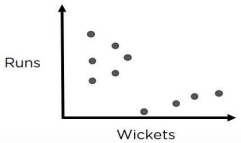
K-MEANS CLUSTERING

Dr. Shailesh S
DCS CUSAT

Assign data points

Here, we have our dataset with x and y coordinates

Now, we want to cluster this data using **K-Means**



What is K-Means Clustering?

Task: Identify bowlers and batsmen

- The data contains runs and wickets gained in the last 10 matches
- So, the bowler will have more wickets and the batsmen will have higher runs

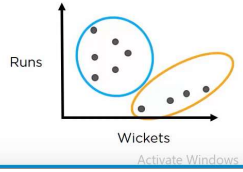
Scores

Cluster 1

We can see that this cluster has players with high runs and low wickets

Cluster 2

And here, we can see that this cluster has players with high wickets and low wickets



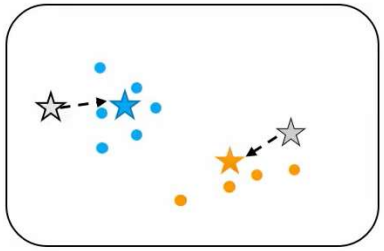
the process of grouping a set of objects into classes of similar object

Clustering

- Clustering: the process of grouping a set of objects into classes of similar object
  - Documents within a cluster should be similar.
  - Documents from different clusters should be dissimilar.
- Goal: maximize intra-cluster similarity and minimize inter-cluster similarity
- **Unsupervised learning = learning from raw data**

What is K-Means Clustering?

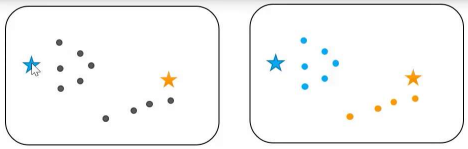
Reposition the two centroids for optimization.



What is K-Means Clustering?

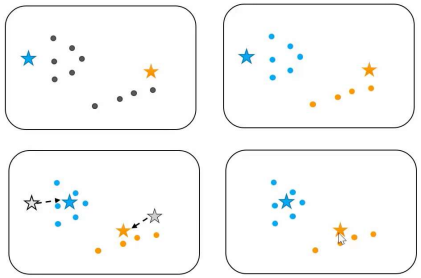
Initially, two centroids are assigned randomly

The Euclidean distance is used to find out which centroid is closest to each data point and the data points are assigned to the corresponding centroids.



What is K-Means Clustering?

The process is iteratively repeated until our centroids become static

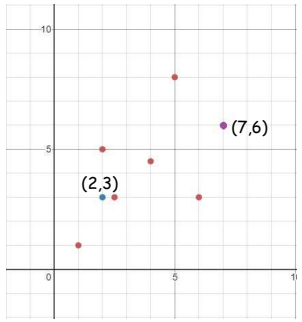


K Means

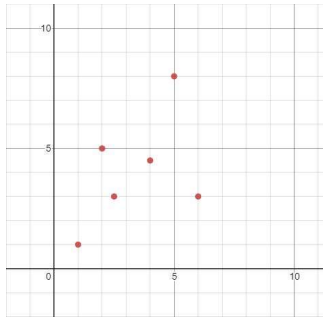
DEMO

Samples	x	y
1	1	1
2	2	5
3	2.5	3
4	5	8
5	4	4.5
6	6	3

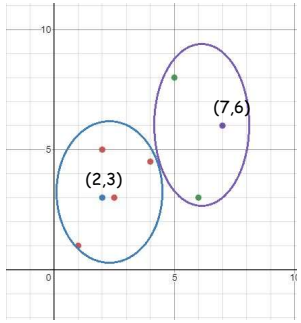
C1 - (2,3)	C2 - (7,6)	Cluster
2.236067977	7.810249676	C1
2	5.099019514	C1
0.5	5.408326913	C1
5.830951895	2.828427125	C2
2.5	3.354101966	C1
4	3.16227766	C2



Samples	x	y
1	1	1
2	2	5
3	2.5	3
4	5	8
5	4	4.5
6	6	3

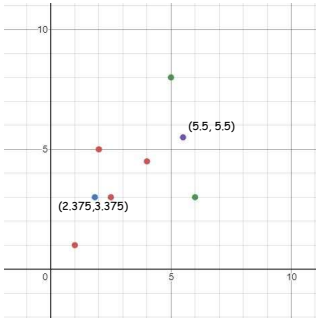


A			B	
1	1		5	8
2	5		6	3
2.5	3			
4	4.5			
2.375	3.375		5.5	5.5



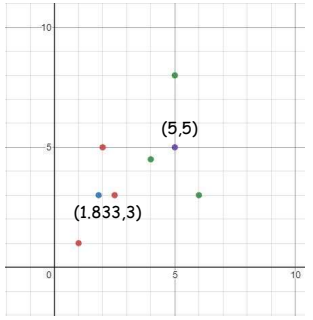
Samples	x	y
1	1	1
2	2	5
3	2.5	3
4	5	8
5	4	4
6	6	3

C1 - (2.375,3.375	C2 - (5.5,5.5)	Cluster
2.7443123	6.363961031	C1
1.667708008	3.535533906	C1
0.395284708	3.905124838	C1
5.318011847	2.549509757	C2
2.061552813	1.802775638	C2
3.644344934	2.549509757	C2

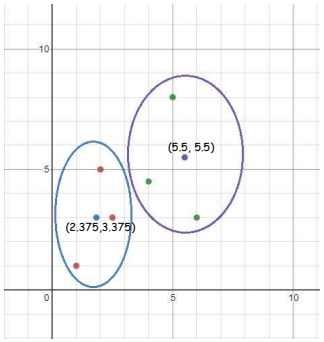


Samples	x	y
1	1	1
2	2	5
3	2.5	3
4	5	8
5	4	4
6	6	3

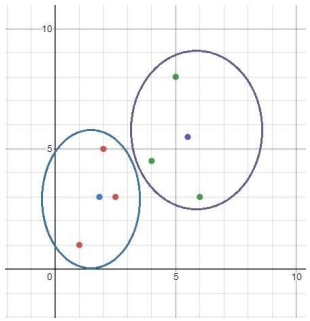
C1 - (1.8333,3)	C2 - (5,5)	Cluster
2.166538483	5.656854249	C1
2.006960139	3	C1
0.667	3.201562119	C1
5.918605326	3	C2
2.635505454	1.118033989	C2
4.167	2.236067977	C2



A			B	
1	1		5	8
2	5		4	4
2.5	3		6	3
1.833333	3		5	5



A			B	
1	1		5	8
2	5		4	4
2.5	3		6	3
1.833333	3		5	5



Applications of K-Means Clustering