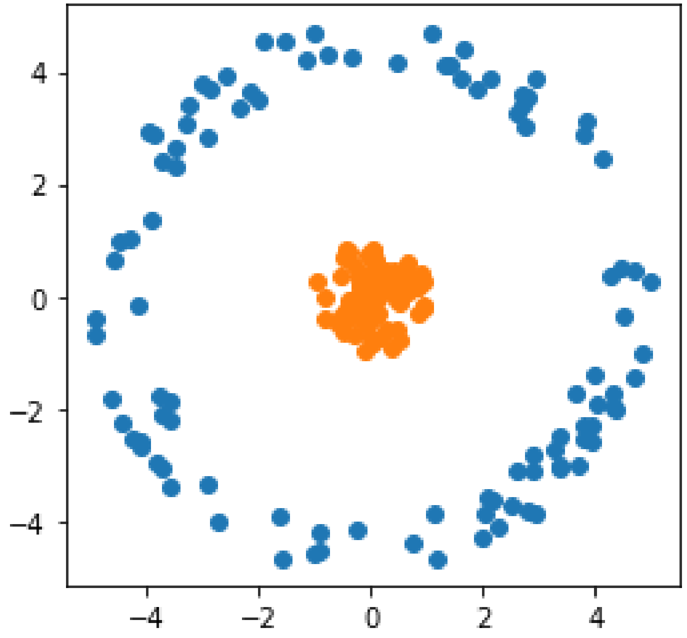
- Two common hierarchy algorithms:
  - Agglomerative clustering
  - Divisive clustering

Limitations of K-means Clustering

- Non-spherical data points
- Prior assumption of similar number of data points in each cluster



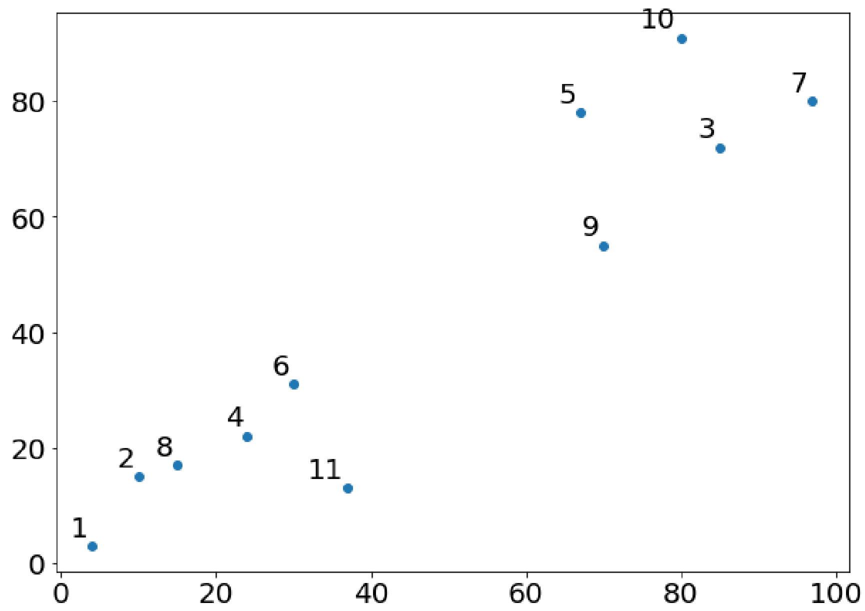
K-means

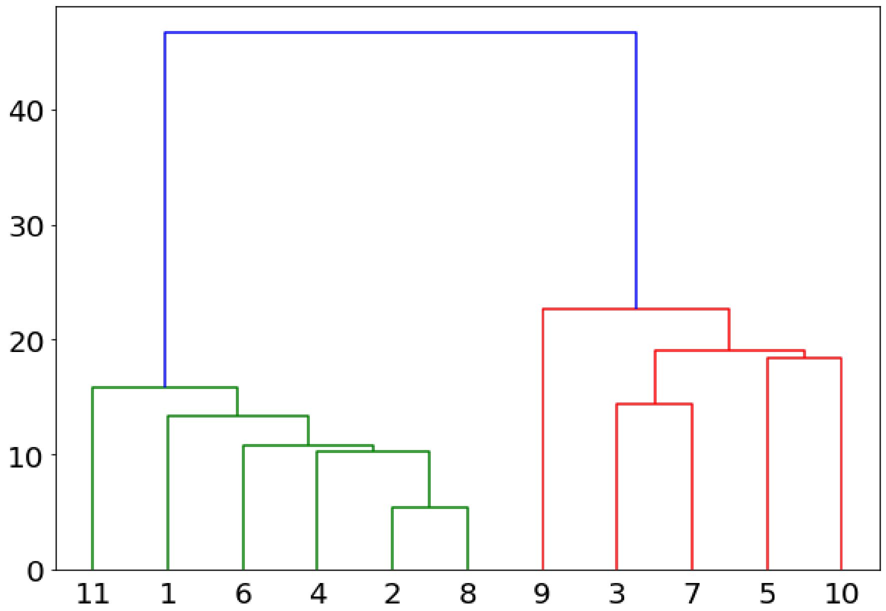


Hierarchical

Dendrogram

- Dendrogram is a tree-like hierarchy which shows the relationship between objects.

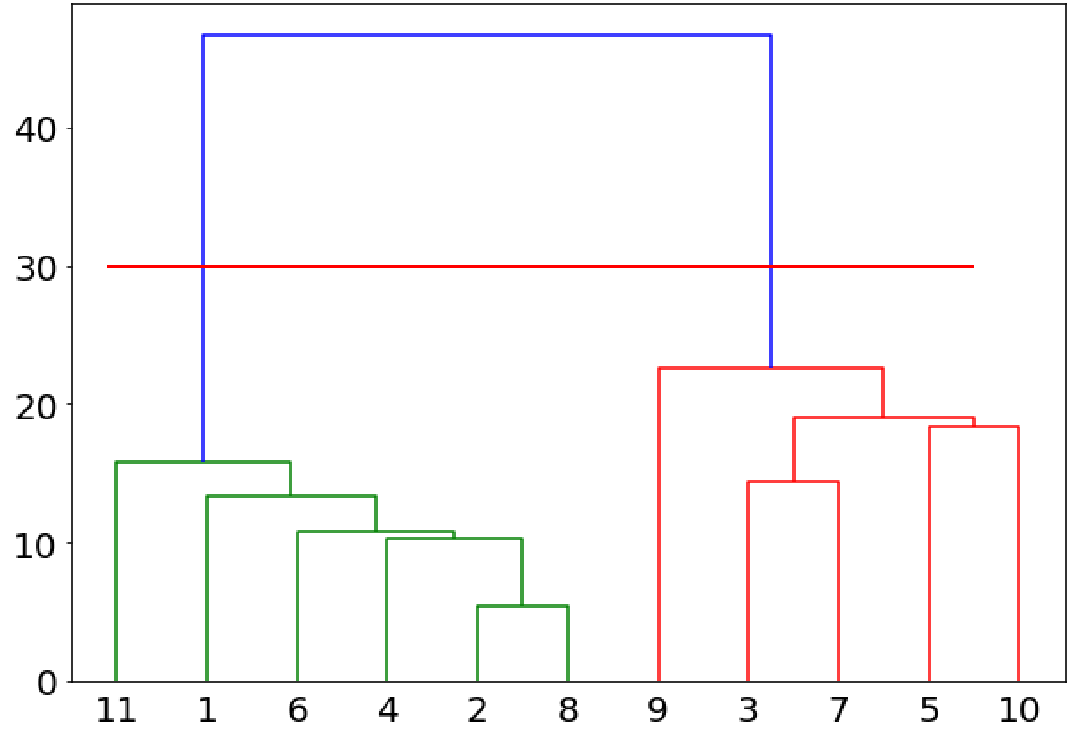




Berkeley
UNIVERSITY OF CALIFORNIA

Dendrogram

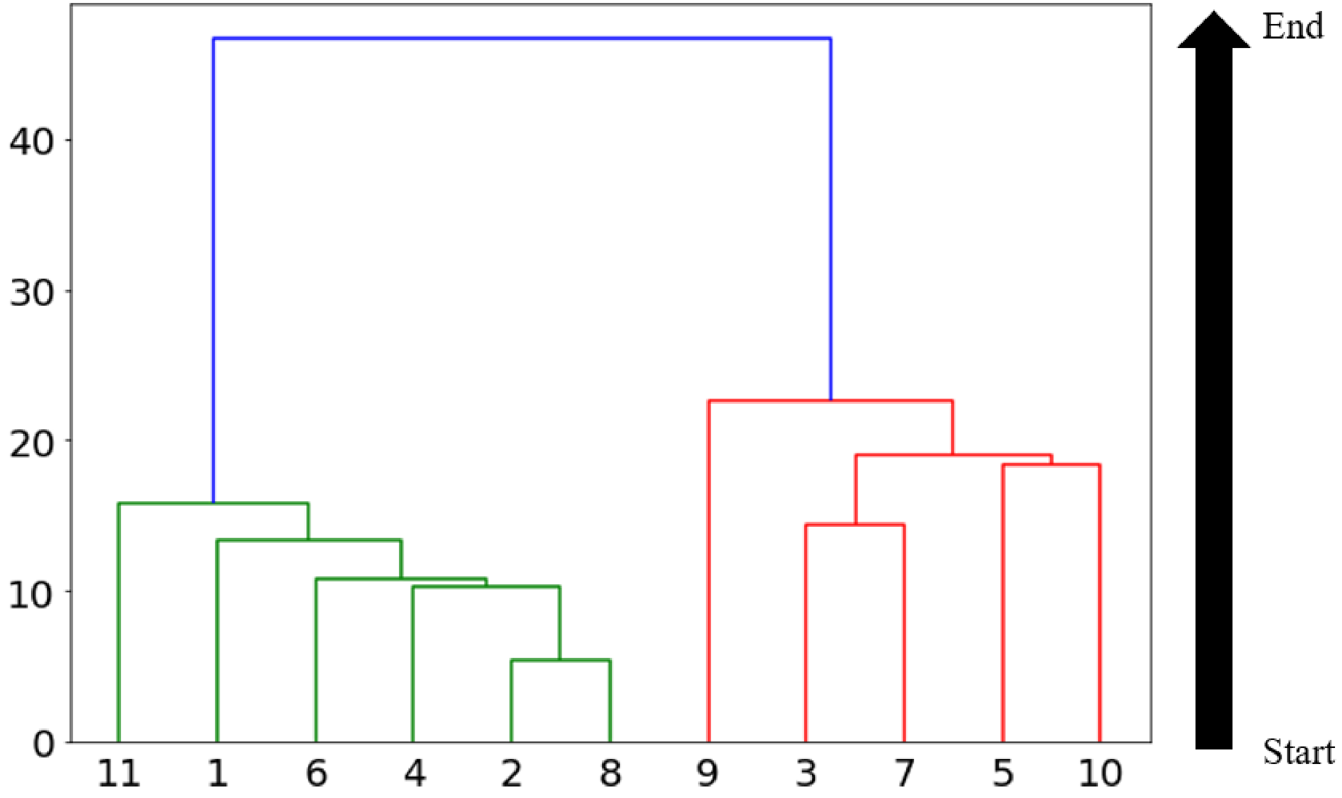
- Dendrogram implicitly contains all possible values of the number of clusters
- Shows relative relations between clusters (points)
- Fails to show all the absolute distances between points



Berkeley
UNIVERSITY OF CALIFORNIA

Agglomerative Clustering

- Start with n clusters containing one single point.
- End up with one cluster containing n objects.



Berkeley
UNIVERSITY OF CALIFORNIA

Agglomerative Clustering [1]

Algorithm 1: Agglomerative Hierarchical Clustering

Input: n data points
Output: final clustering result over n data points

```
Initialize n clusters c_i, i = 1, ..., n;
Initialize the dissimilarity matrix;
for the number of clusters k decreases from n to 1 do
    Merge the two clusters c_i, c_j with the smallest dissimilarity according to dissimilarity matrix;
    Update the dissimilarity matrix;
end for
```

Agglomerative Clustering

- Euclidean distance between points:

$$d(i,j) = \sqrt{\Sigma_{p-1}^{q}\left(x_{ip} - x_{jp}\right)^2}$$

  where q is dimension of the point

- Dissimilarity matrix:

$$\boldsymbol{S} = \begin{bmatrix} d(1,1) & \cdots & d(1,n) \\ \vdots & \ddots & \vdots \\ d(n,1) & \cdots & d(n,n) \end{bmatrix}$$

Berkeley
UNIVERSITY OF CALIFORNIA

Distance between clusters

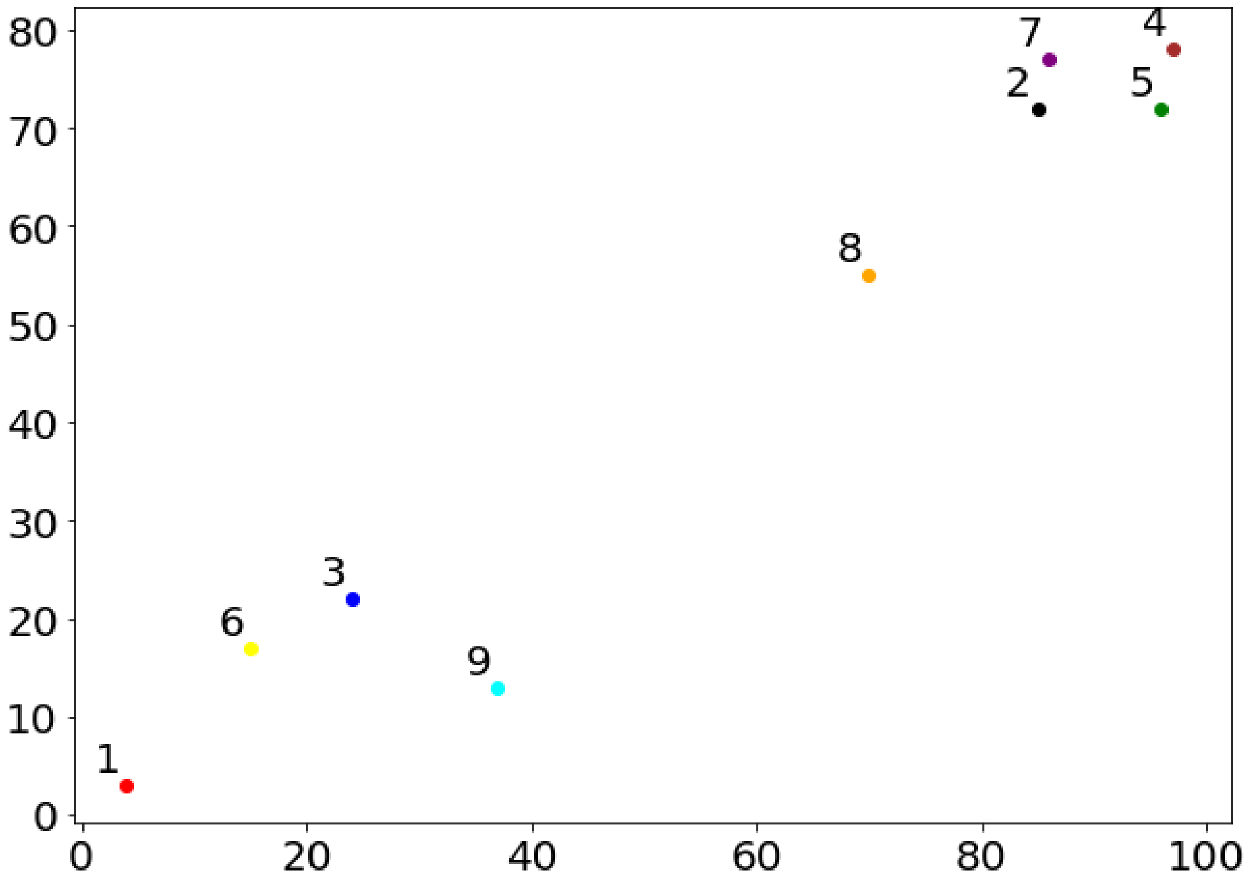
- **Complete linkage:**
  - Maximum distance between clusters
- **Single linkage:**
  - Minimum distance between clusters
- **Average linkage:**
  - Average distance between clusters
- **Centroid linkage:**
  - Distance between centroids of clusters
- **Ward's linkage:**
  - Increase in sum of squares if two clusters are merged

Berkeley
UNIVERSITY OF CALIFORNIA

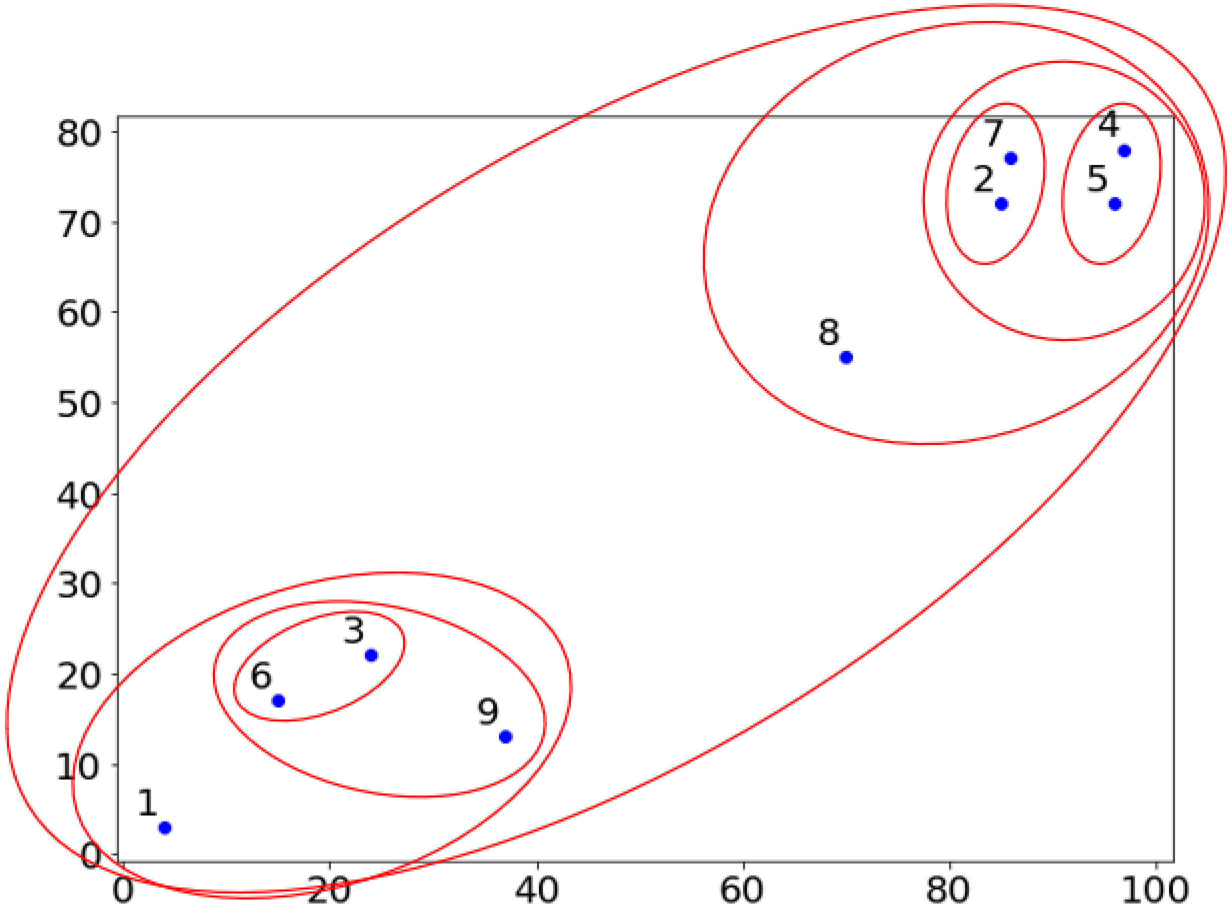
Agglomerative Clustering Example

- Start from 9 clusters
- Complete linkage
- 9×9 dissimilarity matrix

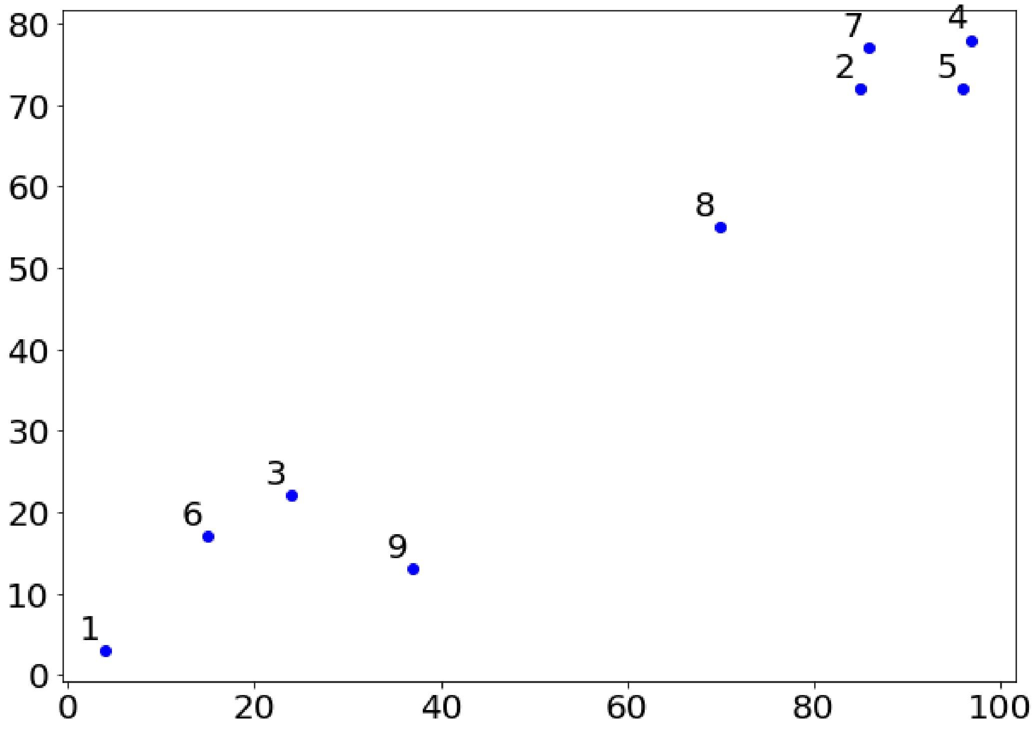
How many distances do we need to calculate? 81?

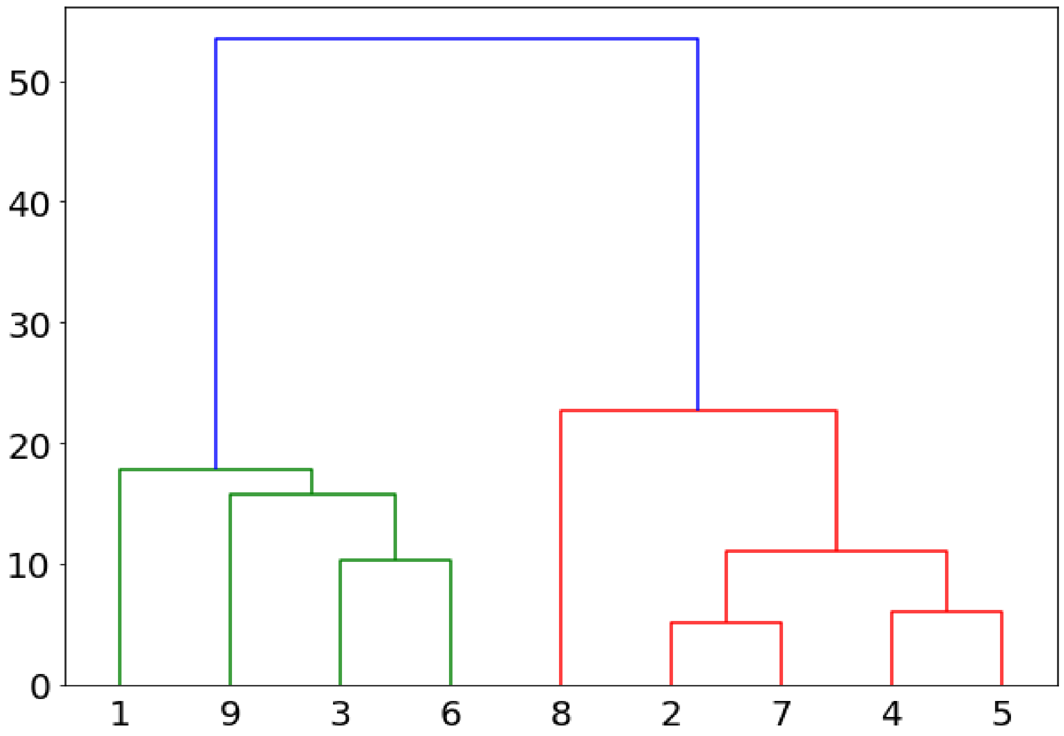


Agglomerative Clustering Example



Agglomerative Clustering Example

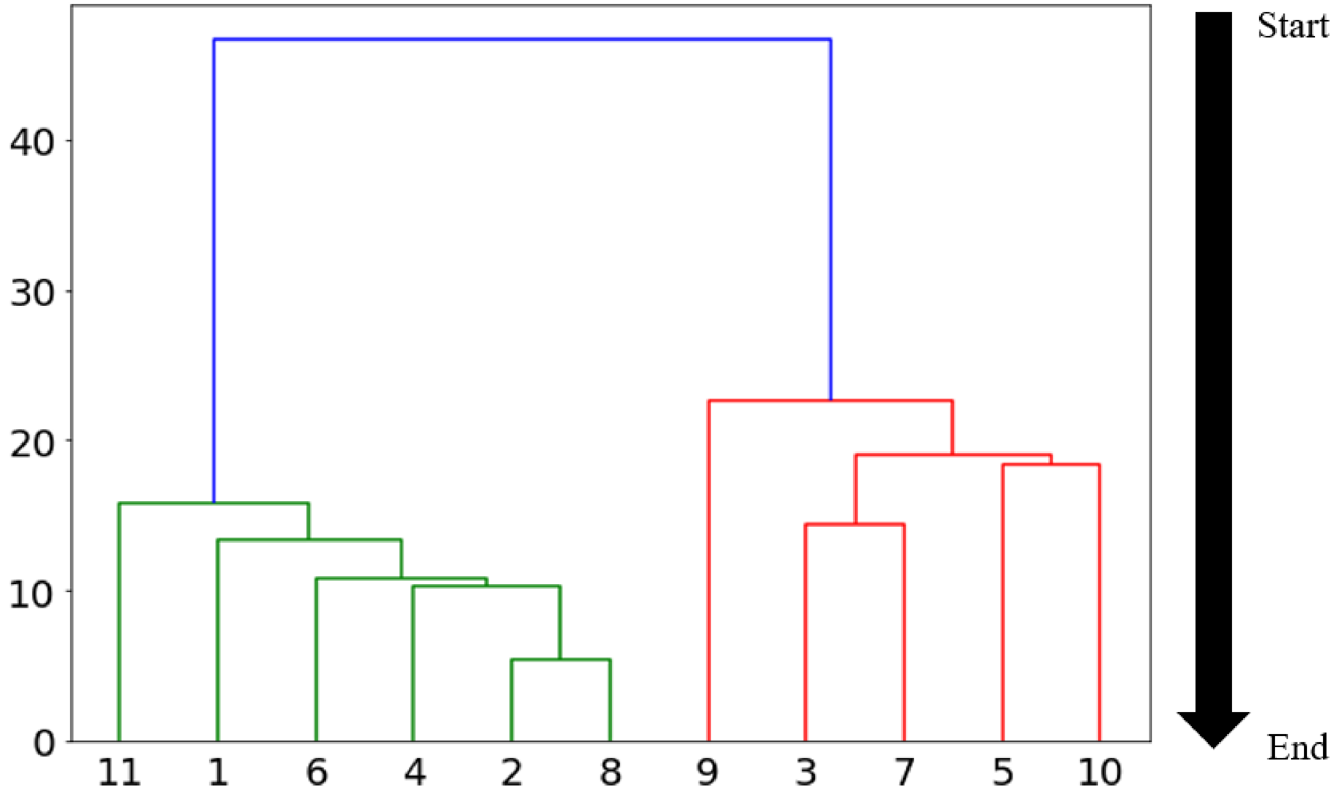




Divisive Clustering (DIANA)

- Start with one cluster containing all n points.

- End up with n clusters containing one object.



Divisive Clustering (DIANA) [1]

Algorithm 2: Divisive Analysis Clustering (DIANA)

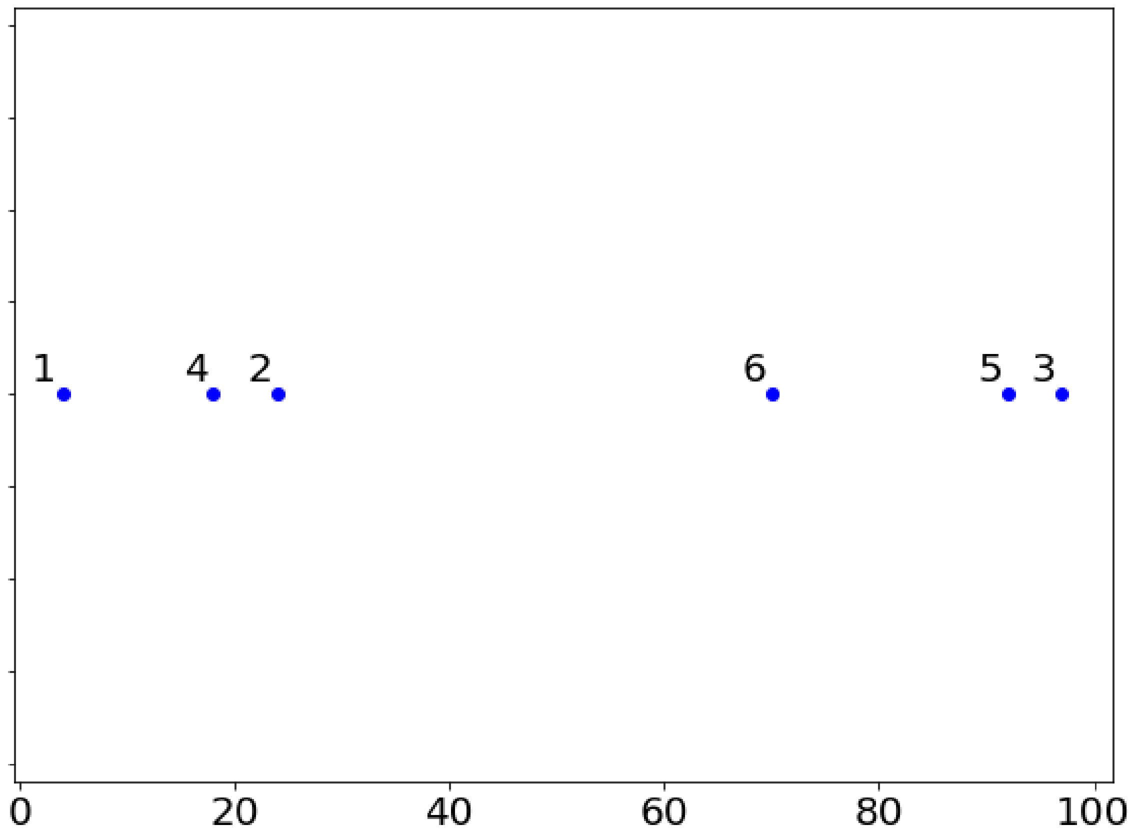
Input: n data points
Output: final clustering result

Initialize one cluster with all objects $\mathbf{c}_1$;
for *the number of clusters k increases from 1 to n* **do**
  Choose the cluster $\mathbf{C}_i$ with the largest diameter value;
  Within $\mathbf{C}_i$, choose the object that has the maximum distance with the other objects as one cluster and split this object as a splinter cluster;
  Update $\mathbf{C}_i$;
  while *True* **do**
    for *each data point j in* $\mathbf{C}_i$ **do**
      Calculate the distance d_1 between the data j and the other objects in $\mathbf{C}_i$ as one cluster;
      Calculate the distance d_2 between the data j and the splinter cluster;
      Calculate the difference $\delta d_j = d_1 - d_2$;
    end for
    if *max* δd_j *is positive* **then**
      Move the the data j with positive δd_j to the splinter cluster and update $\mathbf{C}_i$;
    else
      break;
    end if
  end while
end for

Divisive Clustering Example

- Start from one cluster
- Complete linkage
- 6×6 dissimilarity matrix

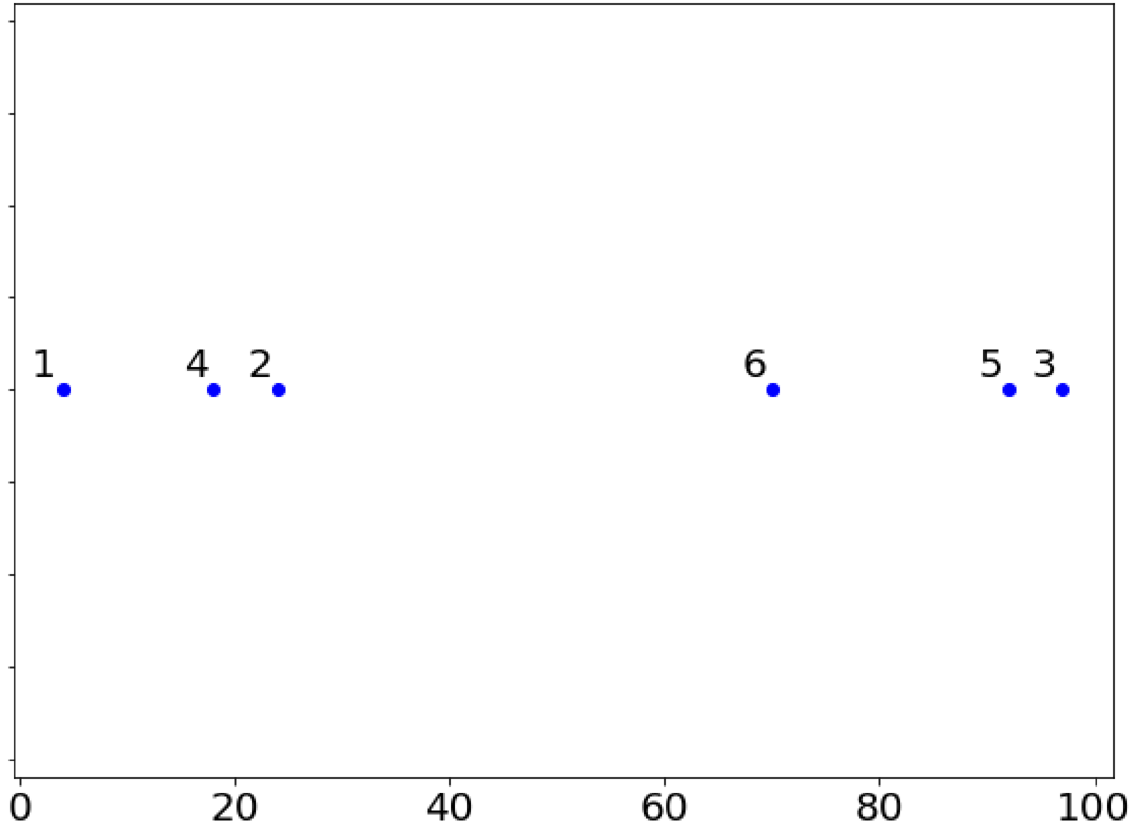
$$S = \begin{bmatrix} 0 & 20 & 93 & 14 & 88 & 66 \\ 20 & 0 & 73 & 6 & 68 & 46 \\ 93 & 73 & 0 & 79 & 5 & 27 \\ 14 & 6 & 79 & 0 & 74 & 52 \\ 88 & 68 & 5 & 74 & 0 & 22 \\ 66 & 46 & 27 & 52 & 22 & 0 \end{bmatrix}$$



Divisive Clustering Example

- Calculate the distance between each point and the other objects

Point	Distance to other points
1	93
2	73
3	93
4	79
5	88
6	66



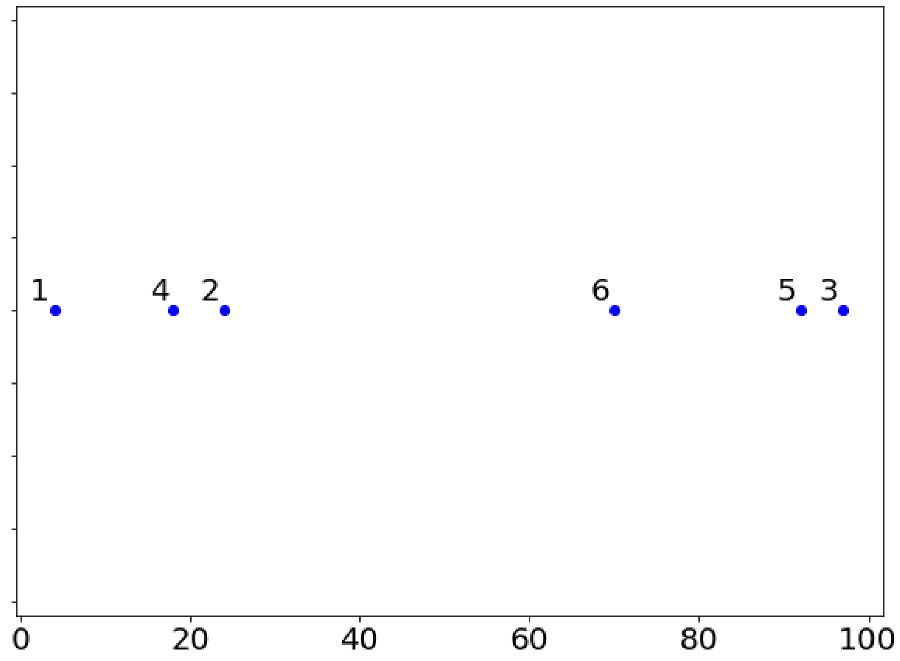
- Splinter cluster {1}

Divisive Clustering Example

- Calculate the distance between each remaining point and the other objects

- Also the distance to the splinter cluster

- Splinter cluster {1,4}

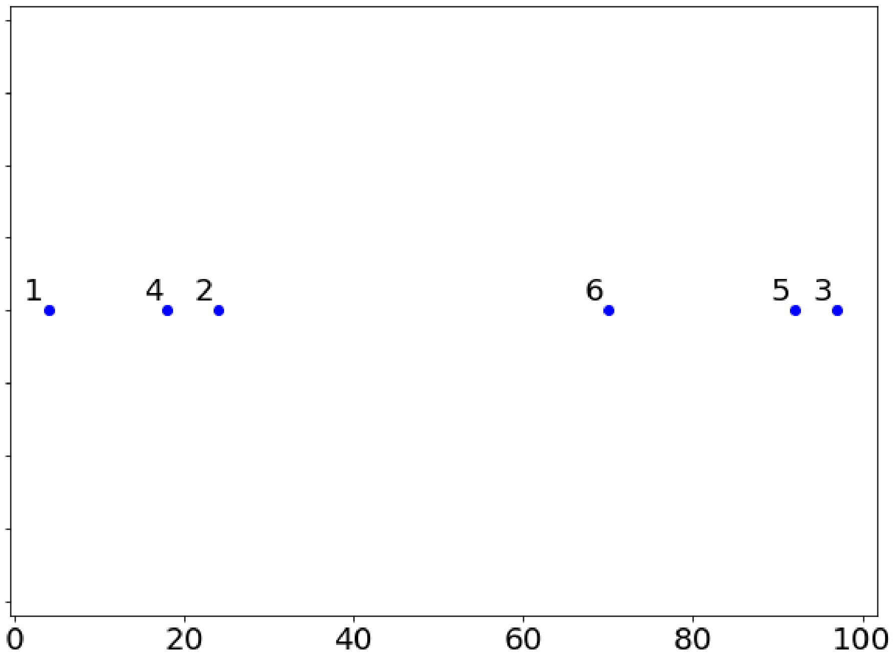


Point	Distance to other points	Distance to the splinter cluster	Difference
2	73	20	53
3	79	93	-14
4	79	14	65
5	74	88	-14
6	52	66	-14

Berkeley
UNIVERSITY OF CALIFORNIA

Divisive Clustering Example

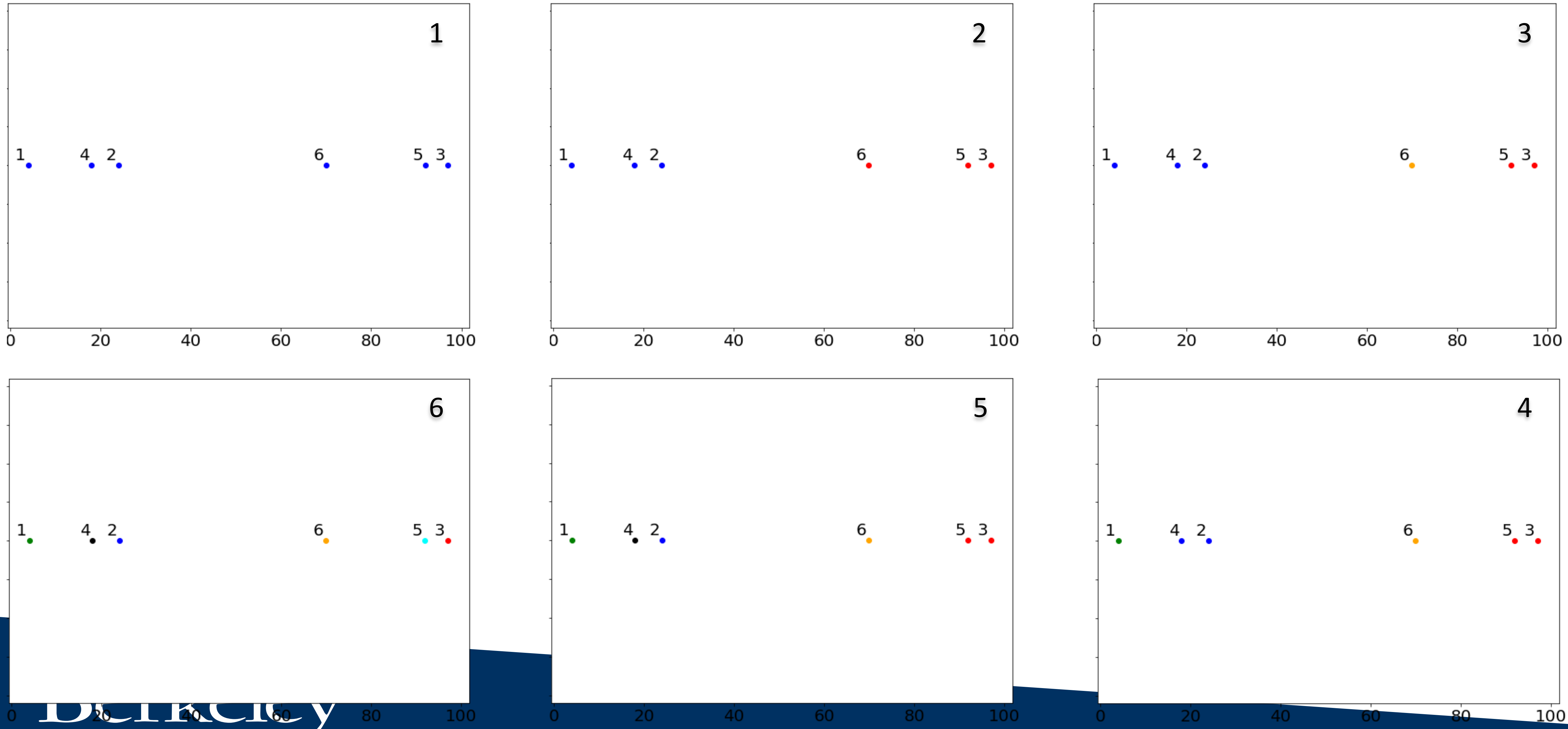
- Repeat the previous step
- Splinter cluster {1, 4, 2}
- {1,2,3,4,5,6} into {1,2,4} and {3,5,6}



Point	Distance to other points	Distance to the splinter cluster	Difference
2	73	20	53
3	73	93	-20
5	68	88	-20
6	46	66	-20

Berkeley
UNIVERSITY OF CALIFORNIA

Divisive Clustering (DIANA)



UNIVERSITY OF CALIFORNIA

Determination of *k* (General)

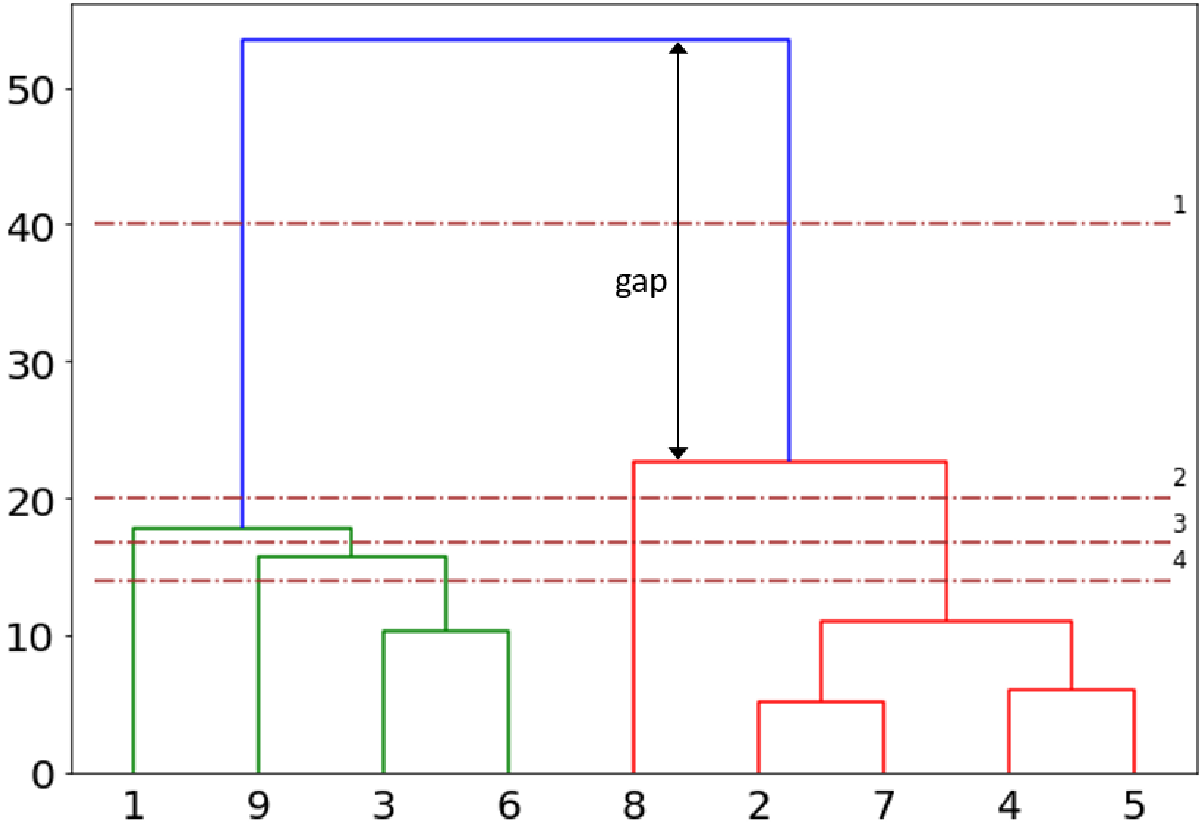
- Elbow method [2]
  - Total within-cluster sum of square (WSS)

- Average silhouette method [3]
  - Average silhouette

- Gap statistic method
  - Gap statistic

Determination of k (from Dendrograom)

- Cut at different dissimilarity levels gives multiple values of k
- Cut at the largest dissimilarity gap gives a roughly reasonable k
- Affected by the linkage type since dissimilarity may change after each iteration.



Specific Hierarchical Algorithms

- Linkage algorithm
  - Single linkage, average linkage, complete linkage

- CURE (Clustering Using REpresentatives)

- BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) (Optional)

Berkeley
UNIVERSITY OF CALIFORNIA

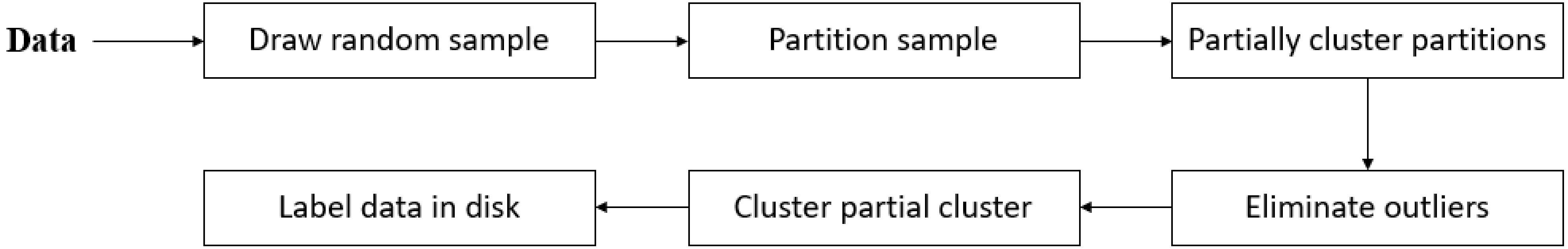
Linkage algorithm [4]

- Single linkage:
  - Time complexity $O[n^3]$ (simplest implementation)
  - Sensitive to outliers
- Complete linkage
  - Time complexity can be reduced to $O[n^2 \log n]$
  - Cluster similar objects
- Average linkage
  - Compromise between single and complete
  - Often fails in complicated cluster shapes

Berkeley
UNIVERSITY OF CALIFORNIA

CURE (Clustering Using REpresentatives) [4]

- A hierarchical based clustering technique



- Representative points and shrinking factor
- Apply to outliers

Berkeley
UNIVERSITY OF CALIFORNIA

Reference

- [1]. Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons, 2009.
- [2]. Bradley Boehmke Brandon Greenwell. *Hands-On Machine Learning with R*. Feb. 2020. URL: https : / / bradleyboehmke . github . io / HOML / hierarchical . html # fig:dendrogram2.
- [3]. Godfrey and Kate. *Determining The Optimal Number Of Clusters: 3 Must Know Methods*. Feb. 2020. URL: https://bradleyboehmke.github.io/HOML/kmeans.html#eq:tot-within-ss.
- [4]. M Kuchaki Rafsanjani, Z Asghari Varzaneh, and N Emami Chukanlo. "A survey of hierarchical clustering algorithms". In: *The Journal of Mathematics and Computer Science* 5.3 (2012), pp. 229–240.

Berkeley
UNIVERSITY OF CALIFORNIA