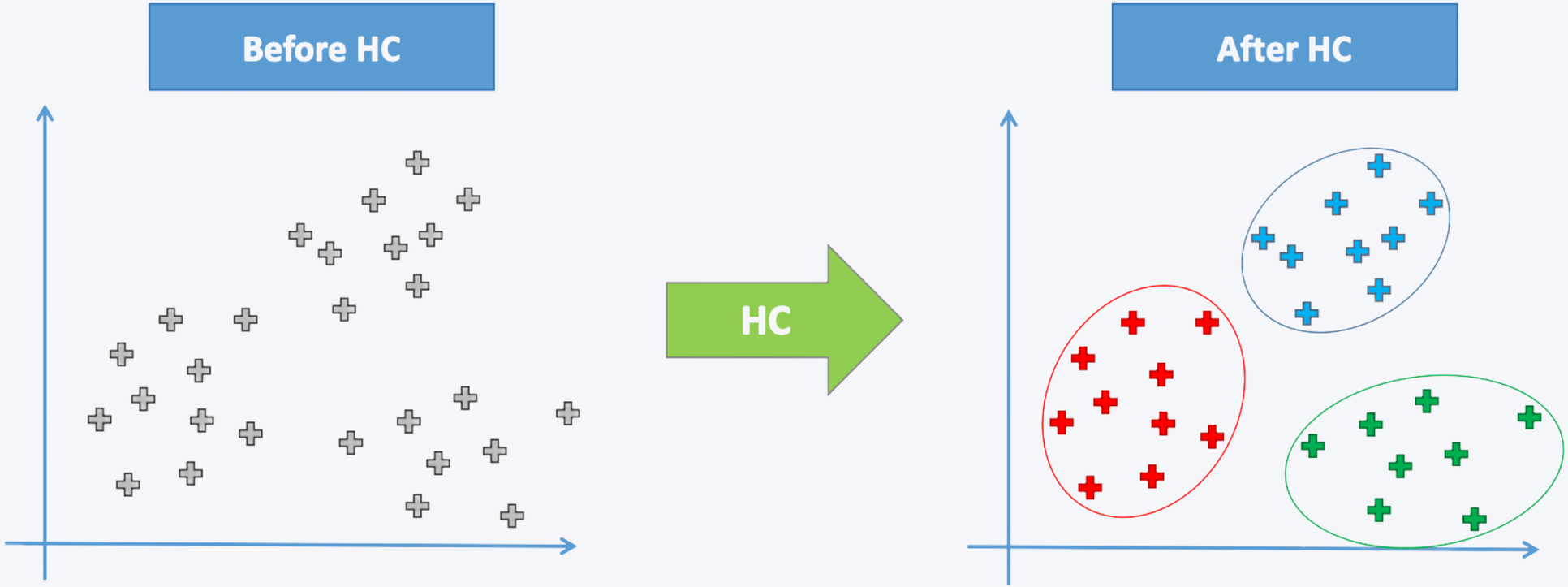
Hierarchical CLustering

Hierarchical CLustering

أكاديمية طويق
Tuwaiq Academy

What HC does ?



Same as K-Means but different process

Image source: mdpi.com/2073-8994/10/12/73

Hierarchical CLustering

أكاديمية طويق
Tuwaiq Academy

Two types of HC:

Agglomerative & Divisive

Bottom-Up Approach

Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

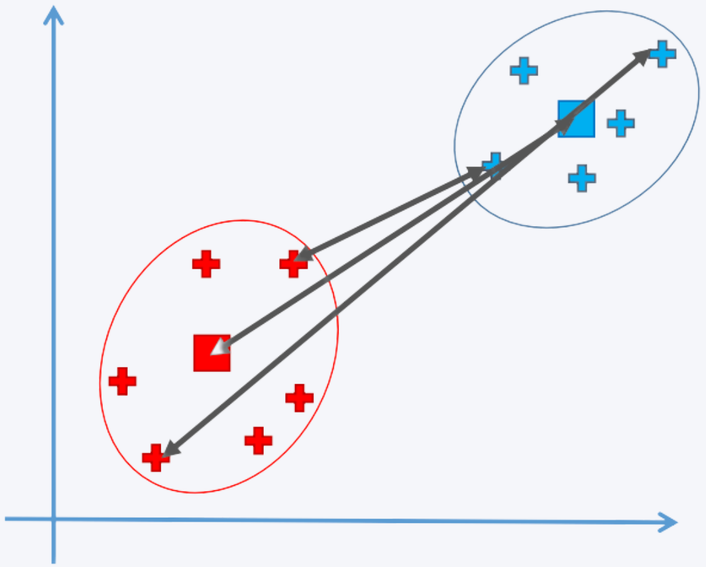
- **STEP 1:** Make each data point a single-point cluster -> That forms N clusters
- **STEP 2:** Take the two closest data points and make them **one cluster** -> That forms N-1 clusters
- **STEP 3**: Take the two **closest clusters** and make them one cluster -> That forms N - 2 clusters
- **STEP 4:**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

Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

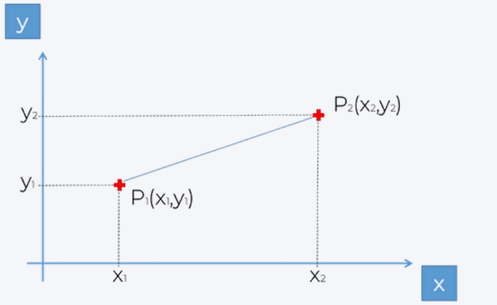
أكاديمية طويق
Tuwaiq Academy

Agglomerative HC



Distance Between Two Clusters:

- Option 1: Closest Points
- Option 2: Furthest Points
- Option 3: Average Distance
- Option 4: Distance Between Centroids



Euclidean Distance between P_1 and $P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

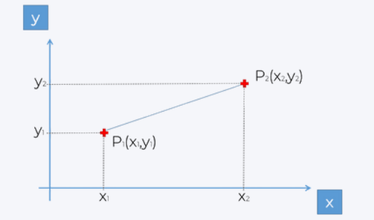
- Euclidean Distance is basically more natural

- **What is actually Distance between clusters ?** shortest ? largest ? average ? or distance between centroids ?
- You need to find that based on business usecase !! choosing distance is **important** !!

Image source: mdpi.com/2073-8994/10/12/73

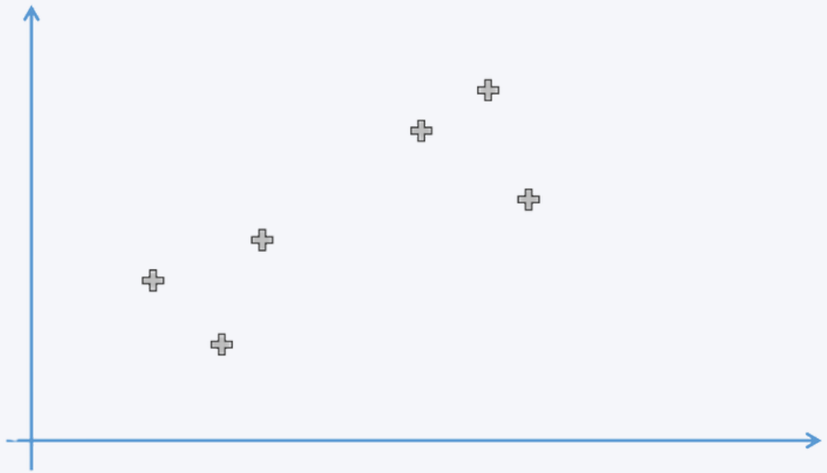
Hierarchical Clustering

Agglomerative HC



Euclidean Distance between P_1 and $P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Consider the following dataset of N = 6 data points



STEP 1: Make each data point a single-point cluster ➞ That forms 6 clusters

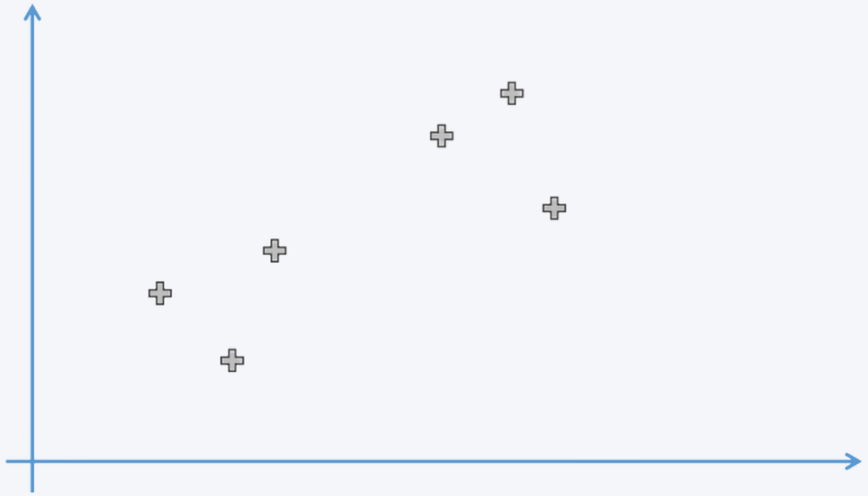


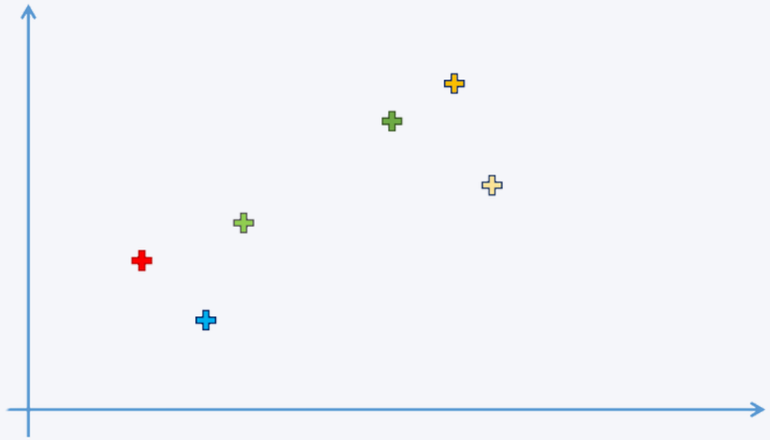
Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

STEP 1: Make each data point a single-point cluster ⟶ That forms 6 clusters



STEP 2: Take the two closest data points and make them one cluster
⟶ That forms 5 clusters

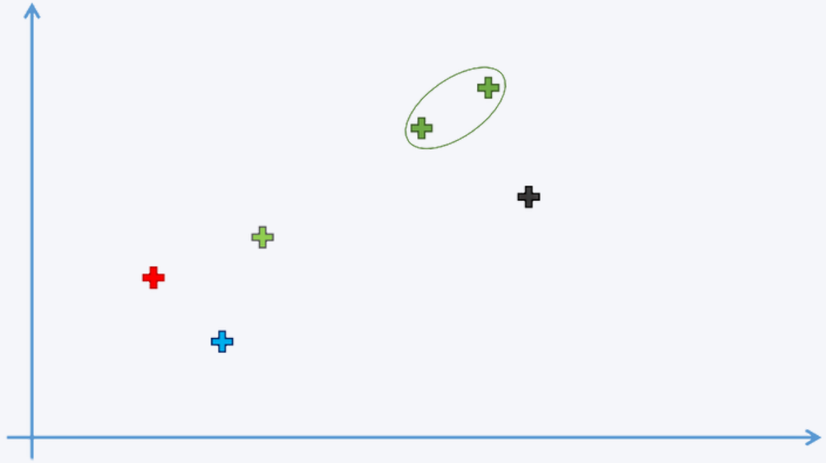


Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

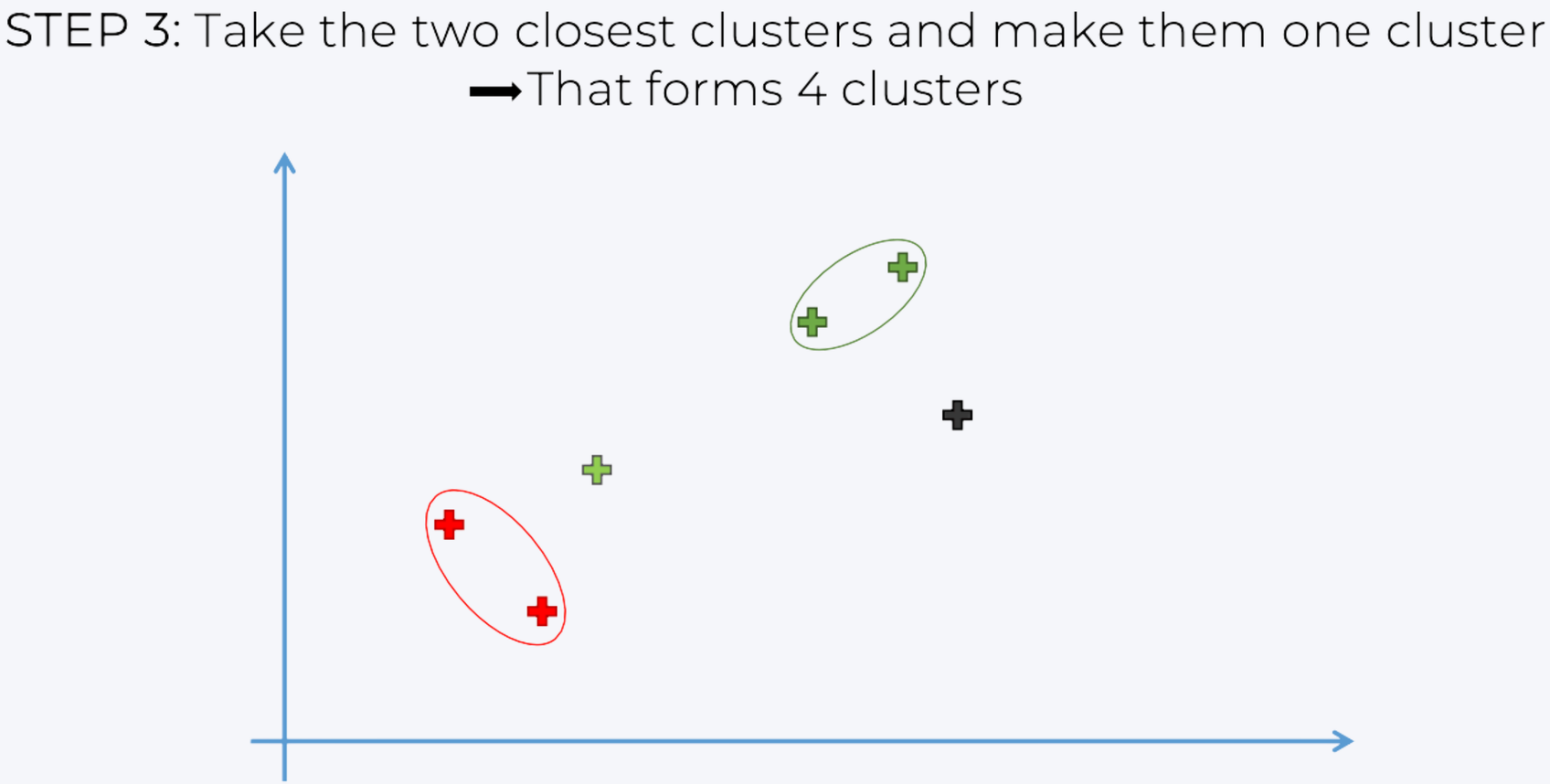


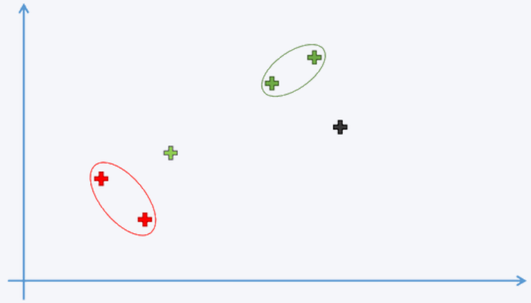
Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

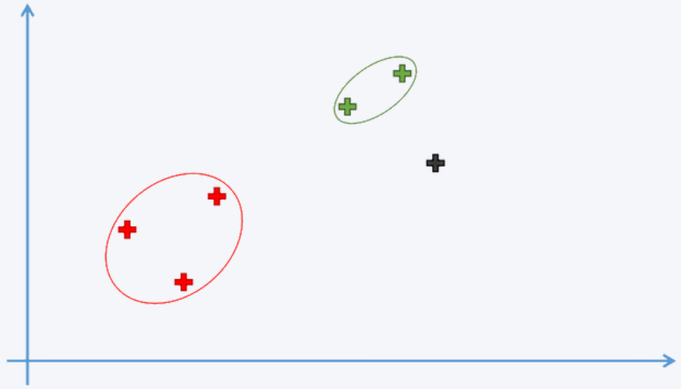
أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

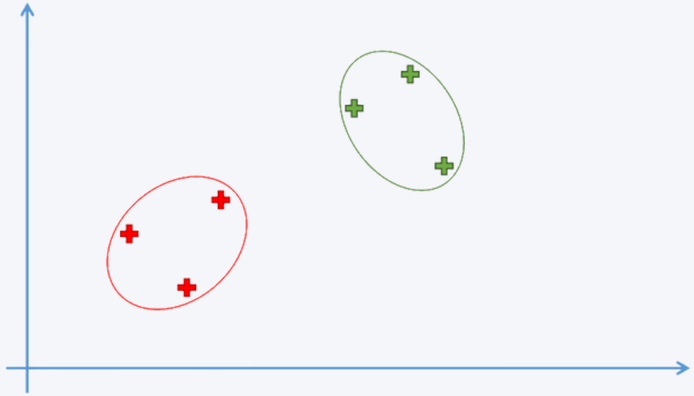
STEP 3: Take the two closest clusters and make them one cluster
➞That forms 4 clusters



STEP 4: Repeat STEP 3 until there is only one cluster



STEP 4: Repeat STEP 3 until there is only one cluster



STEP 4: Repeat STEP 3 until there is only one cluster

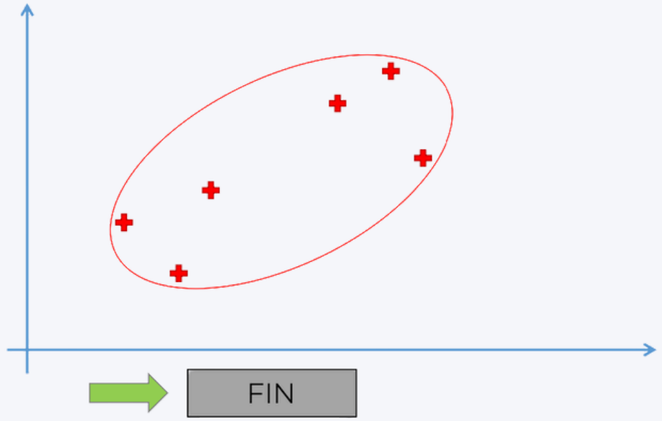


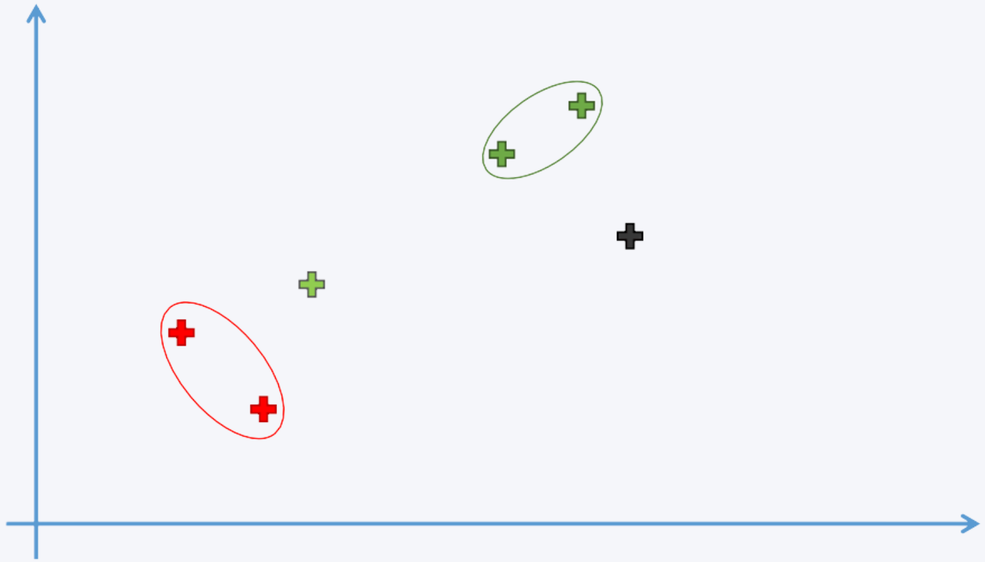
Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

STEP 3: Take the two closest clusters and make them one cluster
➞That forms 4 clusters



STEP 4: Repeat STEP 3 until there is only one cluster

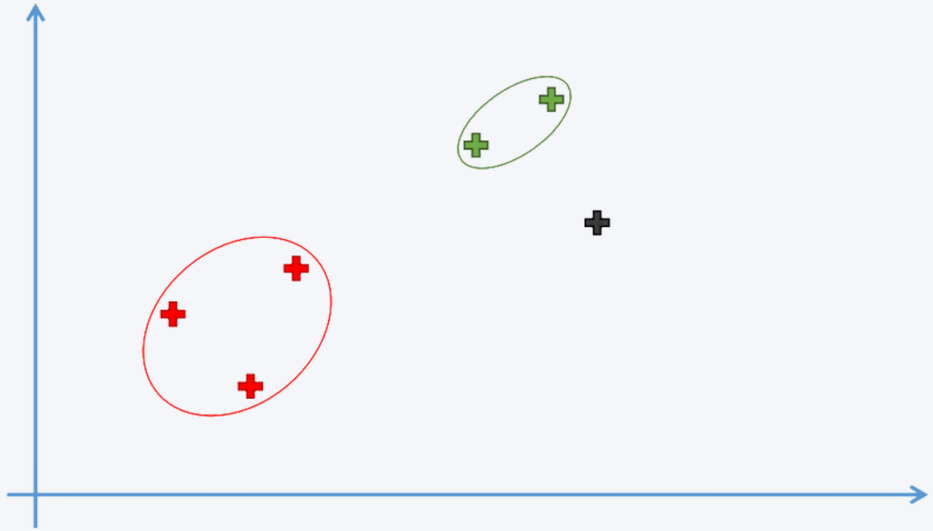


Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

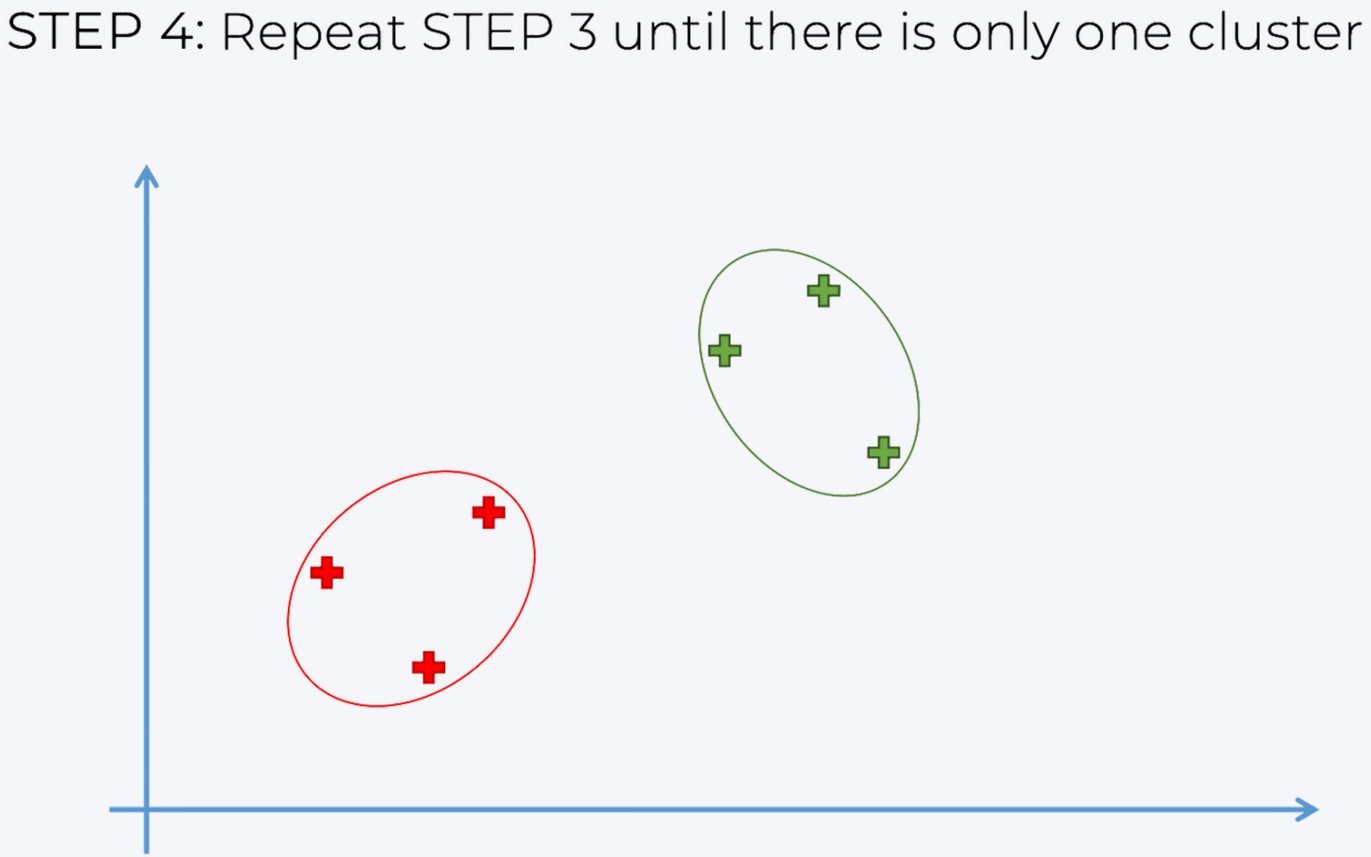
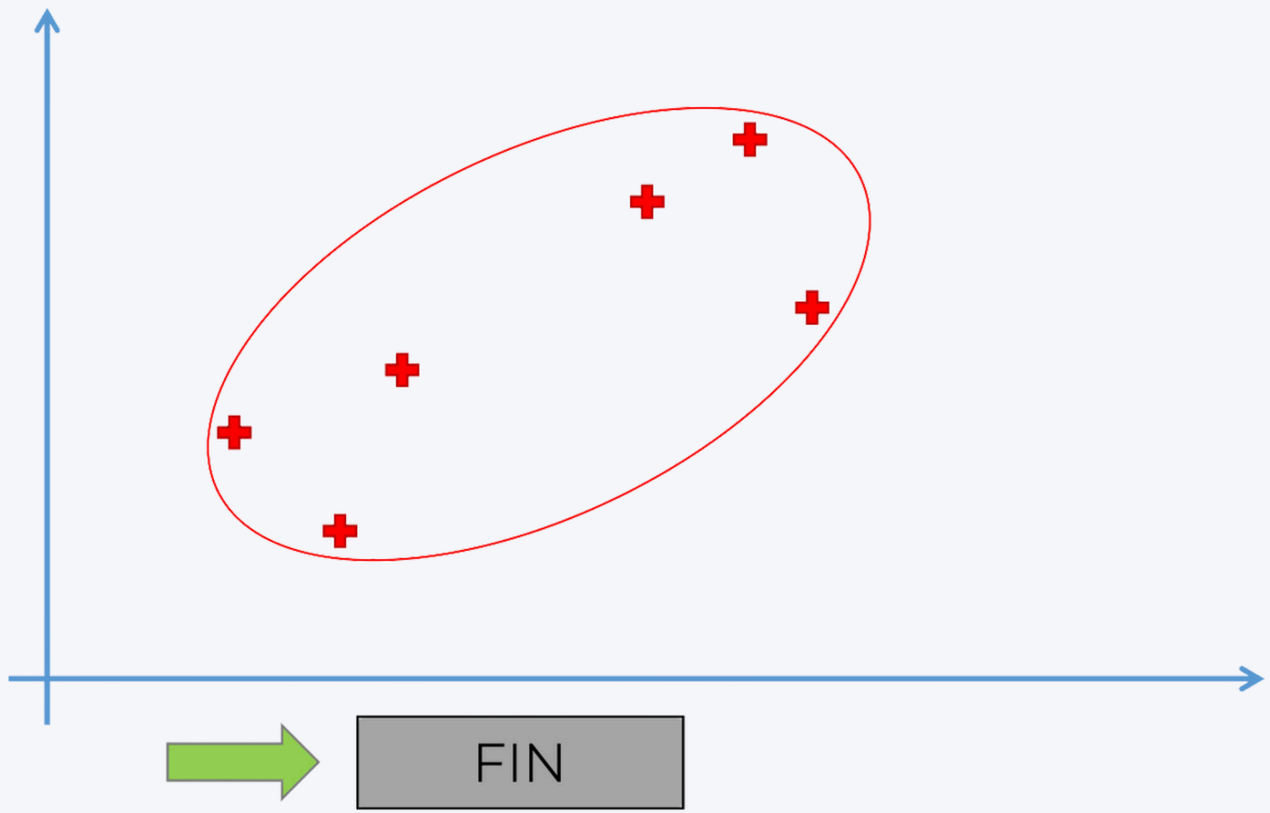


Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Agglomerative HC

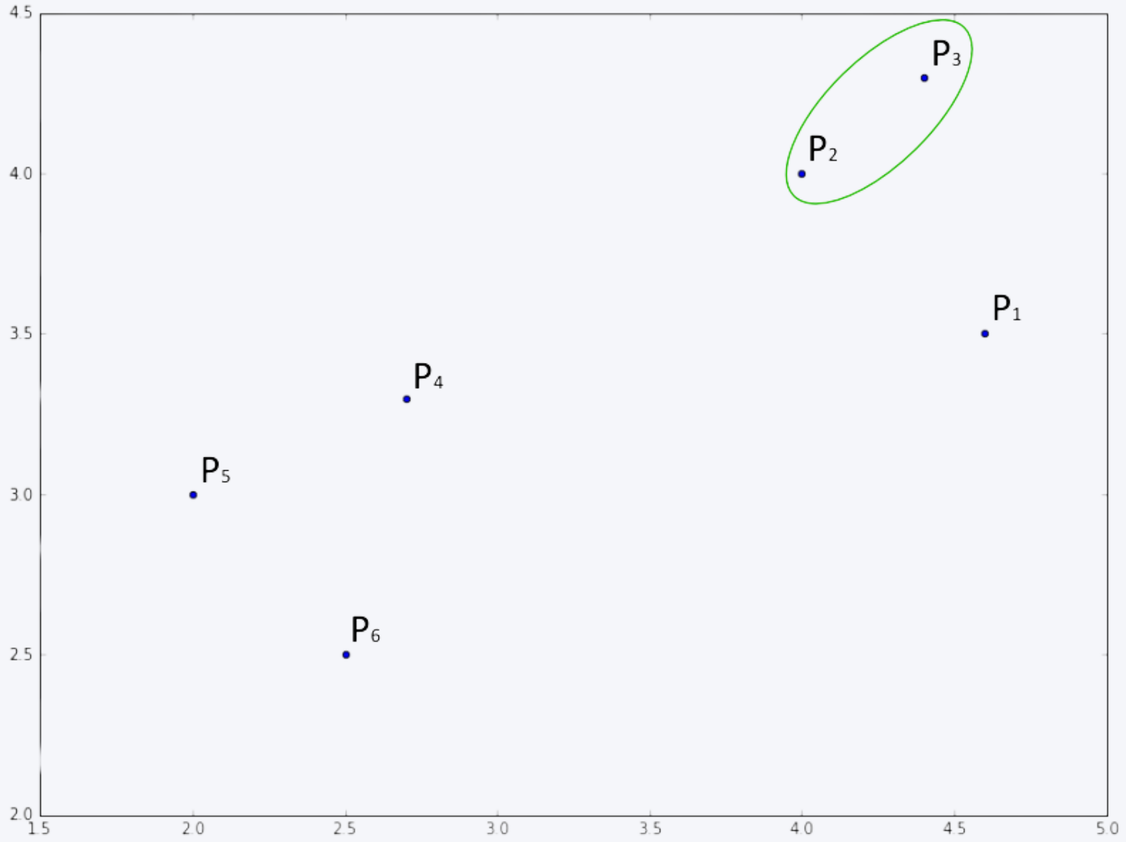


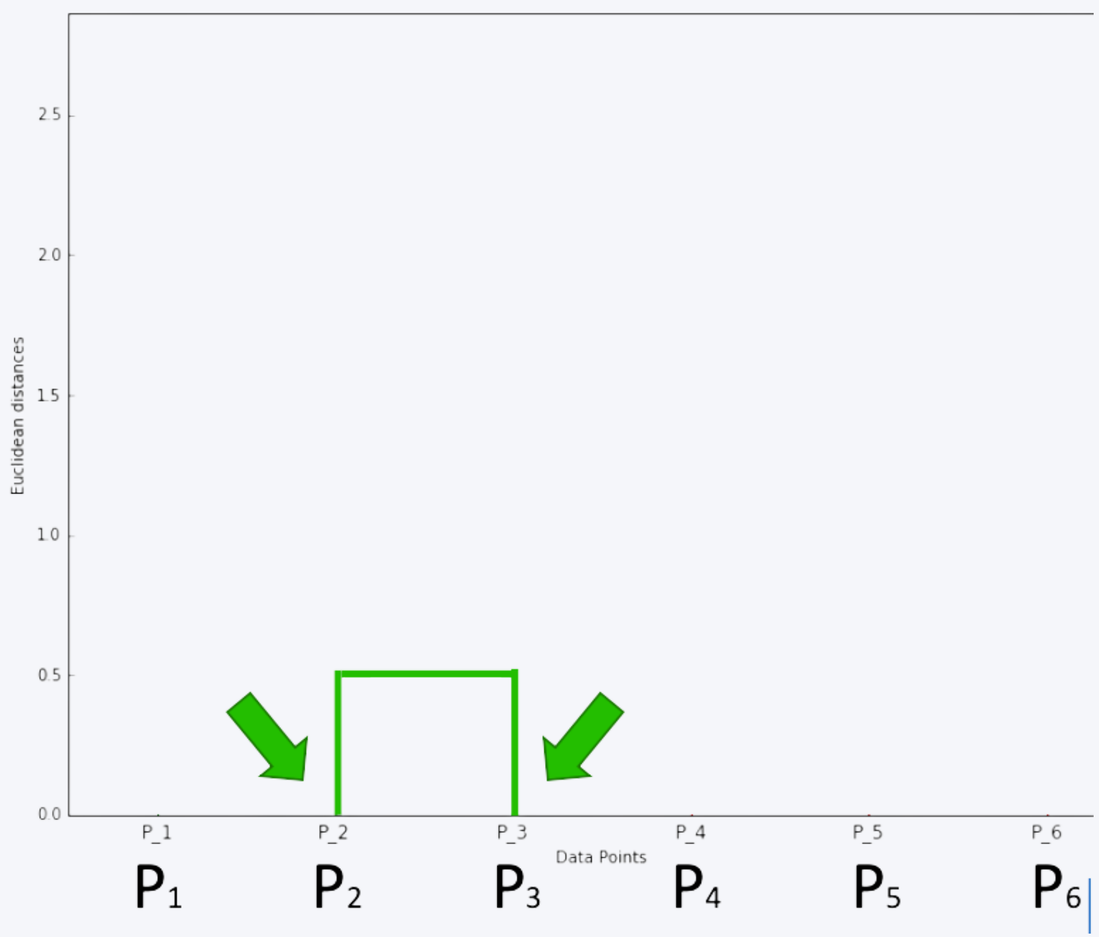
- it maintains like memory of each steps , saving the steps taken with the distances.

Image source: mdpi.com/2073-8994/10/12/73

Hierarchical Clustering

How Do Dendograms Work ?

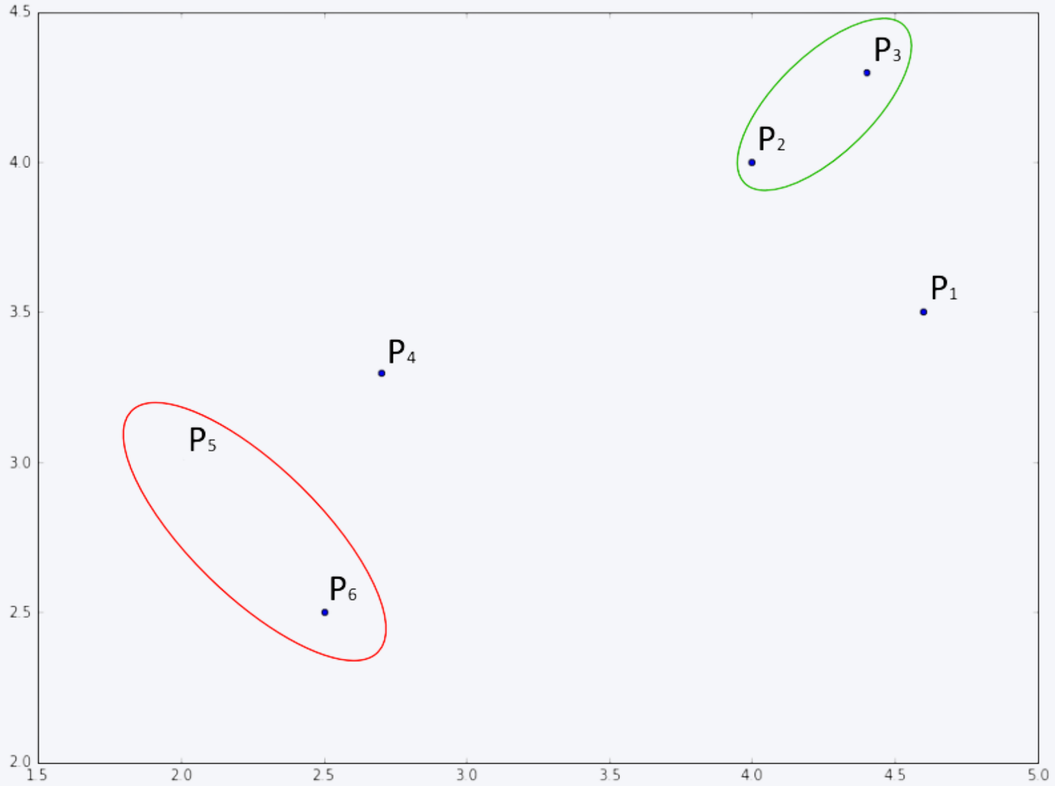


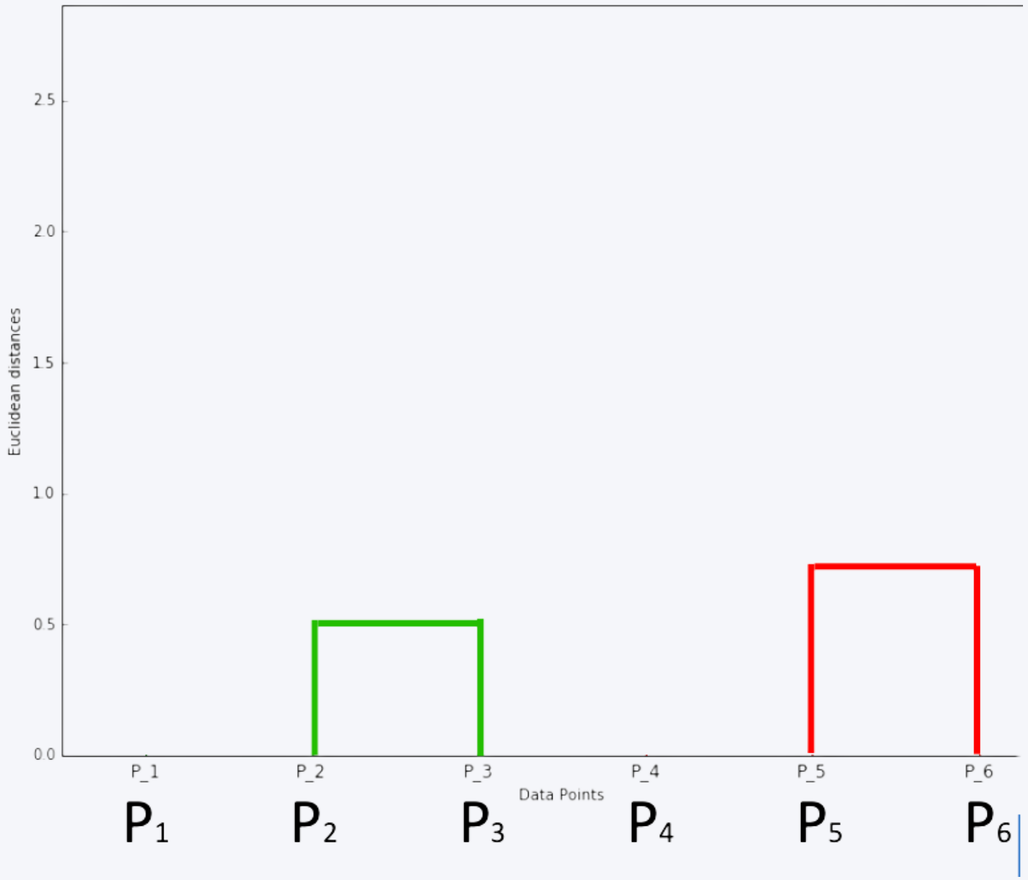


- once we defied first closest two points and made clusters, we can now assign p2 and p3 with the distance (Euclidean), and it is the similarity of these clusters

Hierarchical Clustering

How Do Dendograms Work ?

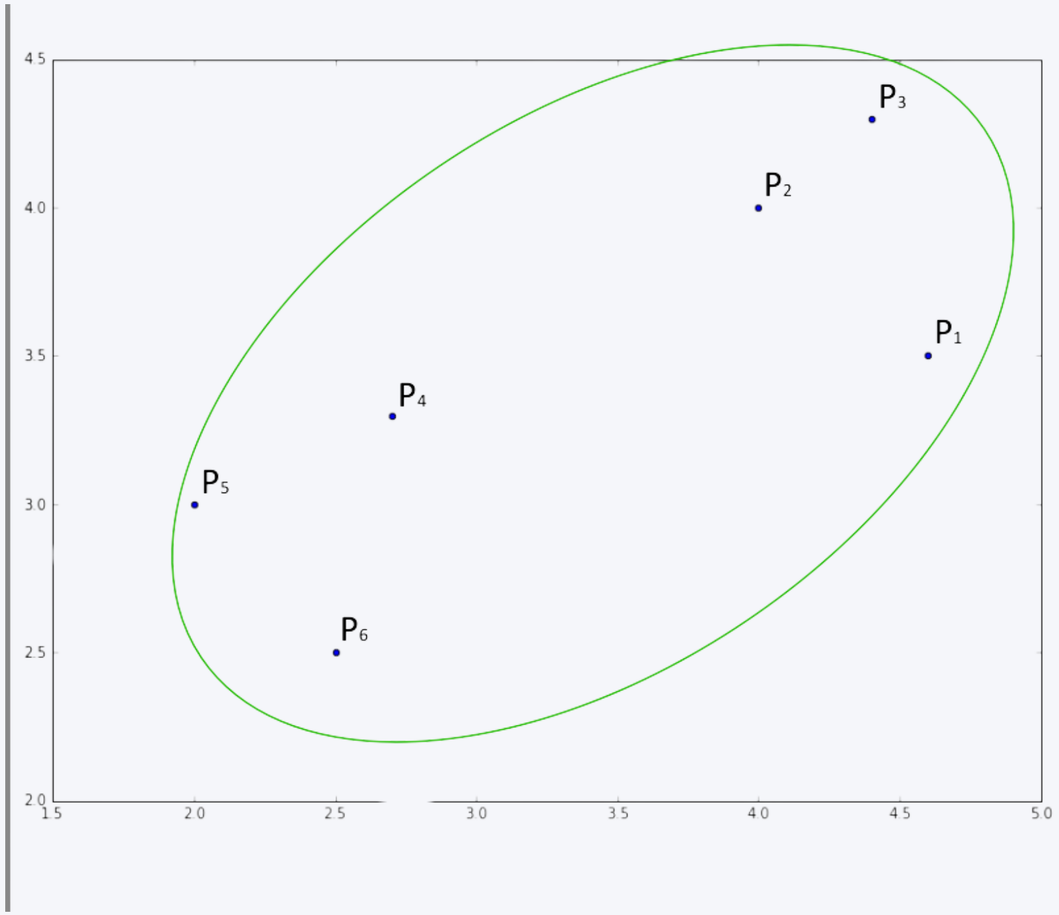


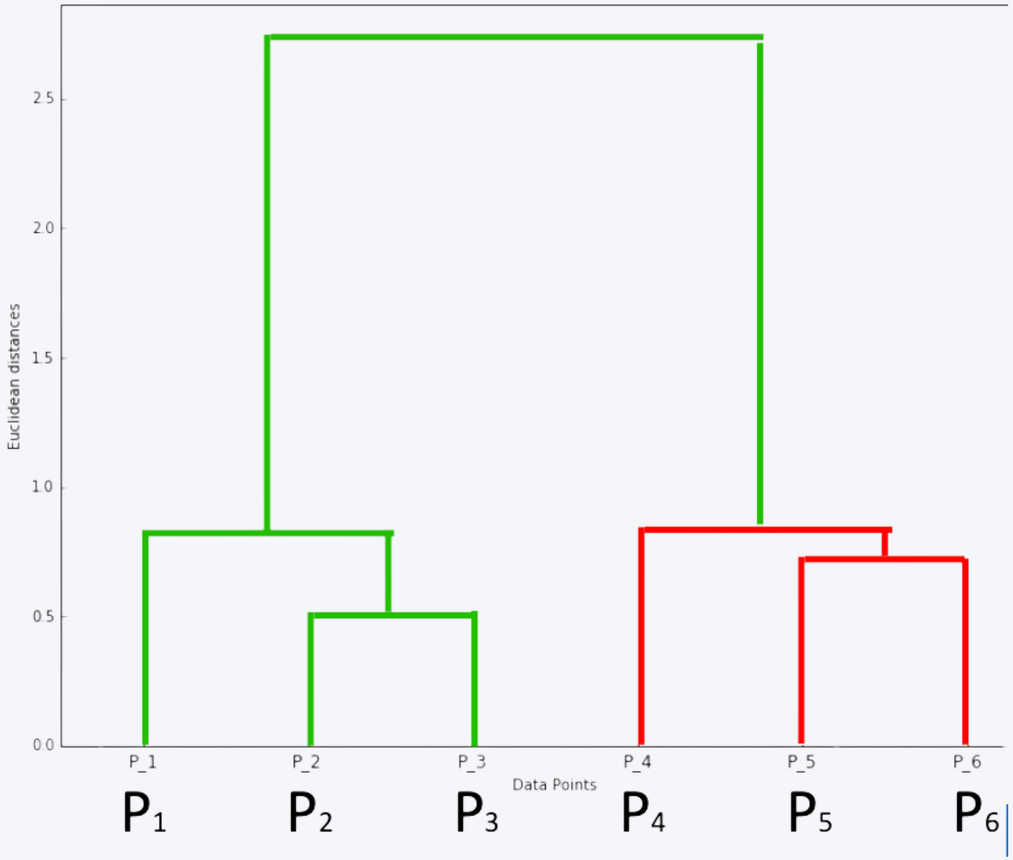


- Next step , the next two closest clusters and plot them as Dendograms

Hierarchical Clustering

How Do Dendograms Work ?

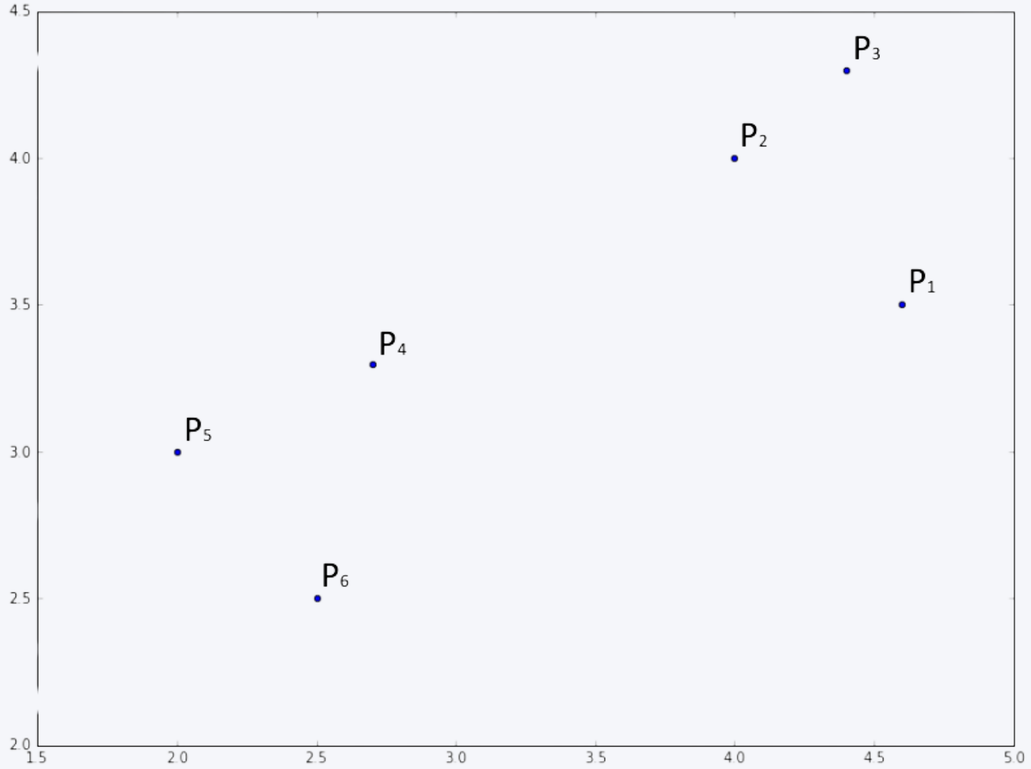


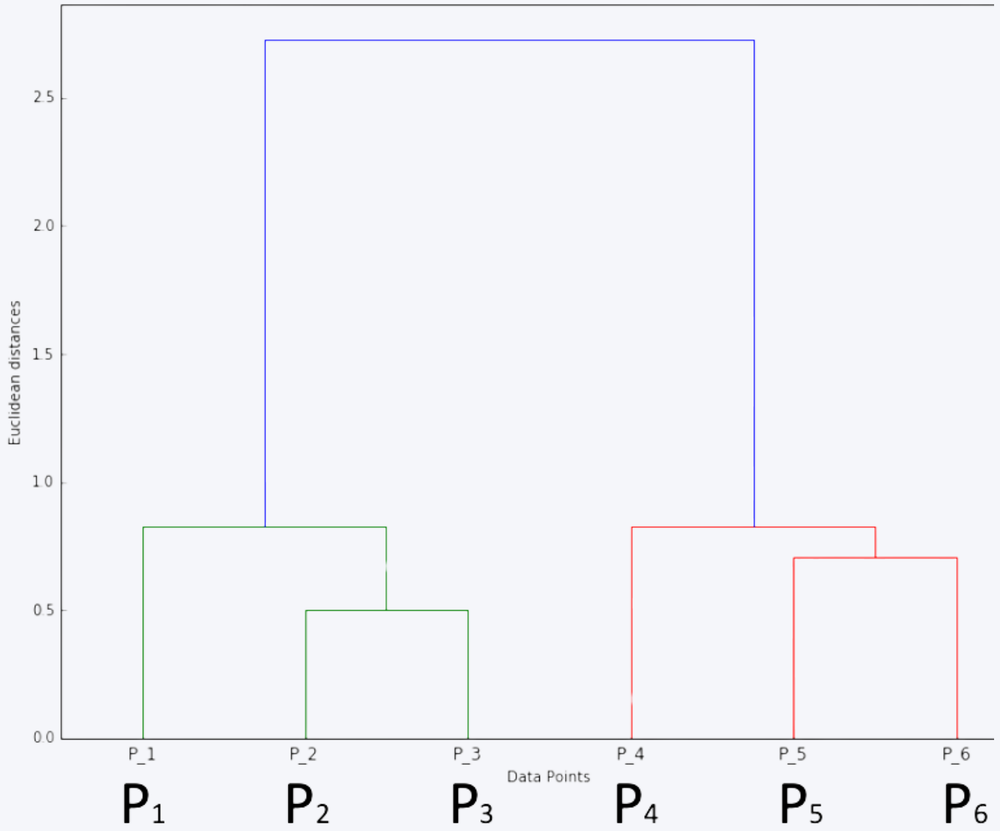


- Repeat the process until you get one cluster

Hierarchical Clustering

How Do Dendograms Work ?



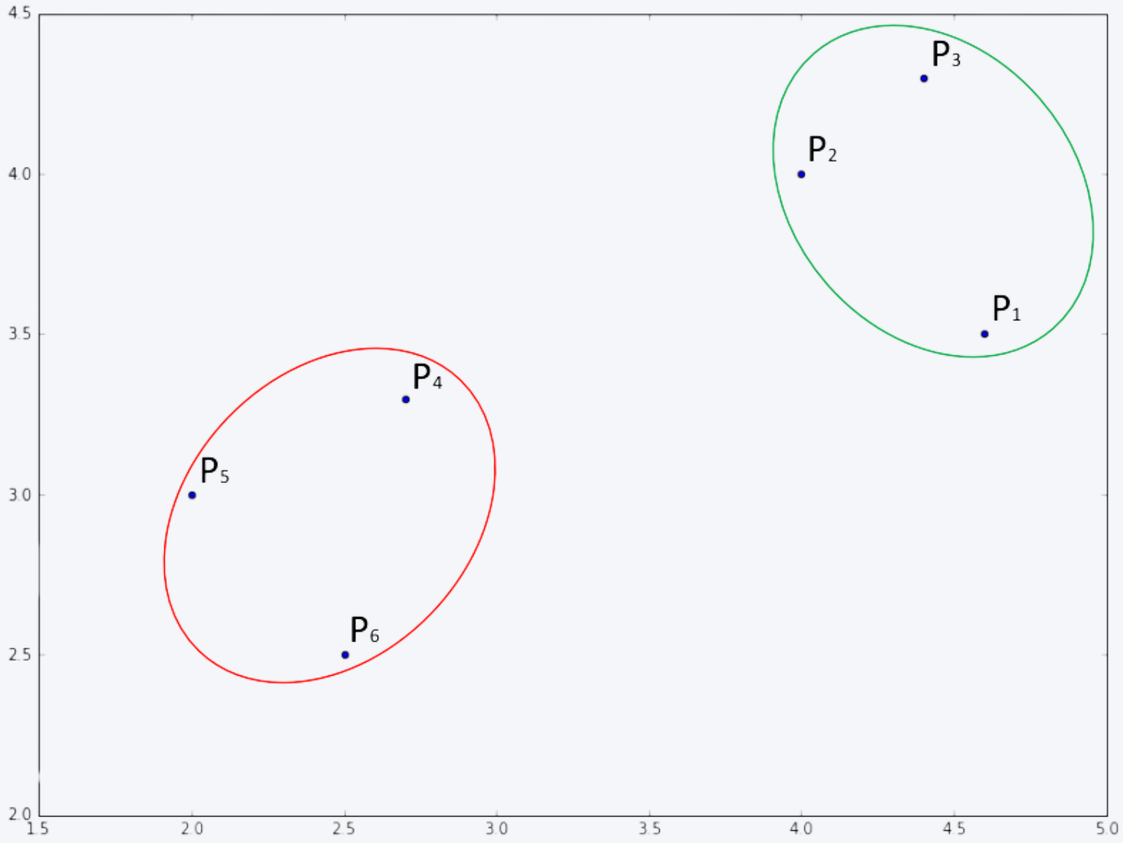


- Final Dendograms with clusters and distances (Memory on how clusters are formed)

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

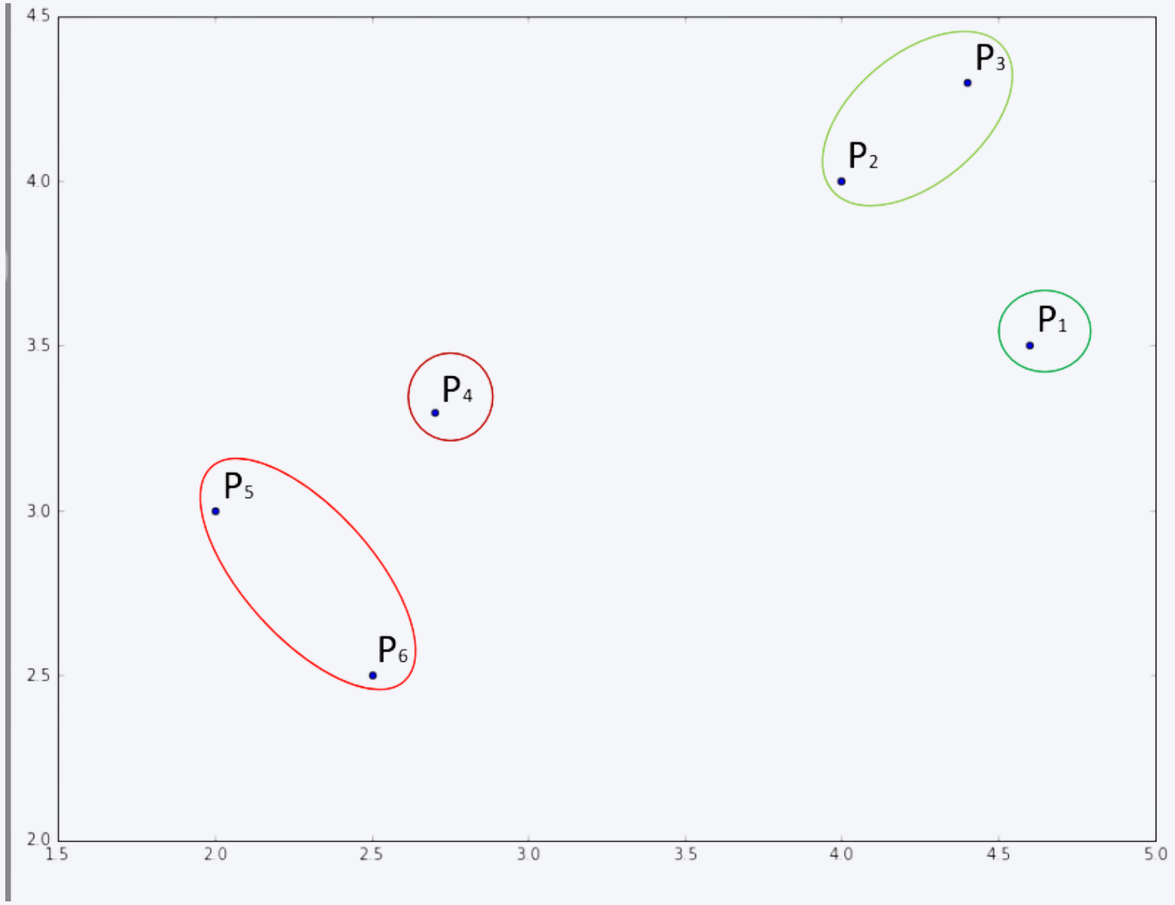
Using Dendograms ?

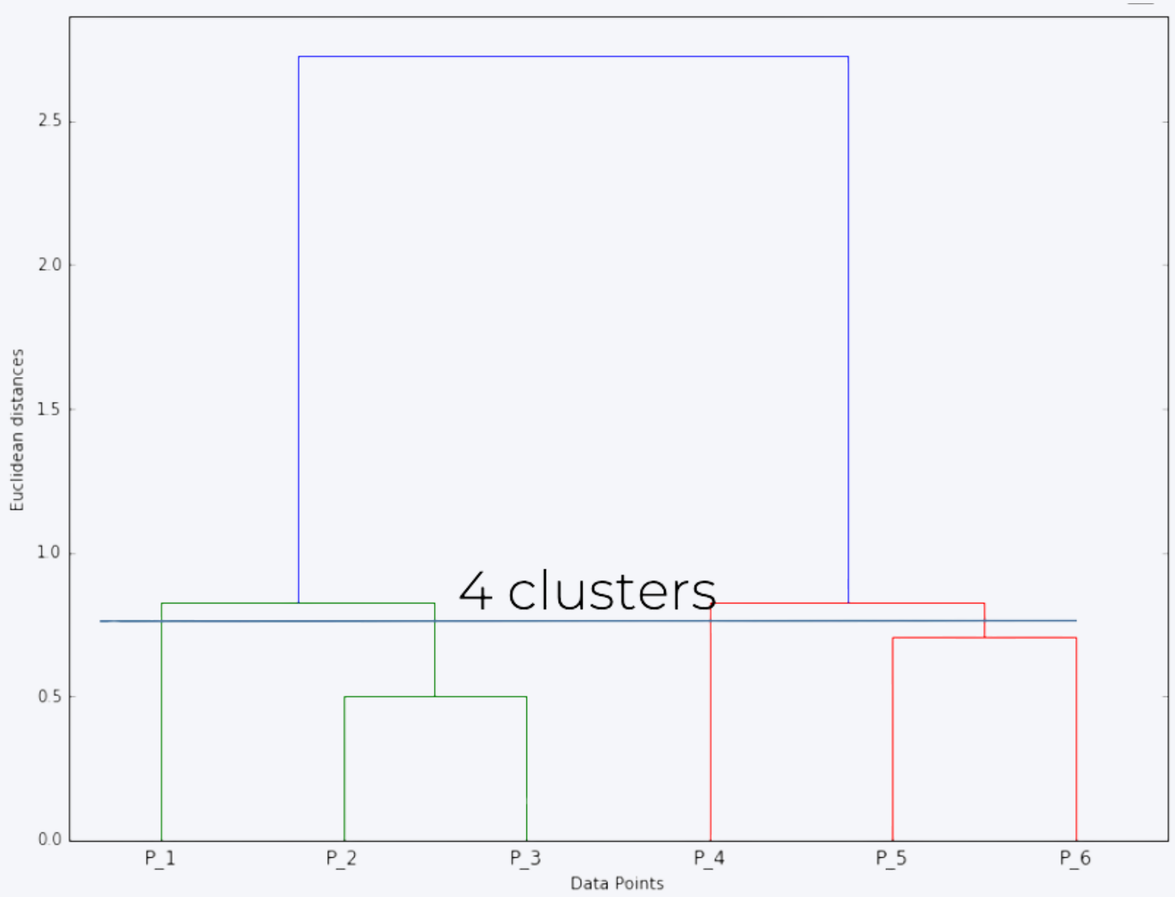


- We can set threshold (distance threshold or dissimilarities) which help in making better clusters by increasing similarity of these clusters // by vertical lines, we can see how many clusters we have.

Hierarchical Clustering

Using Dendograms ?

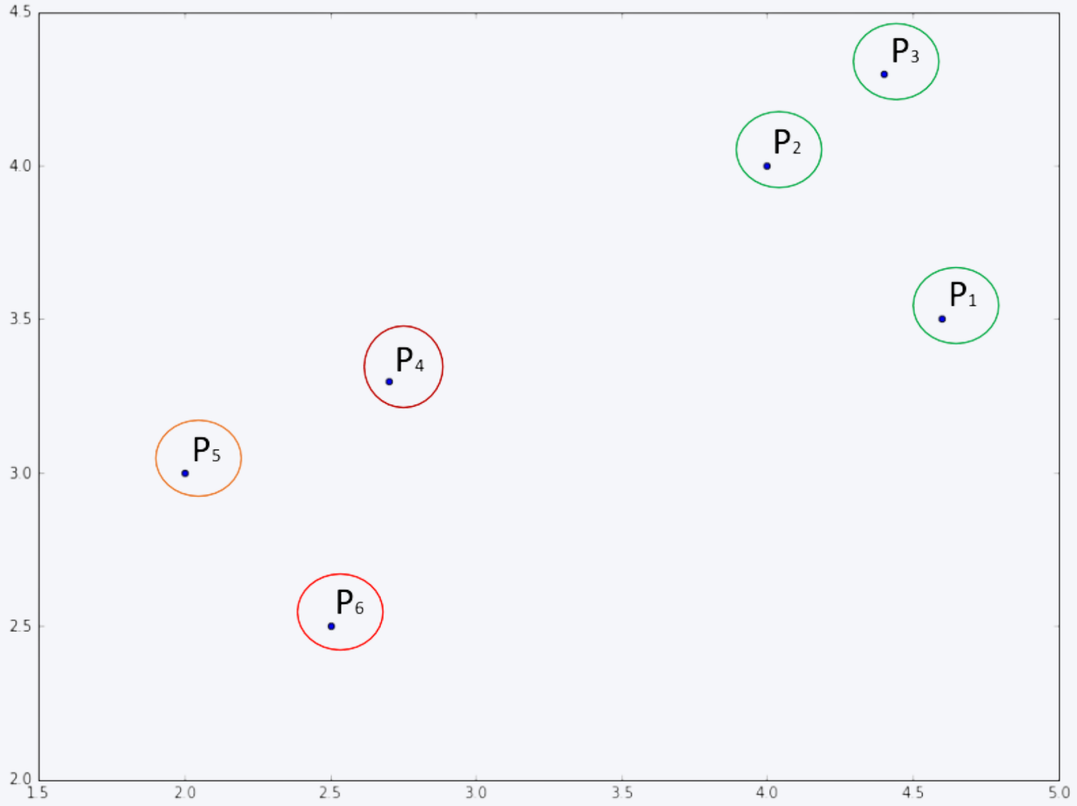


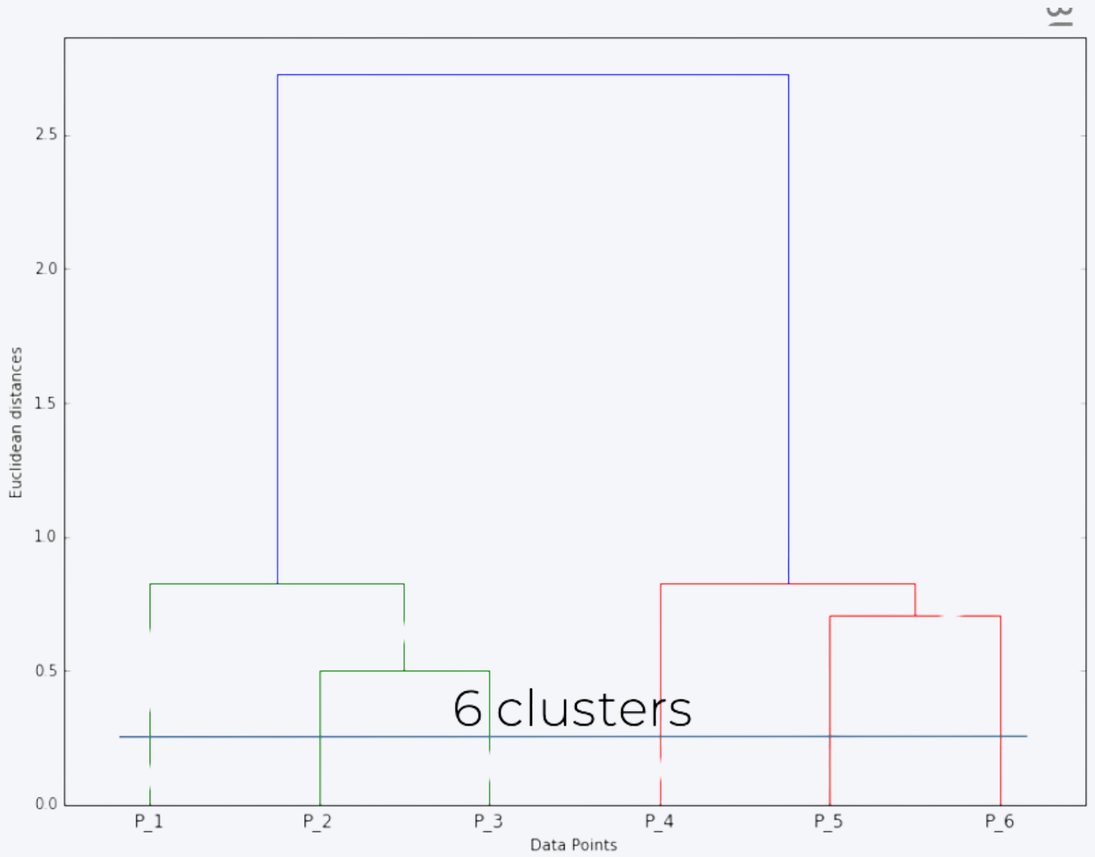


- Another threshold

Hierarchical Clustering

Using Dendograms ?



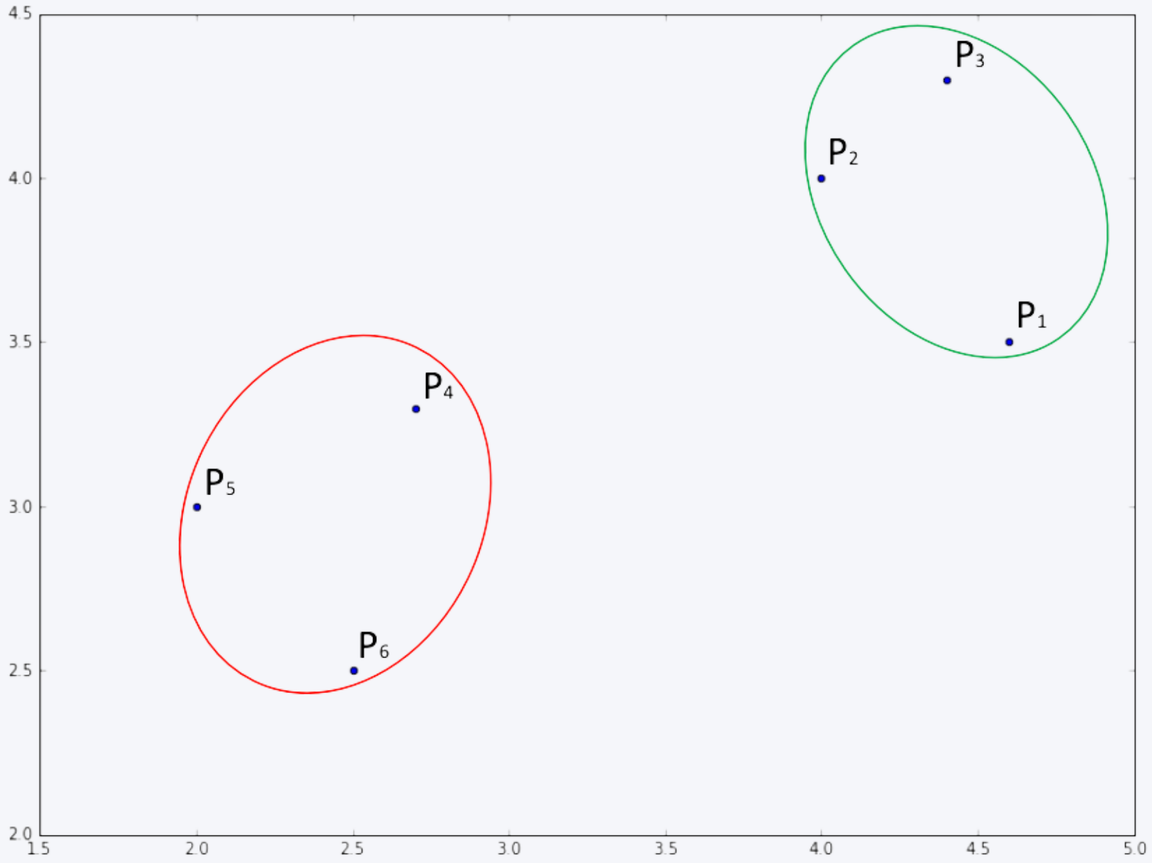


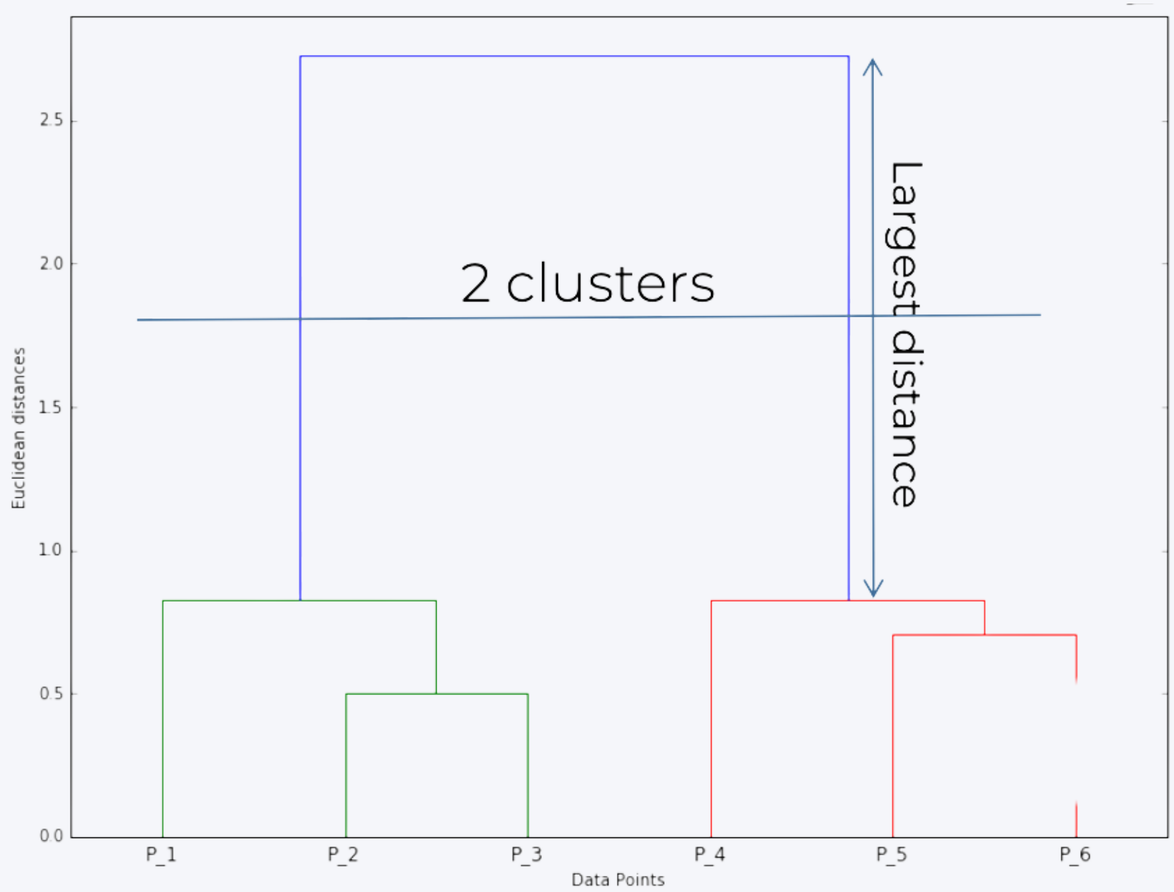
- Threshold below our first clusters

Hierarchical Clustering

أكاديمية طويق
Tuwaiq Academy

Using Dendograms | Optimal Number of clusters:



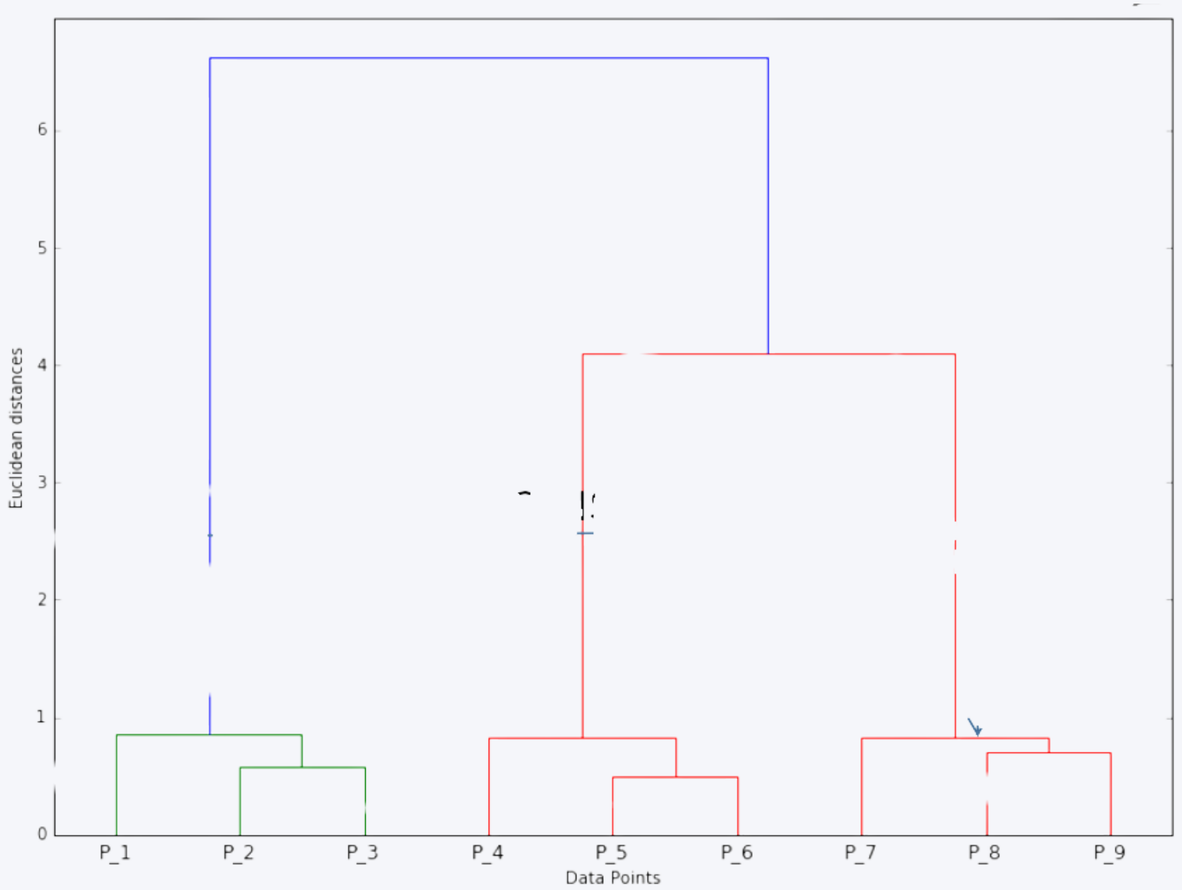


- By using vertical distance that **not** cross any horiztional line and should be **longest** line
-

Hierarchical Clustering

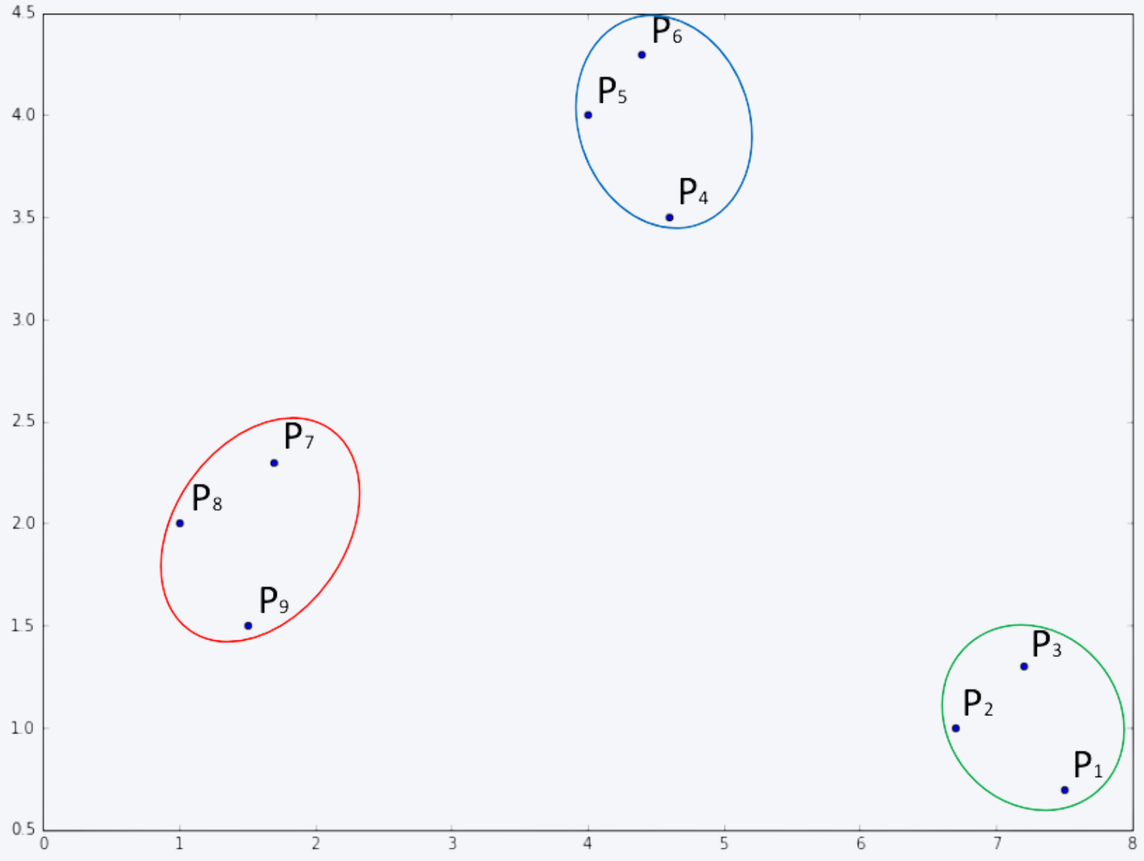
Dendograms | Knowledge Test

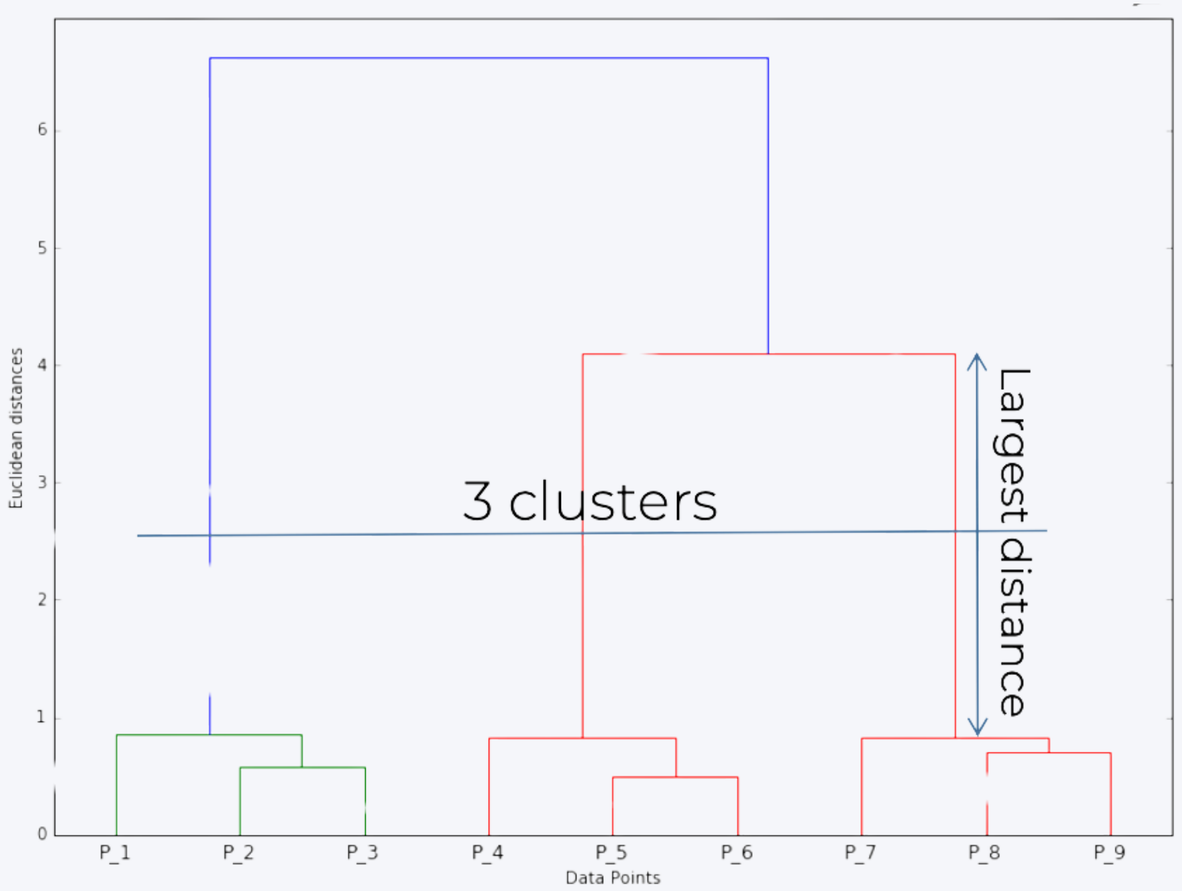
???



Hierarchical Clustering

Dendograms | Knowledge Test





Clustering

أكاديمية طويق
Tuwaiq Academy

Difference Between Clustering Models

Clustering Model	Pros	Cons
K-Means	Simple to understand, easily adaptable, works well on small or large datasets, fast, efficient and performant	Need to choose the number of clusters
Hierarchical Clustering	The optimal number of clusters can be obtained by the model itself, practical visualisation with the dendrogram	Not appropriate for large datasets

أكاديمية طويق
Tuwaiq Academy

Hands-On Code

Clustering Implementation