# Clustering Techniques

# K-Means

Objective : To minimize the sum of distances between the points and their respective cluster centroid.

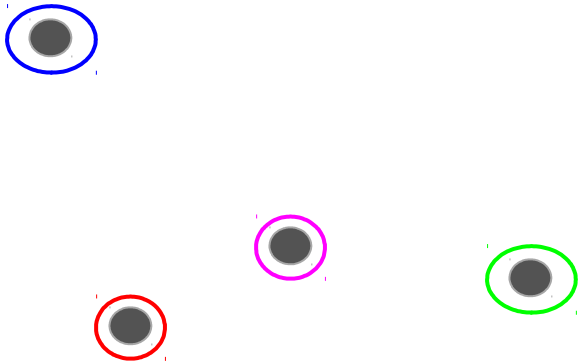
- Pick the number of clusters in the beginning

# Hierarchical Clustering

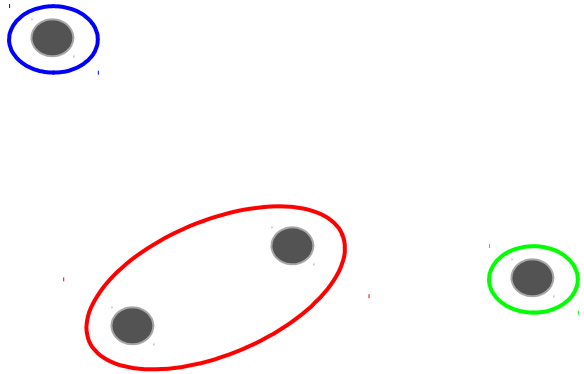
- No need to decide the number of clusters in beginning

# Hierarchical Clustering

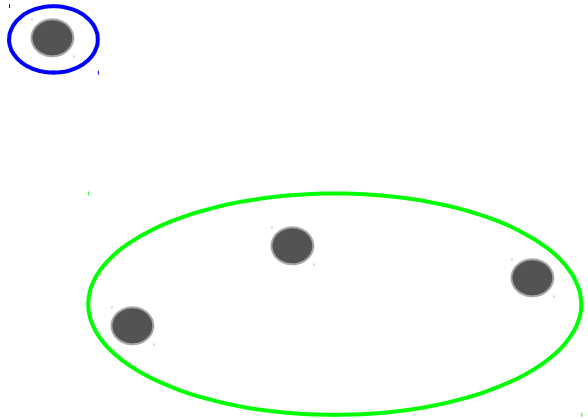
# Hierarchical Clustering



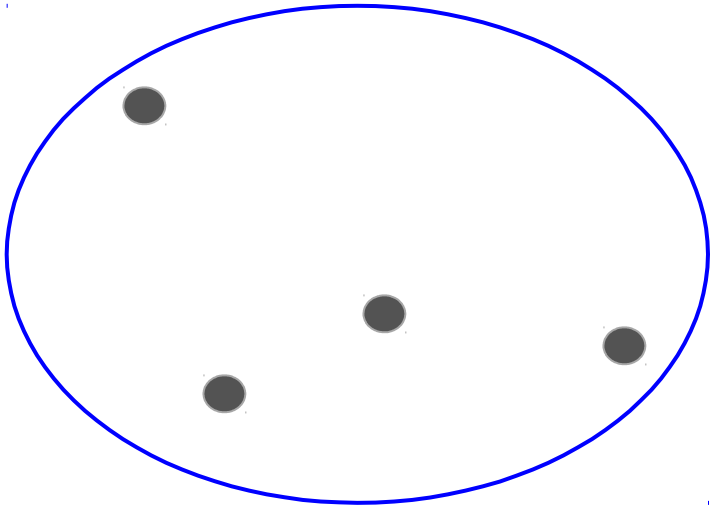
# Hierarchical Clustering



# Hierarchical Clustering



# Hierarchical Clustering

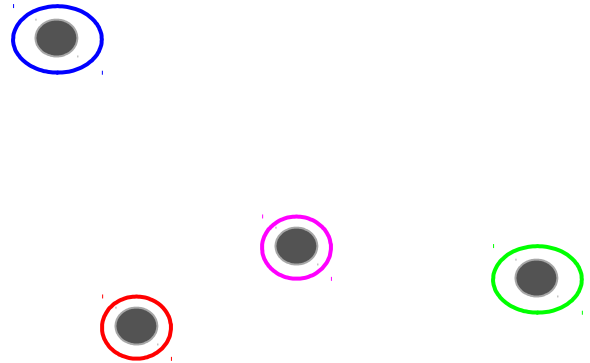


# Types of Hierarchical Clustering

- Agglomerative

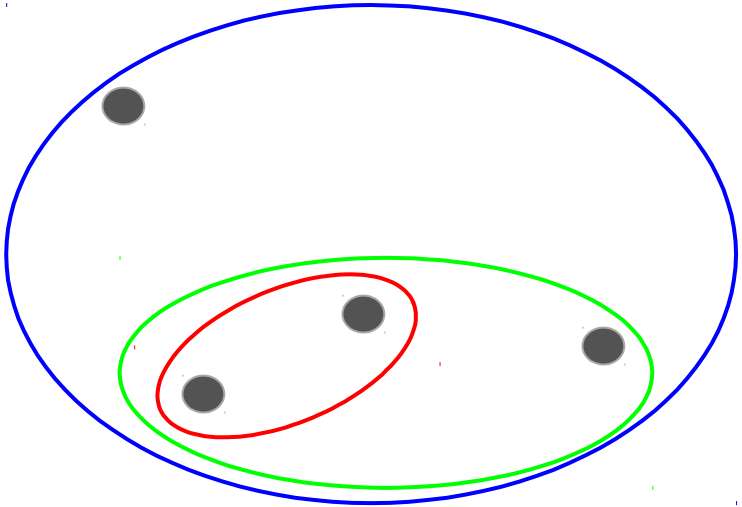
# Types of Hierarchical Clustering

- Agglomerative
  - Assign each point as a separate cluster



# Types of Hierarchical Clustering

- Agglomerative
  - Assign each point as a separate cluster
  - Merge the closest pair of clusters until only single cluster is left

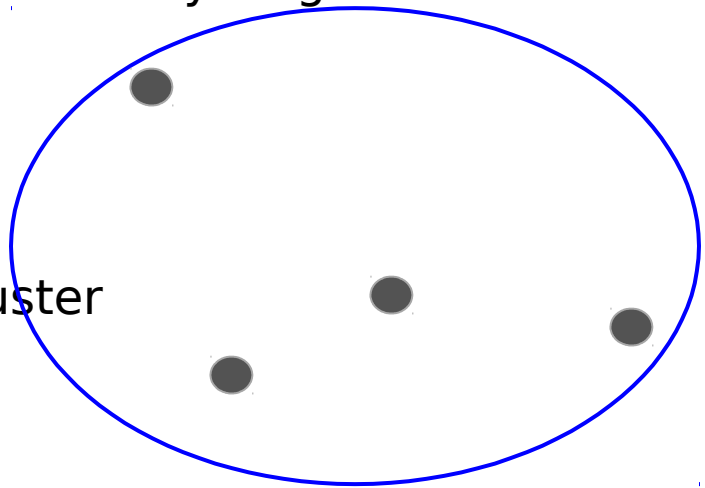


# Types of Hierarchical Clustering

- Agglomerative
  - Assign each point as a separate cluster
  - Merge the closest pair of clusters until only single cluster is left
- Divisive

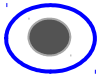
# Types of Hierarchical Clustering

- Agglomerative
  - Assign each point as a separate cluster
  - Merge the closest pair of clusters until only single cluster is left
- Divisive
  - Assign all the points to a single cluster

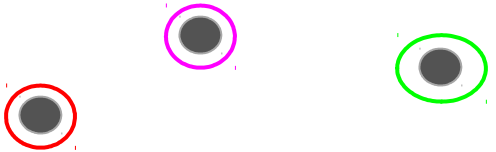


# Types of Hierarchical Clustering

- Agglomerative
  - Assign each point as a separate cluster
  - Merge the closest pair of clusters until only single cluster is left



- Divisive
  - Assign all the points to a single cluster
  - Split the clusters until each cluster only contains a single point



# Steps to perform Hierarchical Clustering

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |

Proximity Matrix

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | | | | |
| 2 | | 0 | | | |
| 3 | | | 0 | | |
| 4 | | | | 0 | |
| 5 | | | | | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 3 | | | |
| 2 | 3 | 0 | | | |
| 3 | | | 0 | | |
| 4 | | | | 0 | |
| 5 | | | | | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1** | 0 | 3 | 18 | | |
| **2** | 3 | 0 | | | |
| **3** | 18 | | 0 | | |
| **4** | | | | 0 | |
| **5** | | | | | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1** | 0 | 3 | 18 | 10 | |
| **2** | 3 | 0 | | | |
| **3** | 18 | | 0 | | |
| **4** | 10 | | | 0 | |
| **5** | | | | | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1** | 0 | 3 | 18 | 10 | 25 |
| **2** | 3 | 0 | | | |
| **3** | 18 | | 0 | | |
| **4** | 10 | | | 0 | |
| **5** | 25 | | | | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

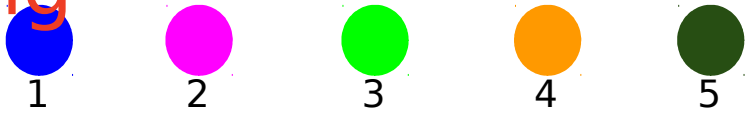
| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1** | 0 | 3 | 18 | 10 | 25 |
| **2** | 3 | 0 | 21 | 13 | 28 |
| **3** | 18 | 21 | 0 | 8 | 7 |
| **4** | 10 | 13 | 8 | 0 | 15 |
| **5** | 25 | 28 | 7 | 15 | 0 |

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

1. Assign each point as a different cluster

# Steps to perform Hierarchical Clustering



# Steps to perform Hierarchical Clustering

1. Assign each point as a different cluster

2. Merge the two closest clusters and update the proximity

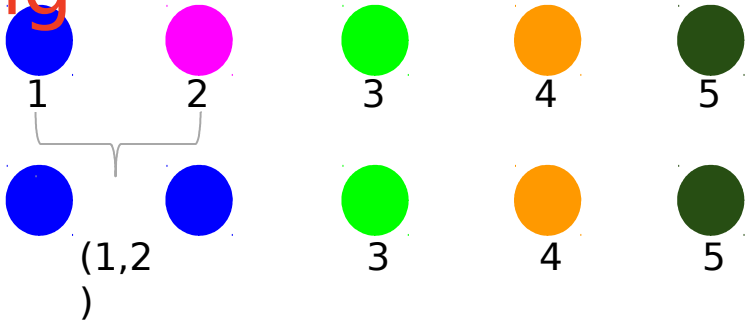
   matrix

# Steps to perform Hierarchical Clustering

| ID | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 3 | 18 | 10 | 25 |
| 2 | 3 | 0 | 21 | 13 | 28 |
| 3 | 18 | 21 | 0 | 8 | 7 |
| 4 | 10 | 13 | 8 | 0 | 15 |
| 5 | 25 | 28 | 7 | 15 | 0 |

Analytics Vidhya
Learn everything about analytics

# Steps to perform Hierarchical Clustering



# Steps to perform Hierarchical Clustering

| Student_ID | Marks |
|---|---|
| 1 | 10 |
| 2 | 7 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| Student_ID | Marks |
|---|---|
| (1,2) | 10 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

| ID | (1,2) | 3 | 4 | 5 |
|---|---|---|---|---|
| **(1,2)** | 0 | 18 | 10 | 25 |
| **3** | 18 | 0 | 8 | 7 |
| **4** | 10 | 8 | 0 | 15 |
| **5** | 25 | 7 | 15 | 0 |

| Student_ID | Marks |
|---|---|
| (1,2) | 10 |
| 3 | 28 |
| 4 | 20 |
| 5 | 35 |

# Steps to perform Hierarchical Clustering

1. Assign each point as a different cluster

2. Merge the two closest clusters and update the proximity

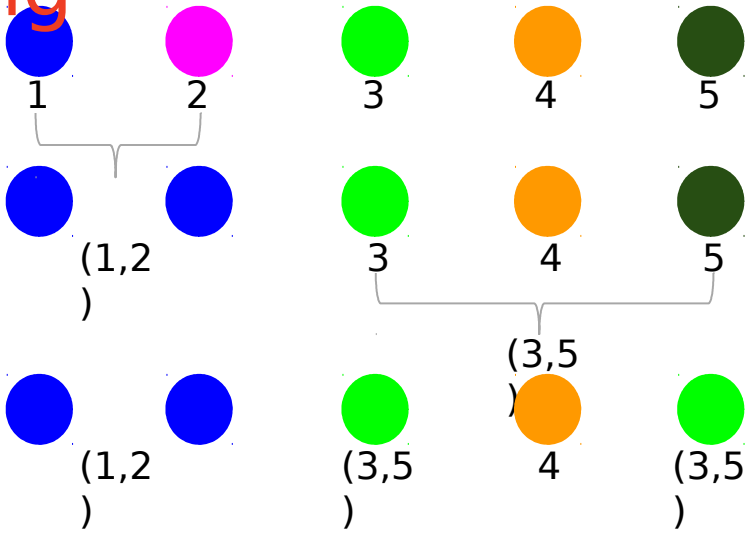
   matrix

3. Repeat step 2 until only a single cluster is left

# Steps to perform Hierarchical Clustering

| ID | (1,2) | 3 | 4 | 5 |
|---|---|---|---|---|
| **(1,2)** | 0 | 18 | 10 | 25 |
| **3** | 18 | 0 | 8 | 7 |
| **4** | 10 | 8 | 0 | 15 |
| **5** | 25 | 7 | 15 | 0 |

# Steps to perform Hierarchical Clustering



# Steps to perform Hierarchical Clustering

| Student_ID | Marks |
|---|---|
| (1,2) | 10 |
| (3,5) | 35 |
| 4 | 20 |

# Steps to perform Hierarchical Clustering

| ID | (1, 2) | (3, 5) | 4 |
|---|---|---|---|
| **(1, 2)** | 0 | 25 | 10 |
| **(3, 5)** | 25 | 0 | 15 |
| **4** | 10 | 15 | 0 |

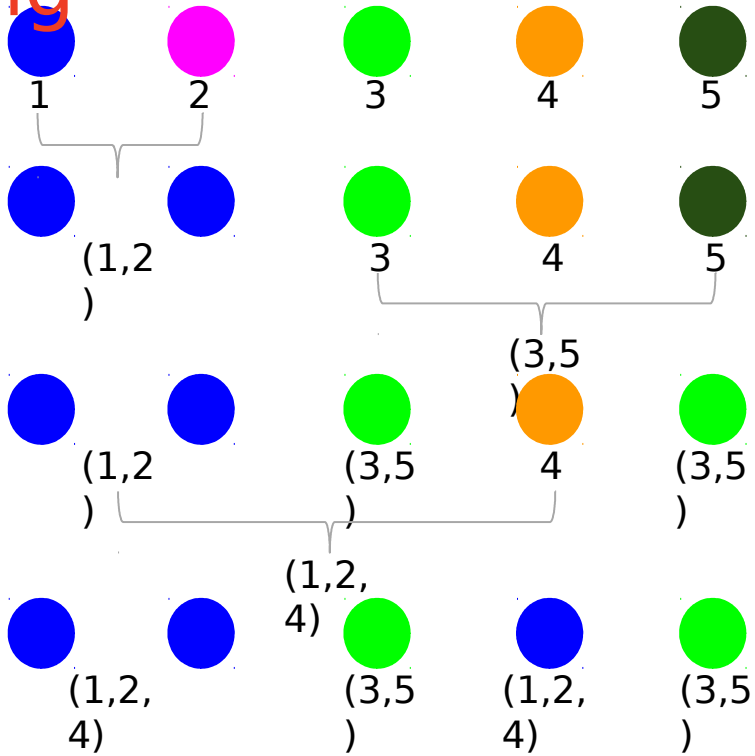
| Student_ID | Marks |
|---|---|
| (1,2) | 10 |
| (3,5) | 35 |
| 4 | 20 |

# Steps to perform Hierarchical Clustering

| ID | (1, 2) | (3, 5) | 4 |
|---|---|---|---|
| (1, 2) | 0 | 25 | 10 |
| (3, 5) | 25 | 0 | 15 |
| 4 | 10 | 15 | 0 |

| Student_ID | Marks |
|---|---|
| (1,2) | 10 |
| (3,5) | 35 |
| 4 | 20 |

# Steps to perform Hierarchical Clustering



# Steps to perform Hierarchical Clustering

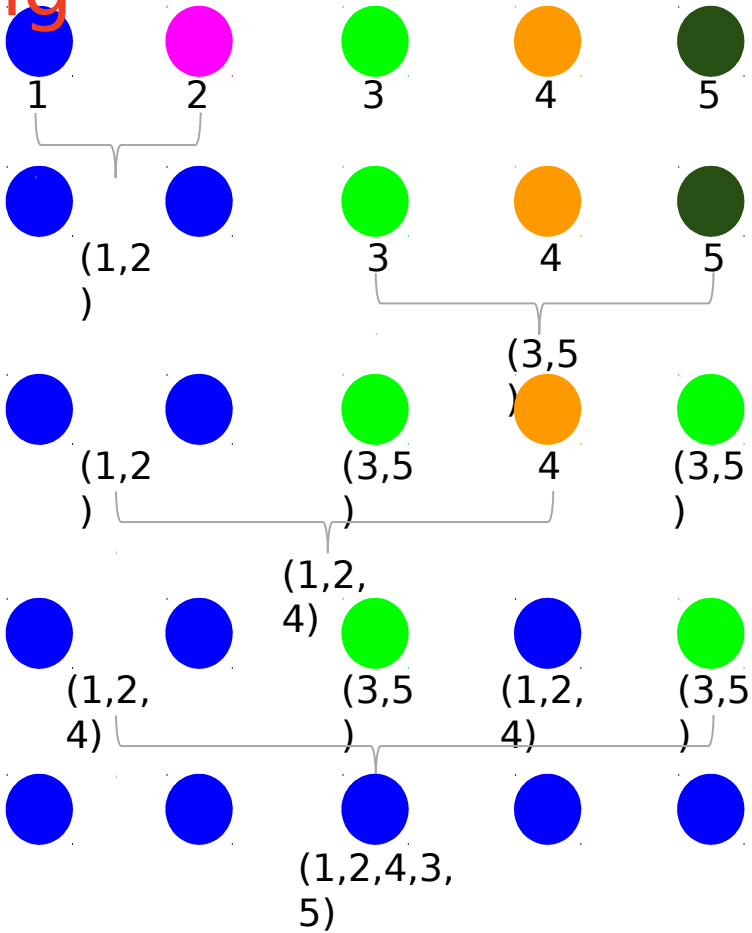
| Student_ID | Marks |
|---|---|
| (1,2,4) | 20 |
| (3,5) | 35 |

# Steps to perform Hierarchical Clustering

| ID | (1,2,4) | (3,5) |
|---|---|---|
| (1,2,4) | 0 | 15 |
| (3,5) | 15 | 0 |

| Student_ID | Marks |
|---|---|
| (1,2,4) | 20 |
| (3,5) | 35 |

# Steps to perform Hierarchical Clustering



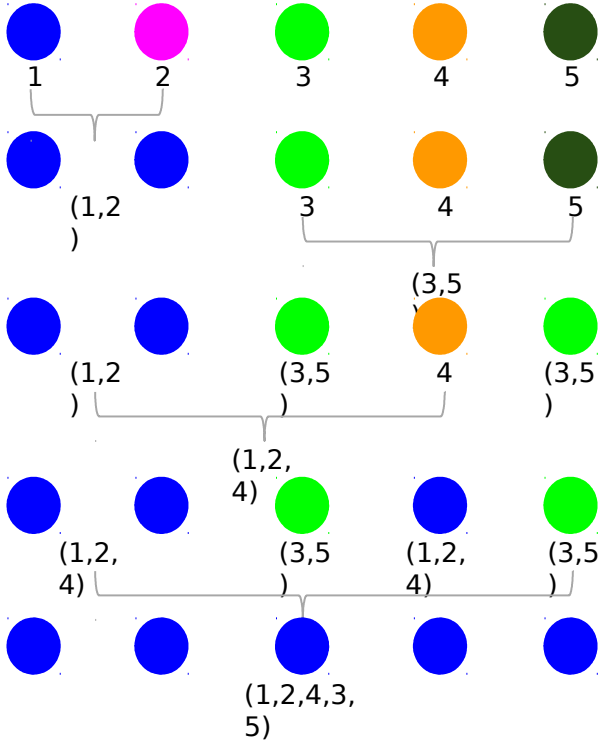

Analytics Vidhya
Learn everything about analytics

# How to decide the number of clusters?

- Dendrogram

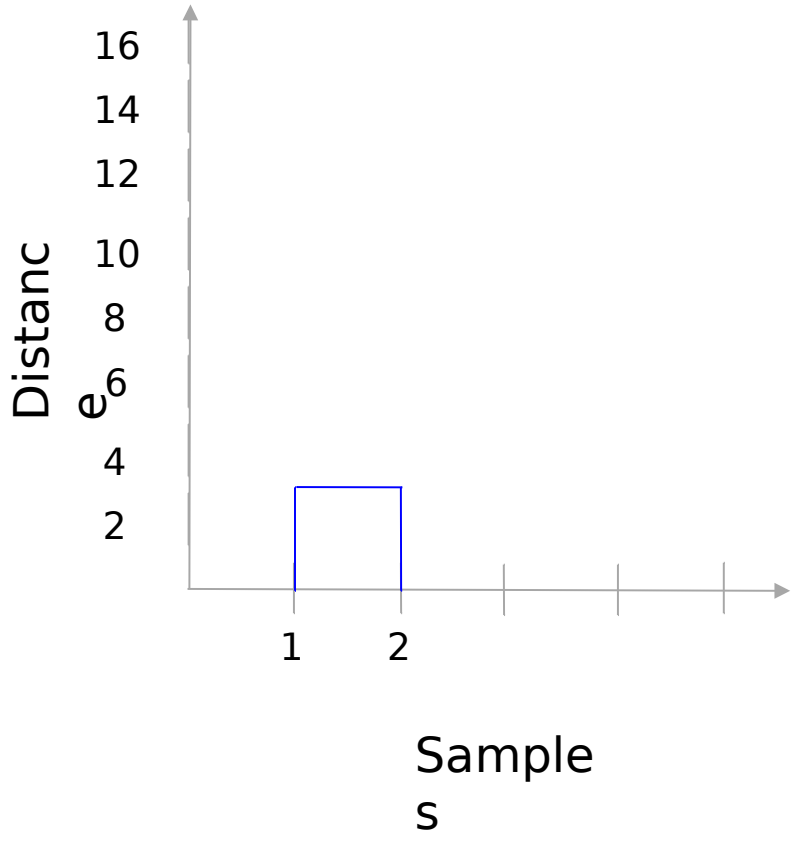
  A dendrogram is a tree like diagram that records the sequences of merges or splits.

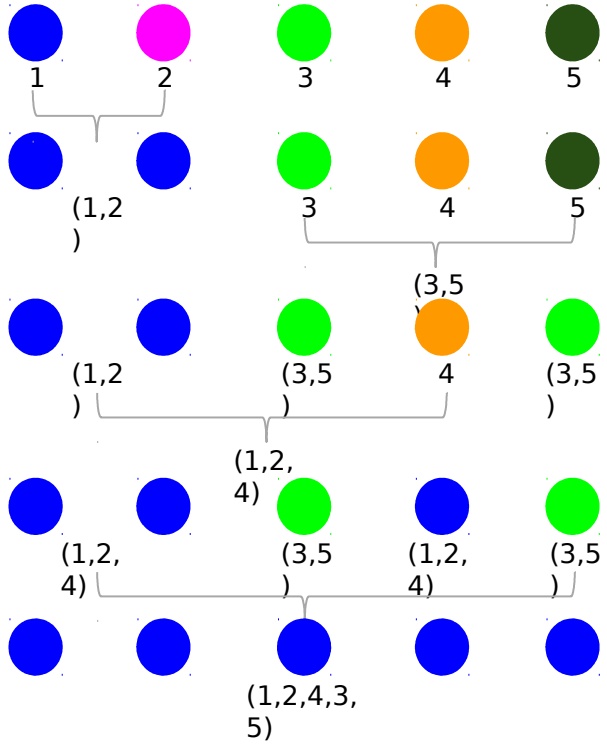
# How to decide the number of clusters?




Analytics Vidhya
Learn everything about analytics

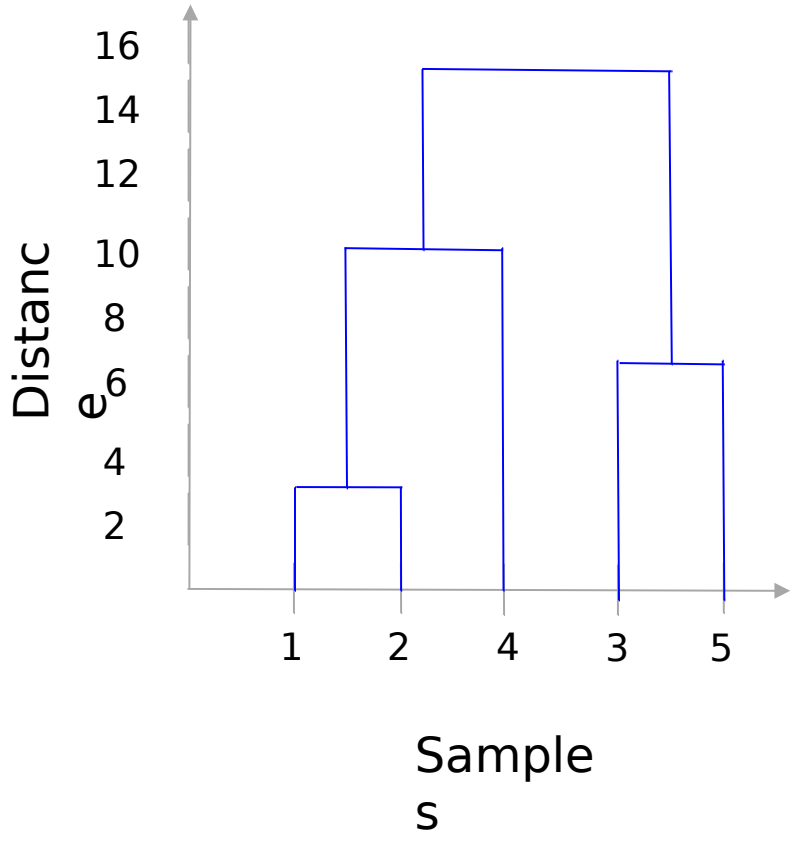
# How to decide the number of clusters?

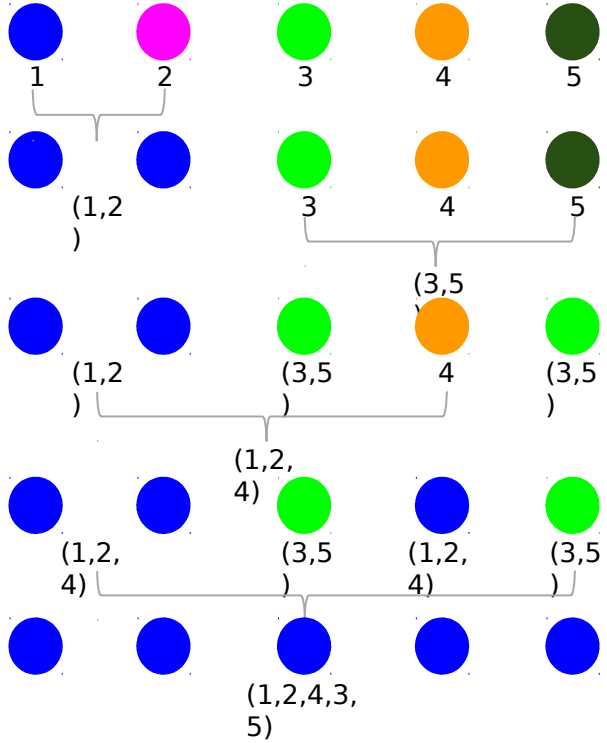



Analytics Vidhya
Learn everything about analytics

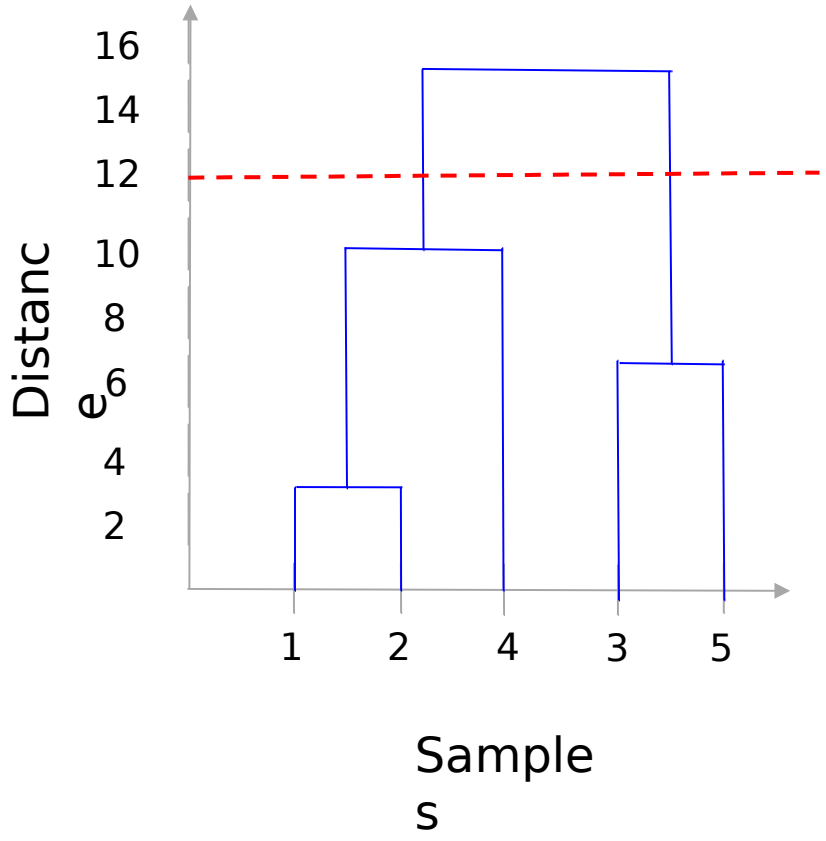
# How to decide the number of clusters?




Analytics Vidhya
Learn everything about analytics

# How to decide the number of clusters?




Number of clusters =

2

Clusters = (1,2,4) &

(3,5)

Analytics Vidhya
Learn everything about analytics

# Challenges with Hierarchical Clustering

- Once a cluster is made, it cannot be undone
- More time and space complexity

Thank
You!

Analytics Vidhya
Learn everything about analytics