




Introduction to Unsupervised Learning





Introduction to Unsupervised Learning

Unsupervised learning is a type of machine learning where the algorithm is trained on unlabeled data, meaning the data does not come with predefined outputs or categories.

Instead, the system tries to discover hidden patterns, structures, or relationships within the data on its own.

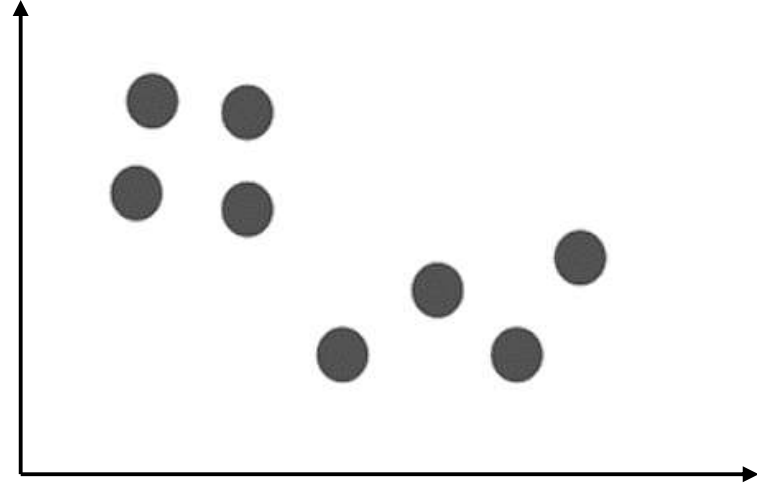
K Means Clustering

K-Means Clustering

- K-means clustering is a popular unsupervised machine learning algorithm used for data segmentation and grouping similar data points into clusters.
- It is widely employed in various fields, including data analysis, image processing, and customer segmentation.

Steps of K – Means Clustering

- What is K?
 - It is the number of clusters
- Let's try to understand the algorithm using a 2-dimensional space data.
- Let's plot them in 2-D space.



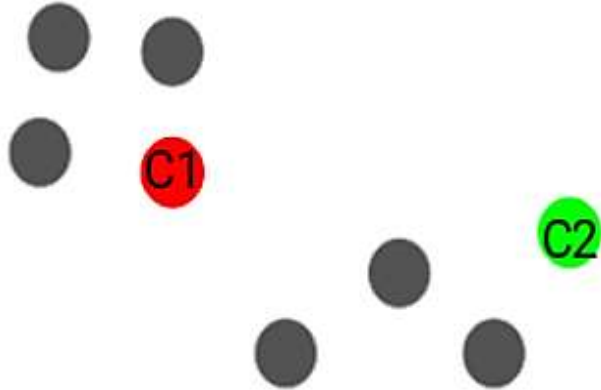
Steps of K – Means Clustering

- Initialization: Start by selecting K initial centroids randomly. Assume K is 2.



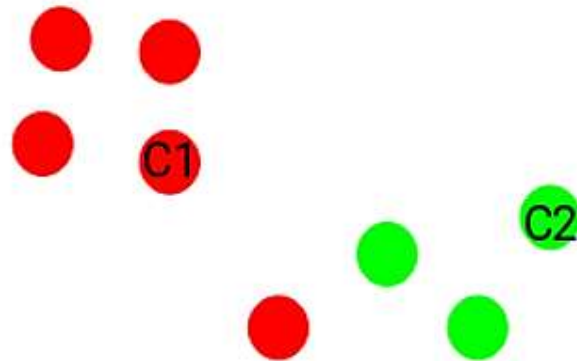
Steps of K – Means Clustering

- Initialization: Start by selecting K initial centroids randomly. Assume K is 2.



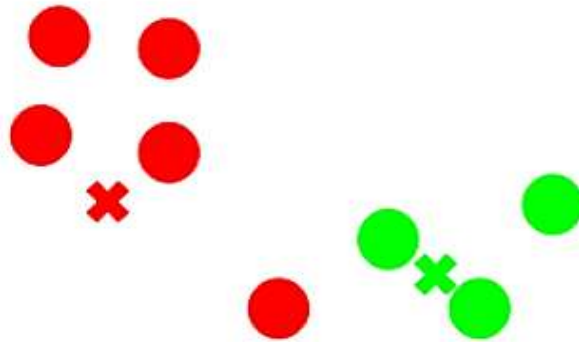
Steps of K – Means Clustering

- Assignment: Assign each data point to the nearest centroid.
- This is done by calculating the Euclidean distance between each data point and each centroid.
- The data point is assigned to the cluster with the nearest centroid.



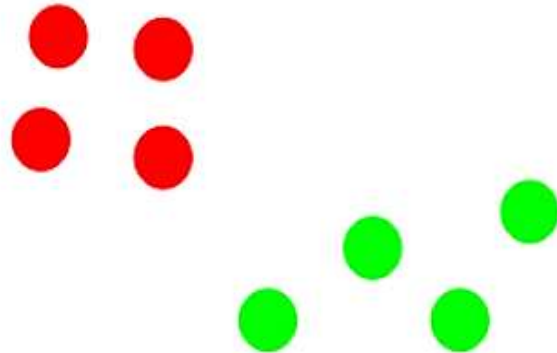
Steps of K – Means Clustering

- Update Centroids: Recalculate the centroids of each cluster by taking the mean of all data points assigned to that cluster.
- These new centroids represent the updated cluster centers.



Steps of K – Means Clustering

- Repeat: Steps 2 and 3 are repeated iteratively until one of the stopping criteria is met.



Steps of K – Means Clustering

- There are several stopping criteria that can be adopted to stop the K-Means algorithm.
 - Centroids of newly formed clusters do not change
 - Points remain in the same cluster
 - Maximum number of iterations are reached

Example

- Consider the following points

	x1	x2
A	2	3
B	6	1
C	1	2
D	3	1
E	6	4

Example

- First choose 2 random points from these points (Assume $K = 2$)

	x1	x2
A	2	3
B	6	1
C	1	2
D	3	1
E	6	4

Example

- Calculate the Euclidean Distance from the selected points to each point separately.

	A	B	C	D	E
A	0	4.47214	1.414214	2.23607	4.123106
B	4.47214	0	5.09902	3	3

Example

- Assign each point to the closest cluster

Cluster 1

Cluster 2

		X1	X2
Cluster 1	A	2	3
	B	6	1
	C	1	2
Cluster 2	D	3	1
	E	6	4

Example

- Then find the new centroids by calculating the average of coordinates

	X1	X2
Centrod1	2	2
Centrod2	6	2.5

Example

- Calculate the Euclidean Distance from new centroids to each point separately

	A	B	C	D	E
C1	1	4.123106	1	1.414214	4.47214
C2	4.03113	1.5	5.02494	3.3541	1.5

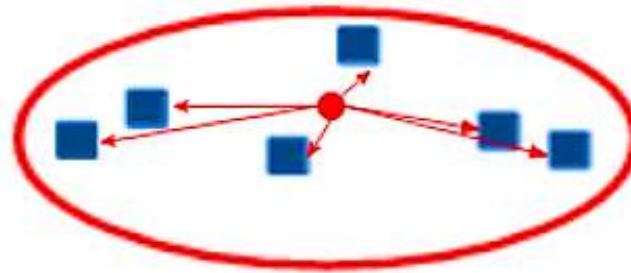
- The same clusters have been found (And the same centroids)
- Then stop

Evaluating Clustering

- There are no many metrics which can be used for measuring the accuracy of the unsupervised learning algorithms as the supervised learning.
- But two metrics are there which can be used for evaluating the clusters.
 - Inertia
 - Dunn Index

Inertia

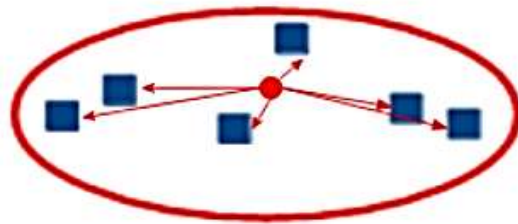
- Calculates the sum of distances of all the points within a cluster from the centroid of that cluster.



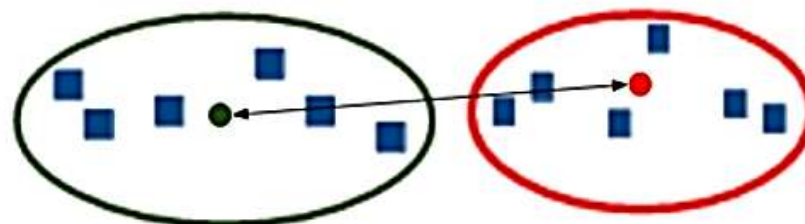
Intra cluster distance

Dunn Index

- Takes into account the distance between two clusters.



Intra cluster distance

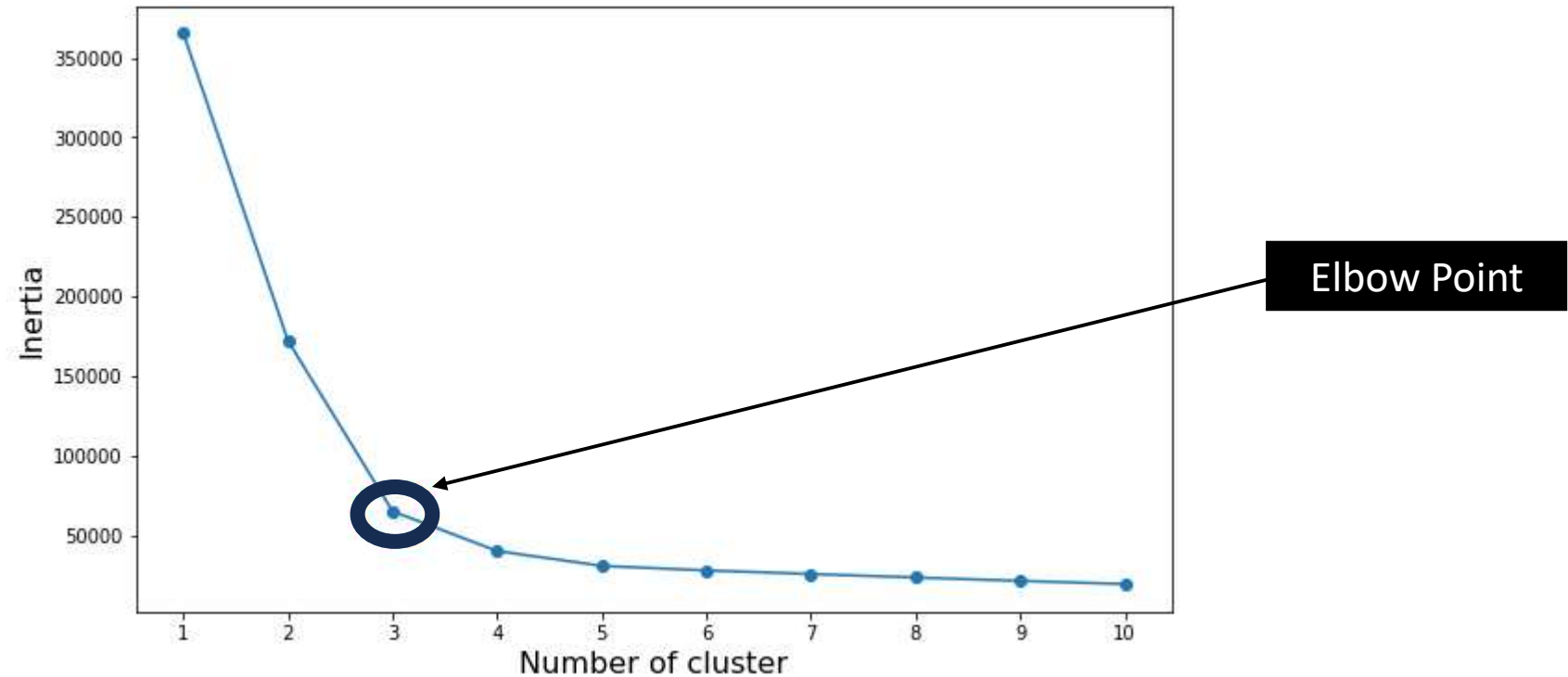


Inter cluster distance

$$\text{Dunn Index} = \frac{\min(\text{Inter cluster distance})}{\max(\text{Intra cluster distance})}$$

Selecting K

- Consider several values for K and find the inertia values for each K.



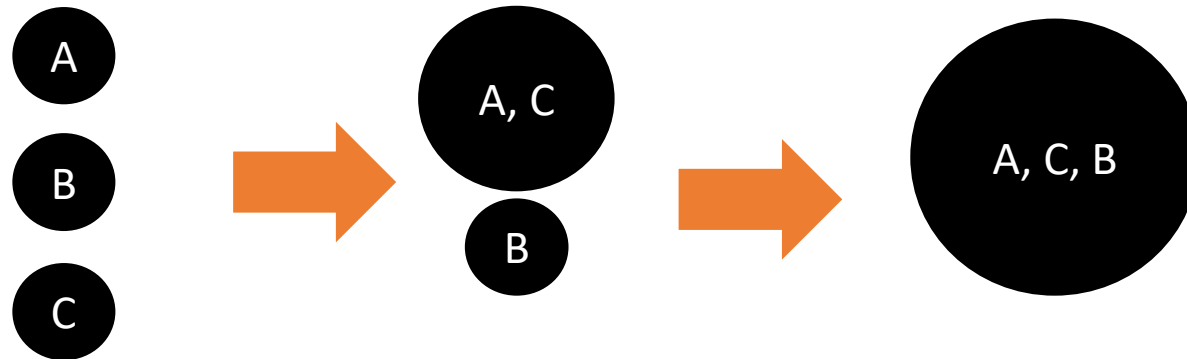
Hierarchical Clustering

Hierarchical Clustering

- Hierarchical clustering is a data analysis technique that organizes data objects into a tree-like structure based on their similarity.
- It's important in various fields, including biology, social sciences, and image analysis, for identifying hierarchical relationships within data, which can aid in understanding groupings, taxonomy, and decision-making based on data similarities.

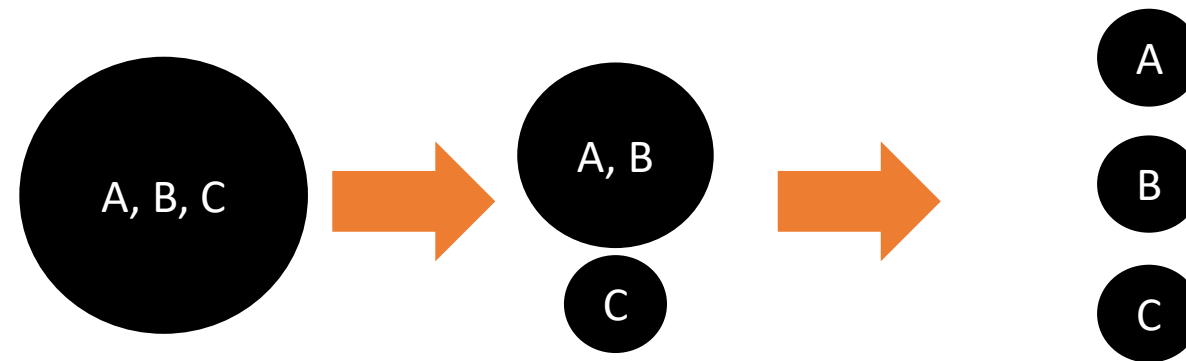
Hierarchical Clustering

- There are two types of Hierarchical Clustering.
 - Agglomerative Hierarchical Clustering – Bottom Up (This is the mostly used one)



Hierarchical Clustering

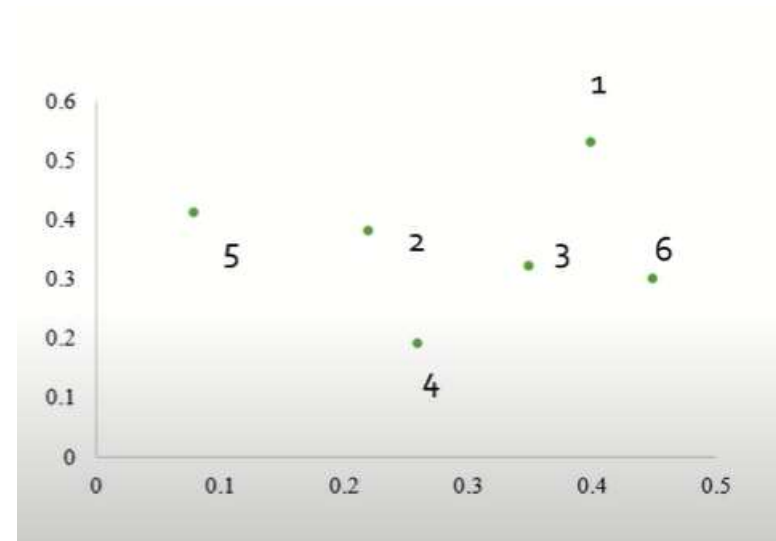
- There are two types of Hierarchical Clustering.
 - Divisive Hierarchical Clustering – Top Down



Agglomerative Hierarchical Clustering

- Consider the following example.

	X1	X2
A	0.40	0.53
B	0.22	0.38
C	0.35	0.32
D	0.26	0.19
E	0.08	0.41
F	0.45	0.30



Agglomerative Hierarchical Clustering

- After calculating the distances between each pair using Euclidean Distance, the distance matrix is given below.

	A	B	C	D	E	F
A	0					
B	0.23	0				
C	0.22	0.15	0			
D	0.37	0.20	0.15	0		
E	0.34	0.14	0.28	0.29	0	
F	0.23	0.25	0.11	0.22	0.39	0

Agglomerative Hierarchical Clustering

- The minimum distance is given for C and F. So, they are grouped.

	A	B	C	D	E	F
A	0					
B	0.23	0				
C	0.22	0.15	0			
D	0.37	0.20	0.15	0		
E	0.34	0.14	0.28	0.29	0	
F	0.23	0.25	0.11	0.22	0.39	0

Agglomerative Hierarchical Clustering

- How can we update the new distance matrix after grouping?

	A	B	C, F	D	E
A	0				
B	0.23	0			
C, F	?	?	0		
D	0.37	0.20	?	0	
E	0.34	0.14	?	0.29	0

Agglomerative Hierarchical Clustering

Linkage		Example
Single Linkage	Smallest distance	Min ($d[A,C]$, $d[A,F]$)
Complete Linkage	Largest distance	Max ($d[A,C]$, $d[A,F]$)
Average Linkage	Average distance	Avg ($d[A,C]$, $d[A,F]$)

Agglomerative Hierarchical Clustering

- Let's use the Single Linkage.

	A	B	C, F	D	E
A	0				
B	0.23	0			
C, F	Min(0.22,0.23)	Min(0.15,0.25)	0		
D	0.37	0.20	Min(0.15,0.22)	0	
E	0.34	0.14	Min(0.28,0.39)	0.29	0

Agglomerative Hierarchical Clustering

- Updated distance matrix is,

	A	B	C, F	D	E
A	0				
B	0.23	0			
C, F	0.22	0.15	0		
D	0.37	0.20	0.15	0	
E	0.34	0.14	0.28	0.29	0

Agglomerative Hierarchical Clustering

- The minimum distance is,

	A	B	C, F	D	E
A	0				
B	0.23	0			
C, F	0.22	0.15	0		
D	0.37	0.20	0.15	0	
E	0.34	0.14	0.28	0.29	0

Agglomerative Hierarchical Clustering

- Then the process is continued.

	A	B, E	C, F	D
A	0			
B, E	0.23	0		
C, F	0.22	0.15	0	
D	0.37	0.20	0.15	0

Agglomerative Hierarchical Clustering

- Then the process is continued.

	A	B, E, C, F	D
A	0		
B, E, C, F	0.22	0	
D	0.37	0.15	0



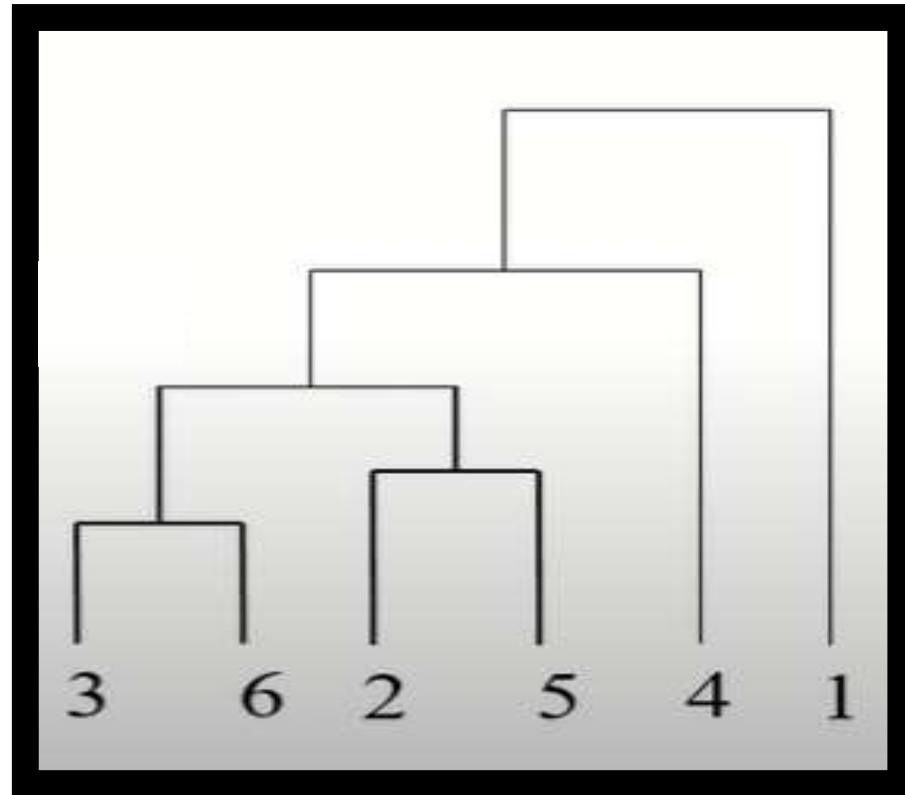
	A	B, E, C, F, D
A	0	
B, E, C, F, D	0.22	0



	B, E, C, F, D, A
B, E, C, F, D, A	0

Agglomerative Hierarchical Clustering

- Dendrogram can be created after the clustering process.





Other Clustering Methods

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
 - Groups together points that are **densely packed**.
 - Can find arbitrarily shaped clusters.
 - Identifies outliers (noise points).
 - Example: Detecting unusual customer spending patterns.
-



Other Clustering Methods

- **OPTICS (Ordering Points To Identify the Clustering Structure)**
 - Extension of DBSCAN.
 - Handles clusters of **varying density** better.
 - Builds a cluster ordering rather than fixed clusters.
-



Other Clustering Methods

- **Mean Shift Clustering**
 - Finds “peaks” (high-density regions) in data.
 - Doesn’t need the number of clusters in advance.
 - Useful for image segmentation and feature analysis.
-



Other Clustering Methods

- **Gaussian Mixture Models (GMMs)**
 - A **probabilistic model** where each cluster is assumed to follow a Gaussian (normal) distribution.
 - Provides **soft clustering** (a point can belong partly to multiple clusters).
 - Example: Speaker recognition, customer segmentation.
-



Other Clustering Methods

- **Spectral Clustering**
 - Uses **graph theory** and eigenvalues of a similarity matrix.
 - Very powerful for **non-convex clusters**.
 - Example: Social network analysis.
-



Other Clustering Methods

- **BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)**
 - Handles **large datasets efficiently**.
 - Builds a tree structure (CF tree) and clusters data incrementally.
-