



# Hierarchical Clustering

Diptangshu Banik

Twitter - @dipbanik



- Two kinds of Hierarchical Clustering –
  - Agglomerative : Starting at the bottom and building our way up.
  - Divisive : Starting at the top and dividing it into multiple clusters.



# Agglomerative HC

- Step 1 : Make each data point a single-point cluster -> forms  $N$  clusters
- Step 2 : Take two closest data points and make them one cluster -> forms  $N-1$  clusters
- Step 3 : Take the two closest clusters and make them one clusters -> forms  $N-2$  clusters
- Step 4: Repeat Step 3 until there is only one cluster.

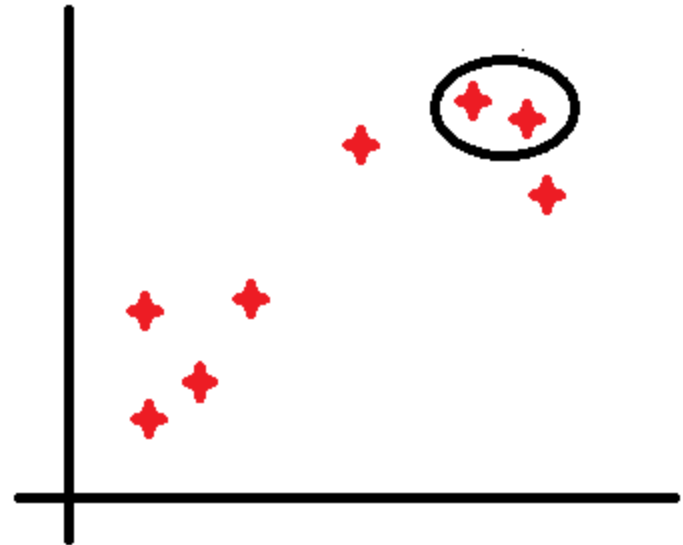
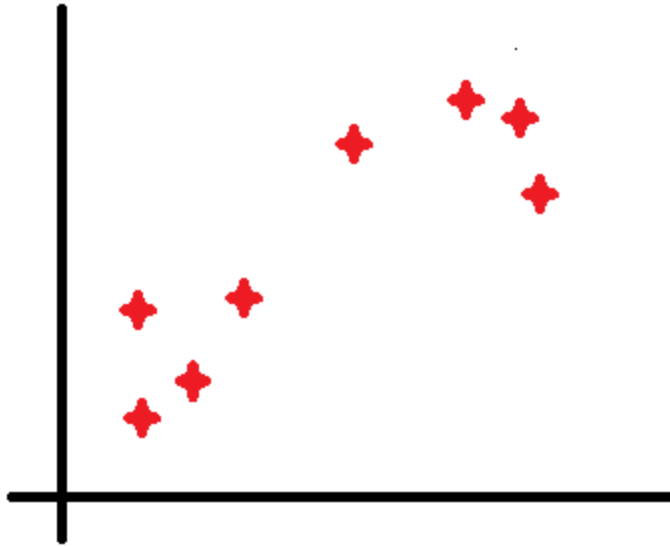


# Calculating distance

- Methods to calculate the distance between two clusters-
  - Option 1 : Closest points
  - Option 2 : Furthest Points
  - Option 3 : Average Distance
  - Option 4 : Distance between centroids

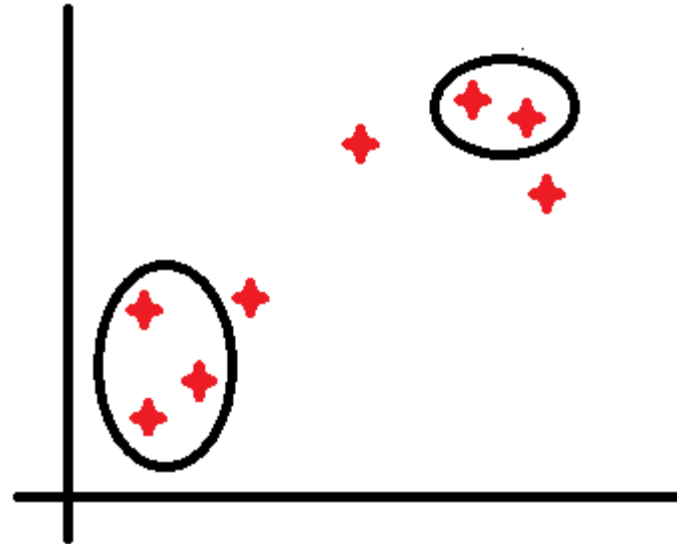
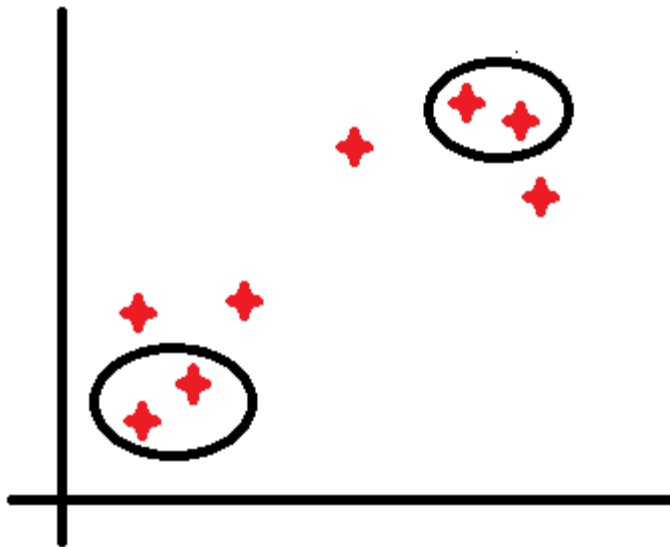


- Every point is one single cluster
- Find the closest clusters and make them one.



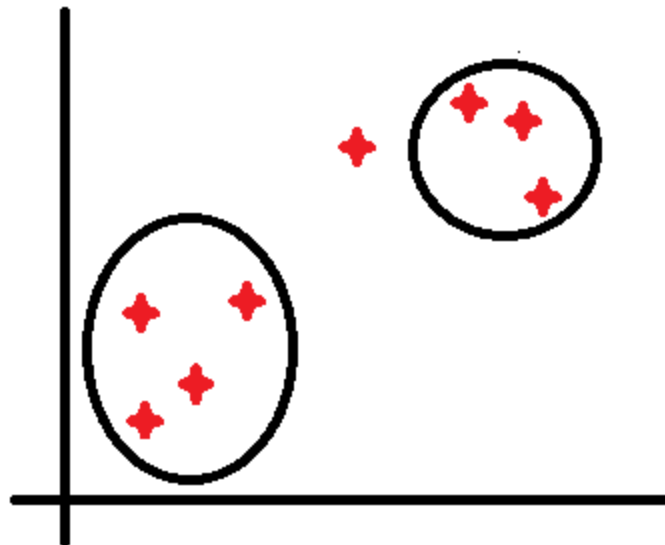
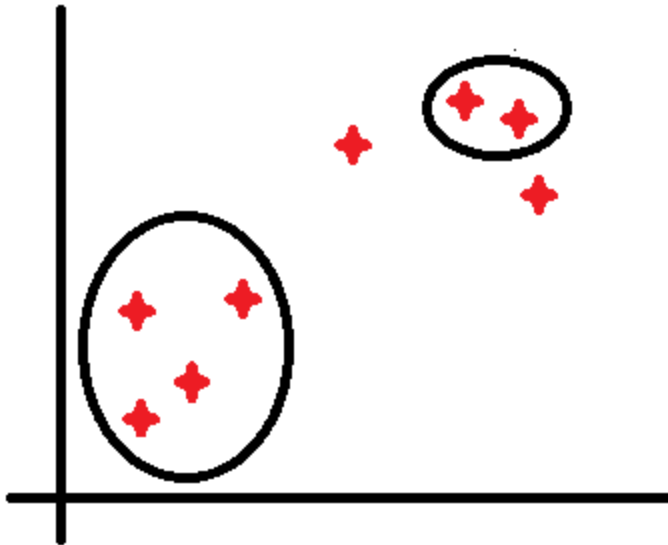


- Step 3



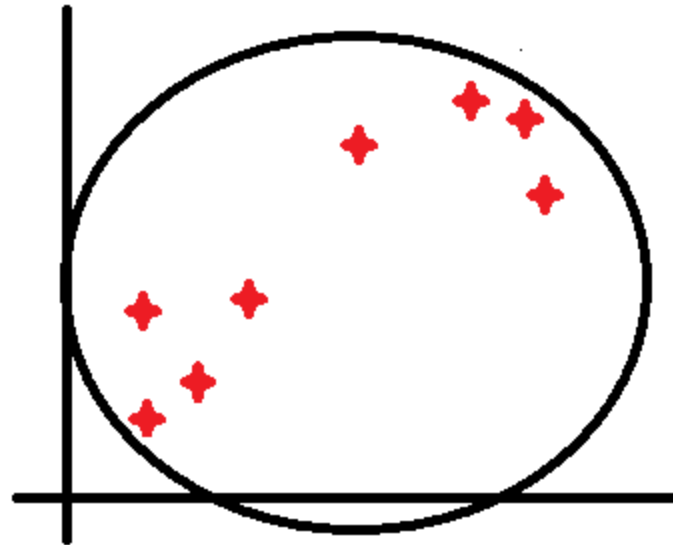
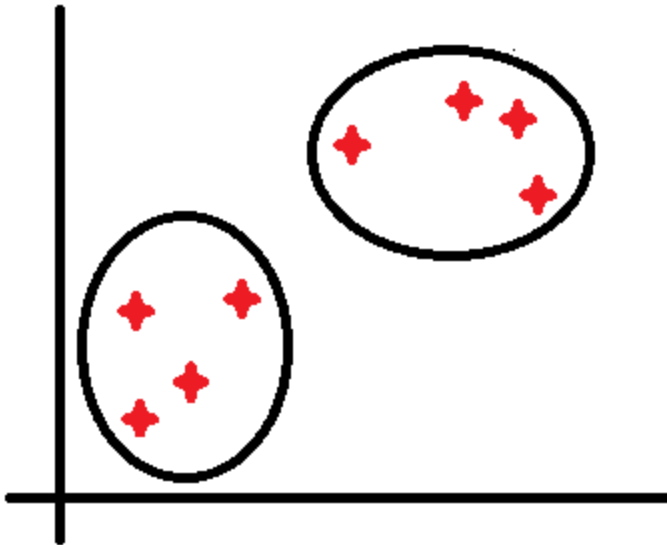


- Repeat Step 3 until you have only one cluster.





- Repeat Step 3 until you have only one cluster.

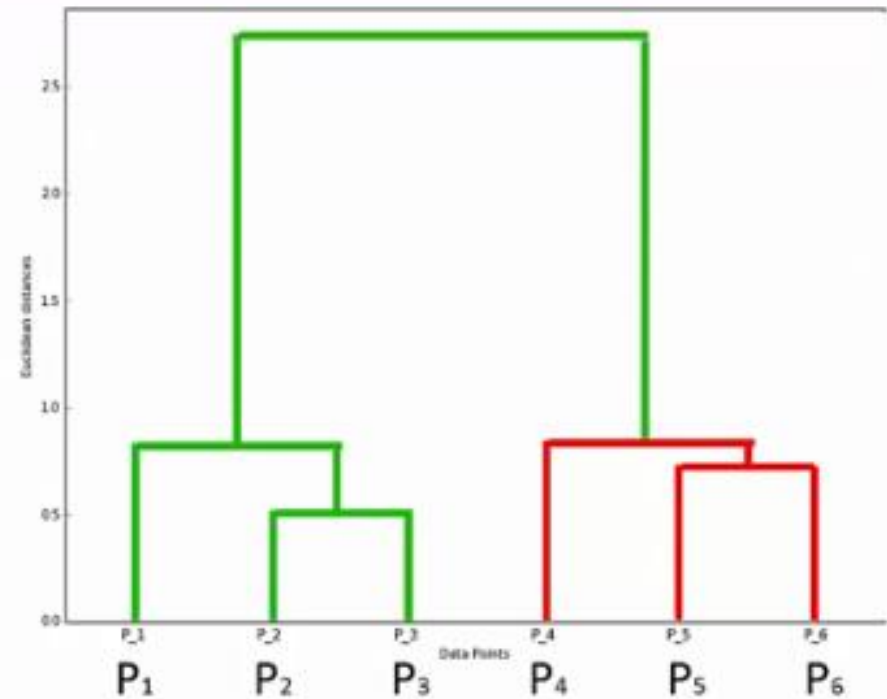
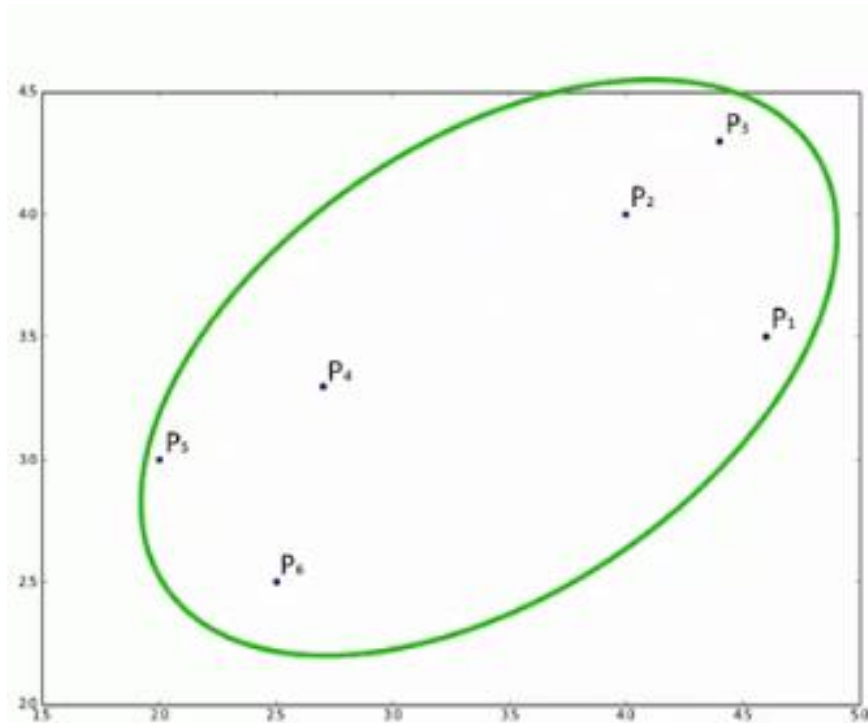






# Dendograms

- Height denotes the distance

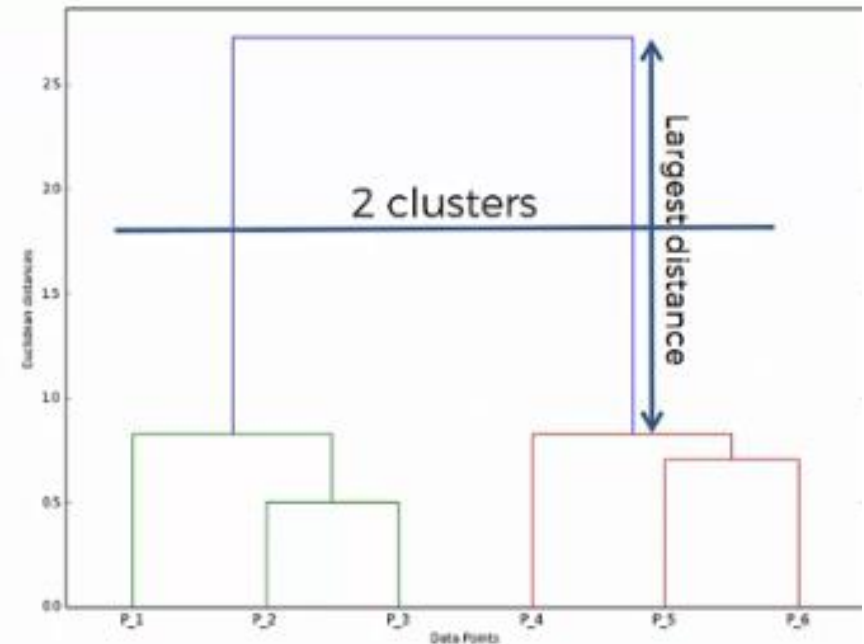
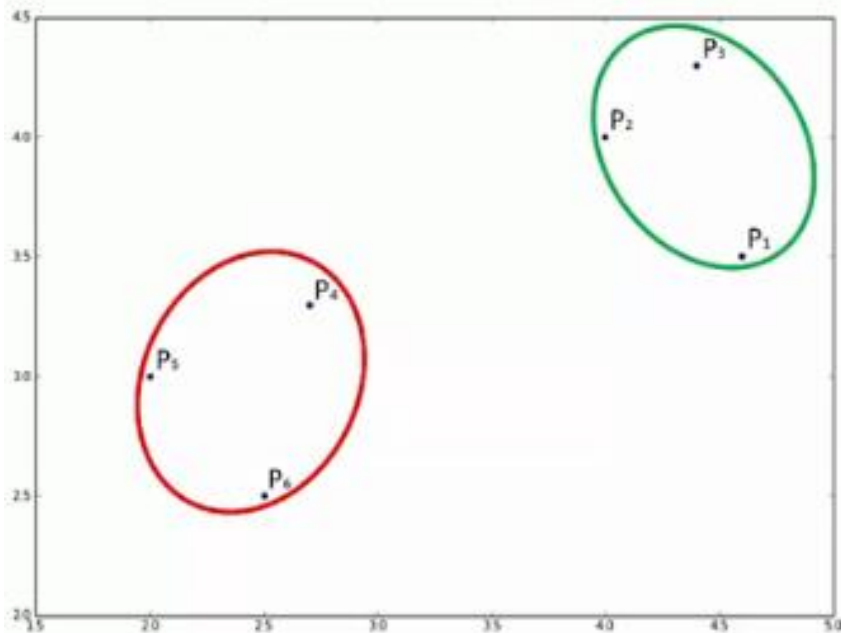




# Dendograms

## Optimal Clusters

- Longest vertical line that does not cross any extended horizontal line.





# Dendograms

## Optimal Clusters

- Count the number of vertical lines at that point to find the optimal number of clusters.

