Unsupervised Learning – K-means implementation

Project discussion

Unsupervised Learning

dataset will contain only features. There exists no label.

Goal of Unsupervised Learning

Set (feature) → Unsupervised Algo. →

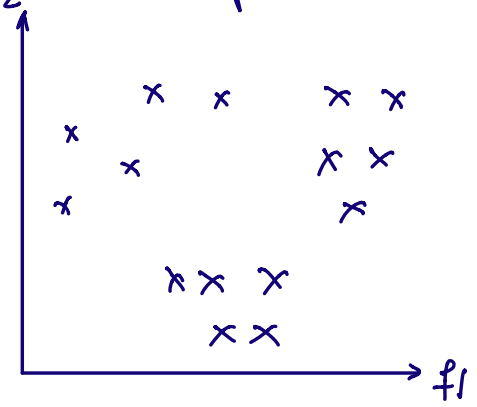
- ✓ Find & Discover Hidden pattern
- Find & Discover Associations in the data
- Find & Discover Genetic behaviour
- Optimize & Extract / Select features from dataset.

– ML engineer

K-means algo.

- K-means is a clustering algorithm.
- Your dataset will be categorized as clusters (groups) on the basis of your assumption.

eg:



scatter plot

Basis of assumption for n_clusters:

1. Visual EDA (Exploratory DA)
   - scatter plot
   - pair plot
2. Elbow method. (check a range of n_clusters and identify where the error was less)

Findings from pair plot for dataset Mall_Customers.csv.

① Customer_Id v/s Spending Score (n_clusters: 5)

② Annual Income v/s Spending Score (n_clusters: 5)

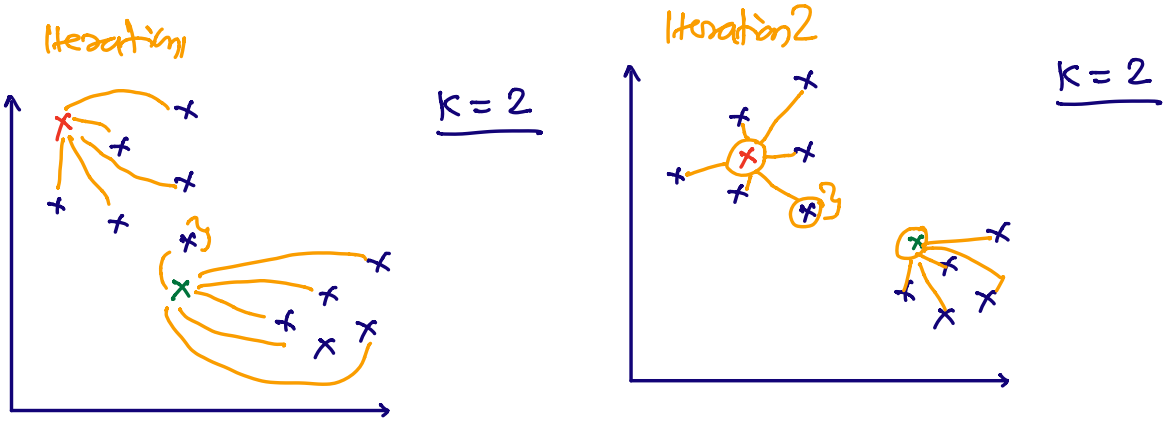
③ Age v/s Spending Score (n_clusters: 2)

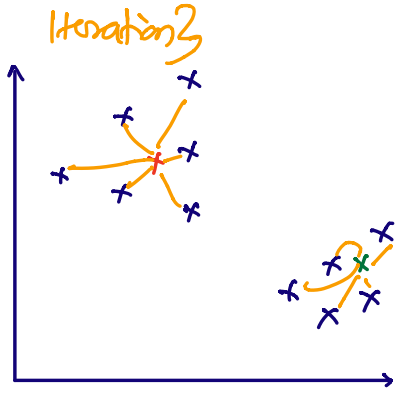
K-means Algo :

① Assume k- value. (n_clusters = k)

② On the basis of k value randomly assign 'k' number of cluster centers.

③ Group all data pts with the nearby cluster centers. (Distance formula (Euclidean))





$K = 2$

End of iteration
Finalize this as the cluster !

④ Perform step 3 till all points belong to correct cluster centers.

⑤ Finalize the cluster.