Session 35

Unsupervised Learning

Table of Content

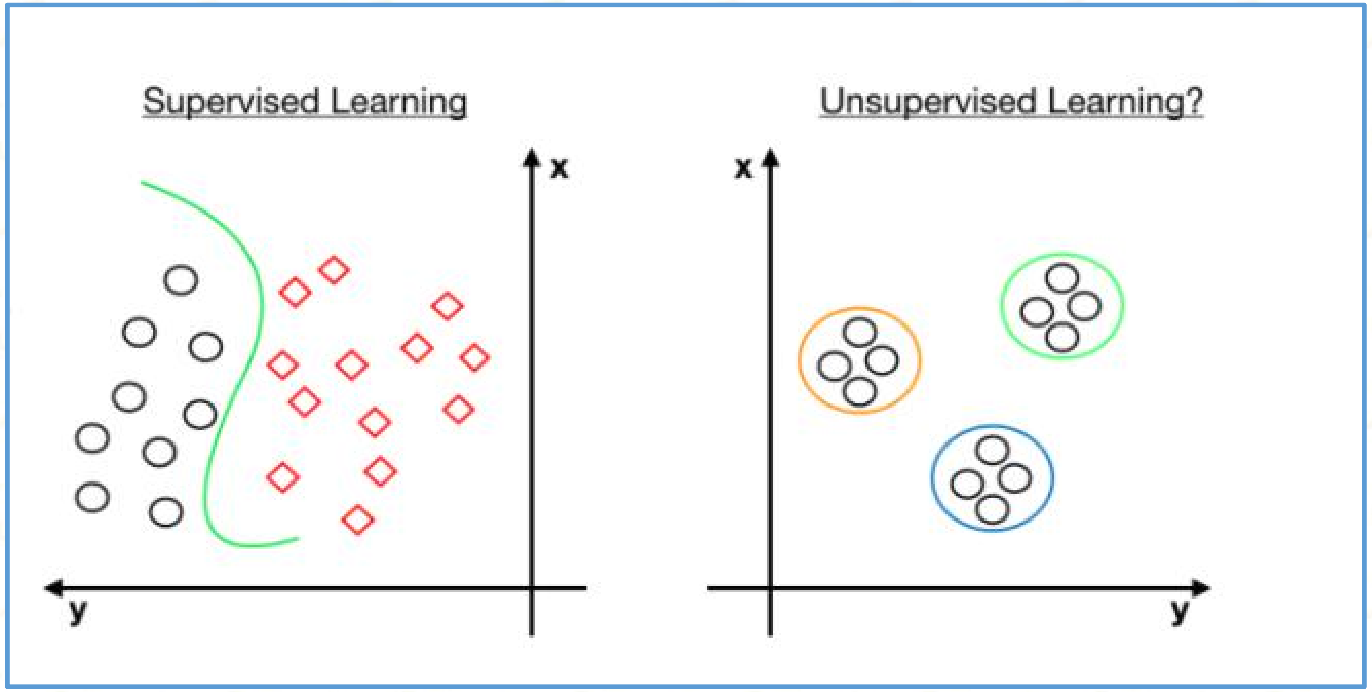
What will We Learn Today?

1. Unsupervised Learning
2. K-Means
3. K-Medoids

Unsupervised Learning

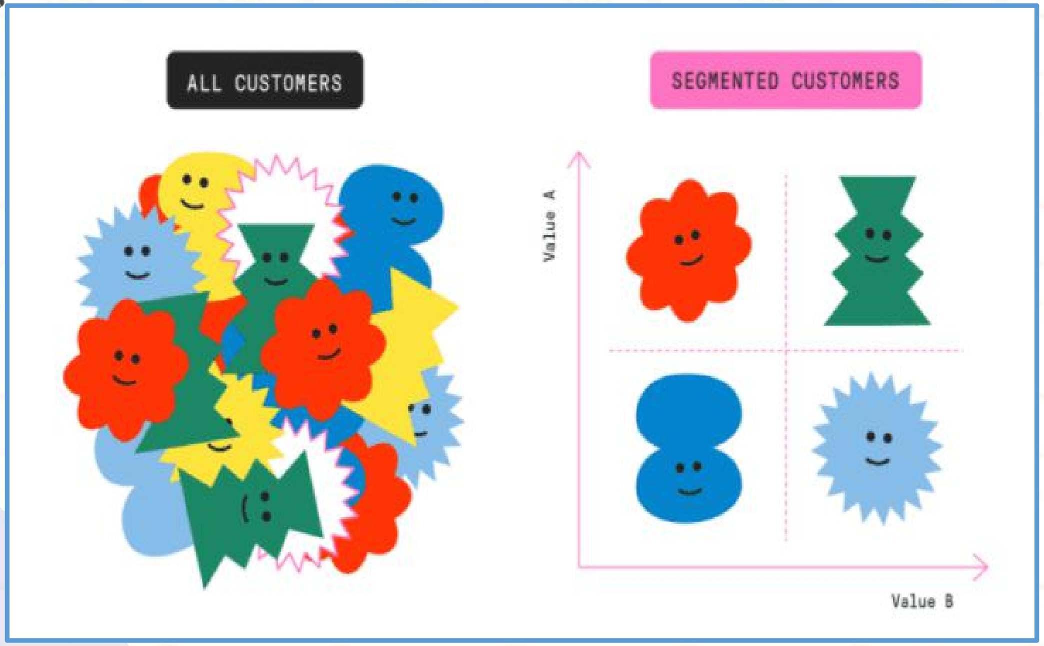
Supervised vs Unsupervised

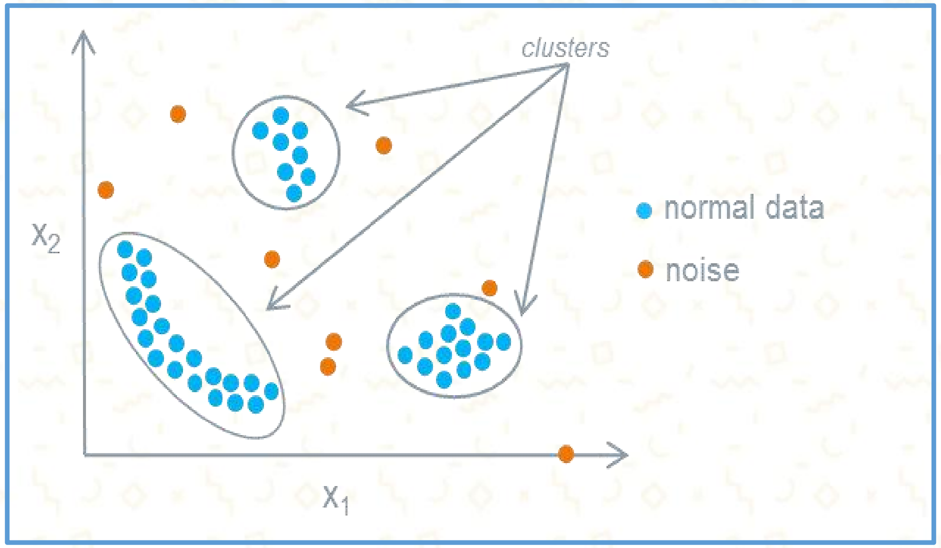
- **Supervised** = Learn to predict the outcome.
  - We know the target label, so we make the model that try to predict the label.
- **Unsupervised** = Finding pattern/ characteristic from data.
  - We do not know our target label, so we make model that try to group the data.



Application of Unsupervised Learning

- Customer segmentation.
  - Understanding different customer groups around which to build marketing or other business strategies.
- Anomaly detection.
- Recommender systems, which involve grouping together users with similar viewing patterns in order to recommend similar content.

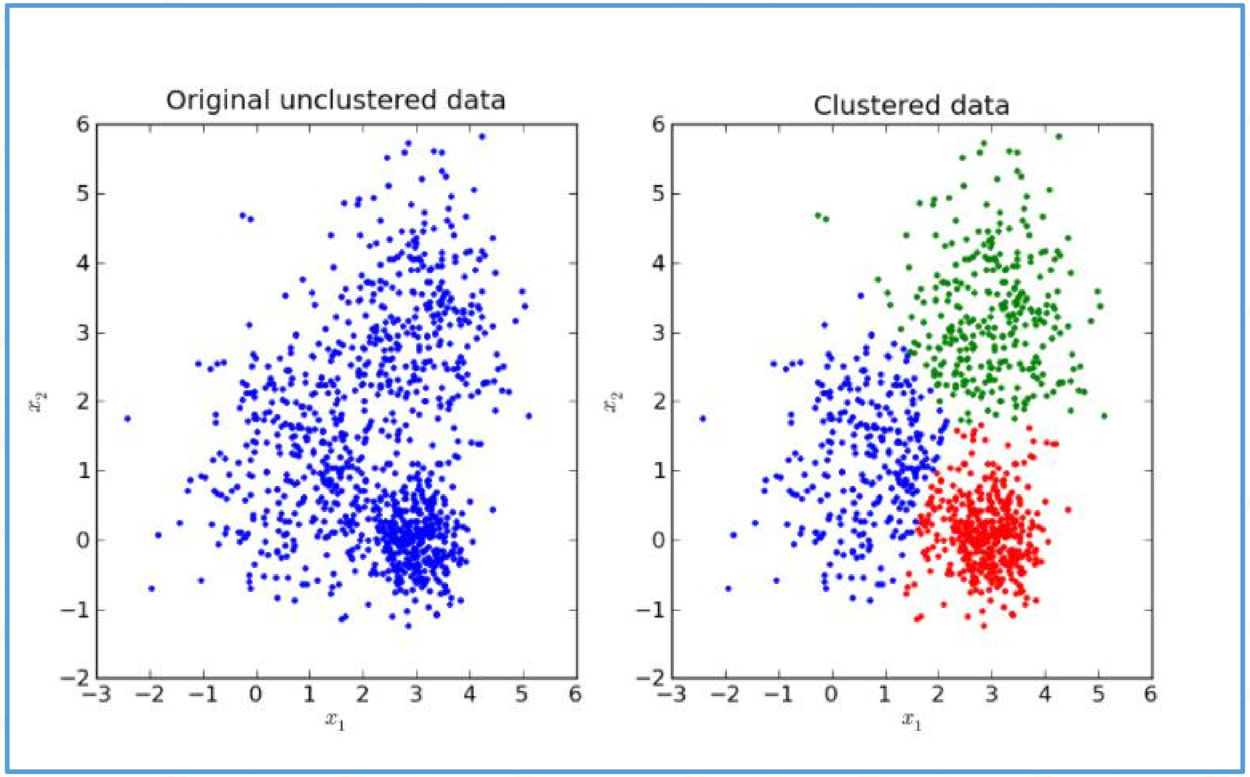




K-Means

K-Means

- K-means clustering algorithm tries to group similar items in the form of clusters.
- The number of groups is represented by *K*.

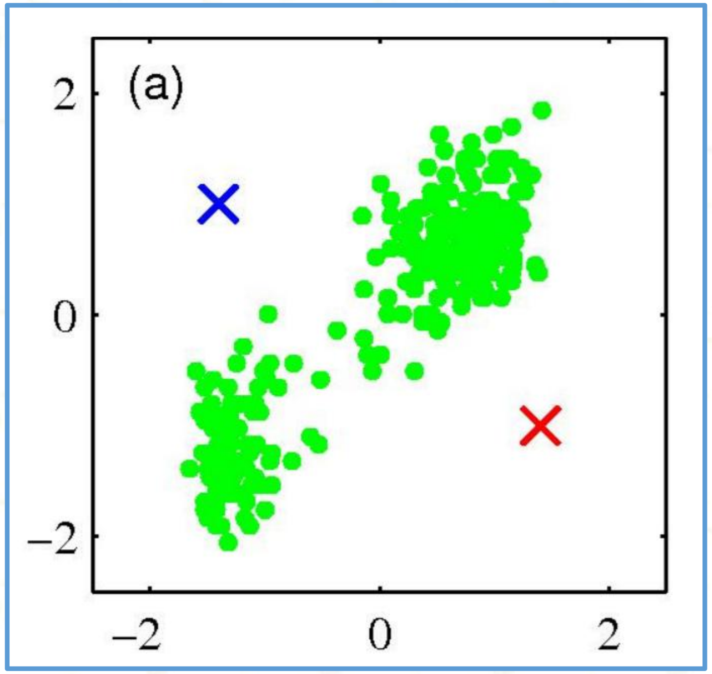


How K-Means works?

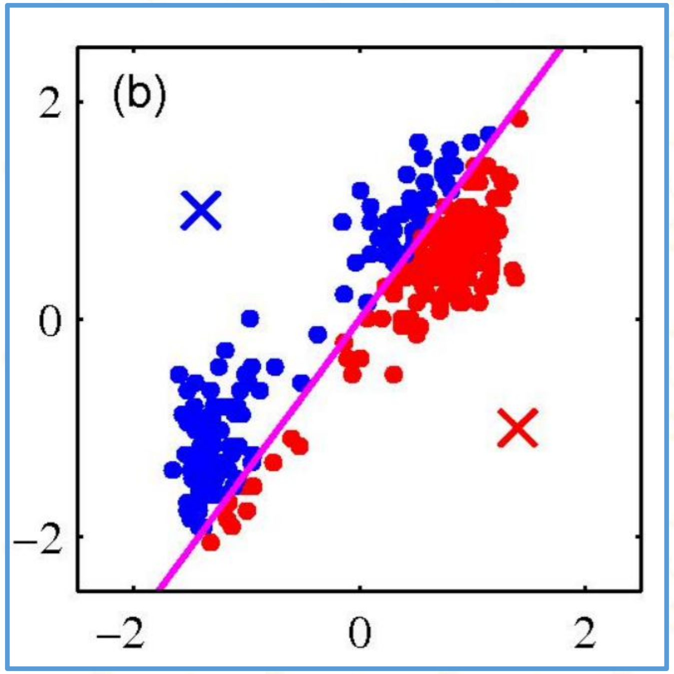
1. **Choose k objects** as initial cluster centers
2. **Assign each object** to the cluster with the nearest center
3. **Update cluster centers** as the mean point of the cluster
4. Go back to Step 2, stop when there is no change

How K-Means works?

- Pick *K* random points as cluster centers (means)
  - Shown here for K=2

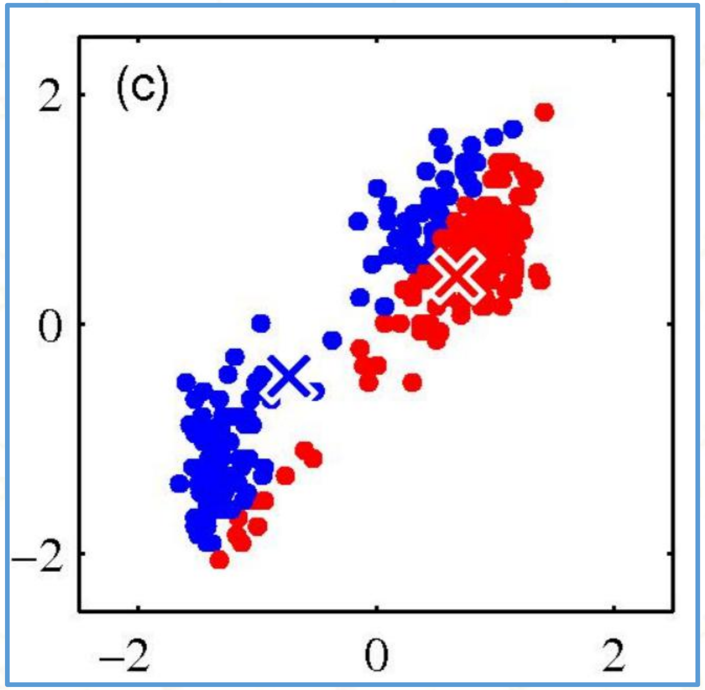


- Iterative Step 1
- • Assign data points to closest cluster center

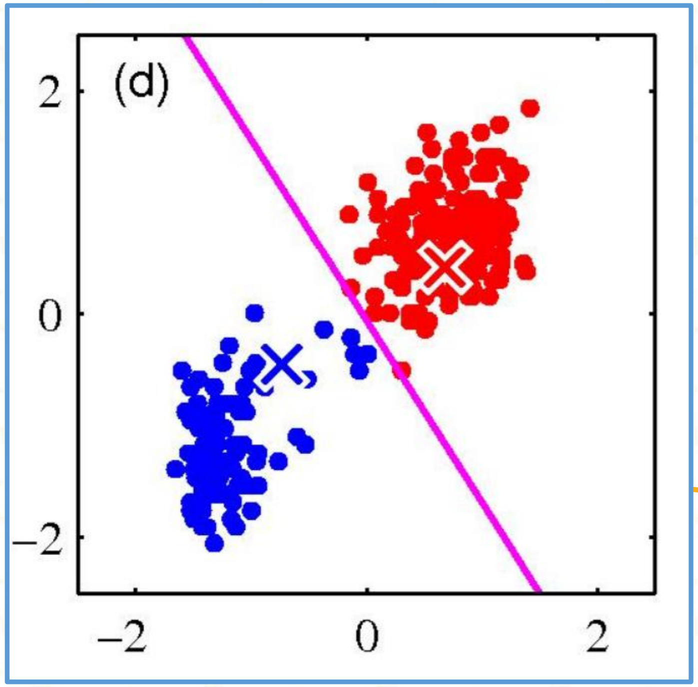


How K-Means works?

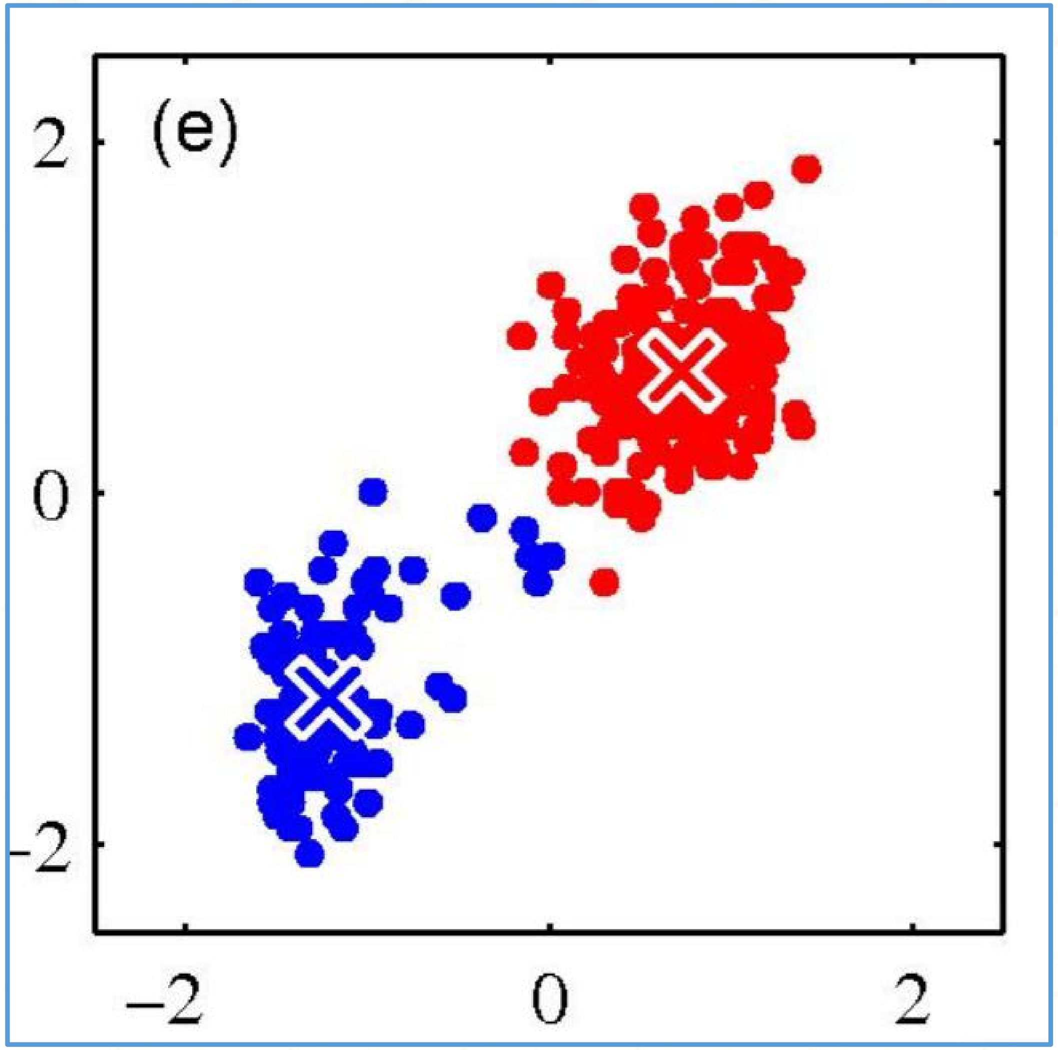
- Iterative Step 2
  - Update cluster center
  - Change the cluster center to the average of the assigned points



- Repeat until convergence



How K-Means works?



Evaluating clustering performance

1. **Inertia**
   - Sum of squared distance from each point (xi) to its cluster (Ck).
   - If the inertia is small, it means that the points are close each other.

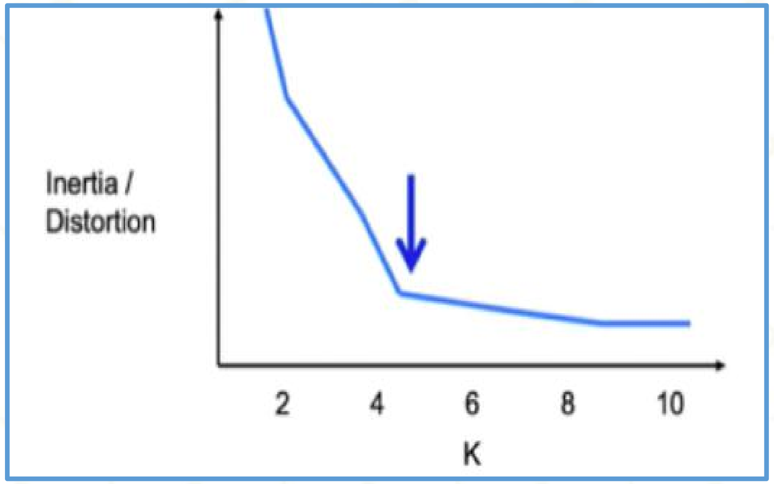
$$\sum_{i=1}^{n}(x_i - C_k)^2$$

2. **Silhoutte score**
   - a : mean distance to all other other points in its cluster.
   - b : mean distance to all other points in the next nearest cluster.
   - The score range between -1 to 1. It is better when the score is near to 1.

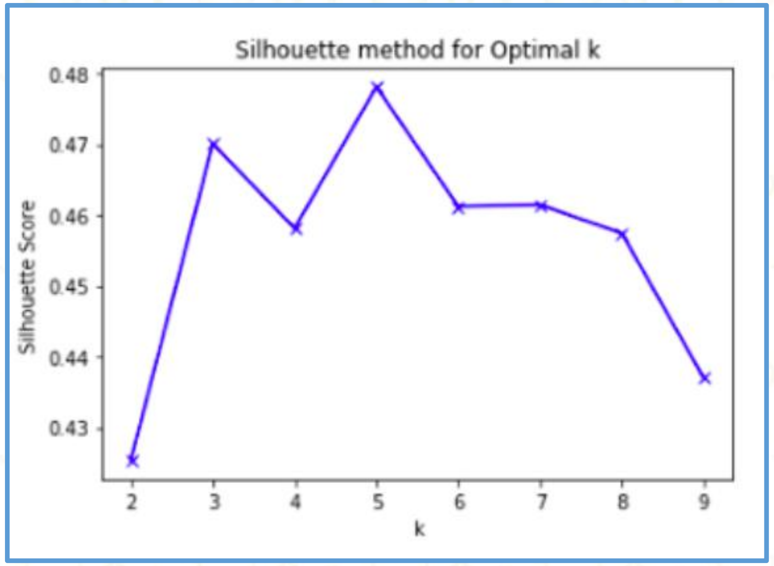
$$SC = \frac{b-a}{\max(a,b)}$$

How to choose the K?

1. **Elbow method**



2. **Silhoutte score**
   - High score is better



Discussion on the K-means

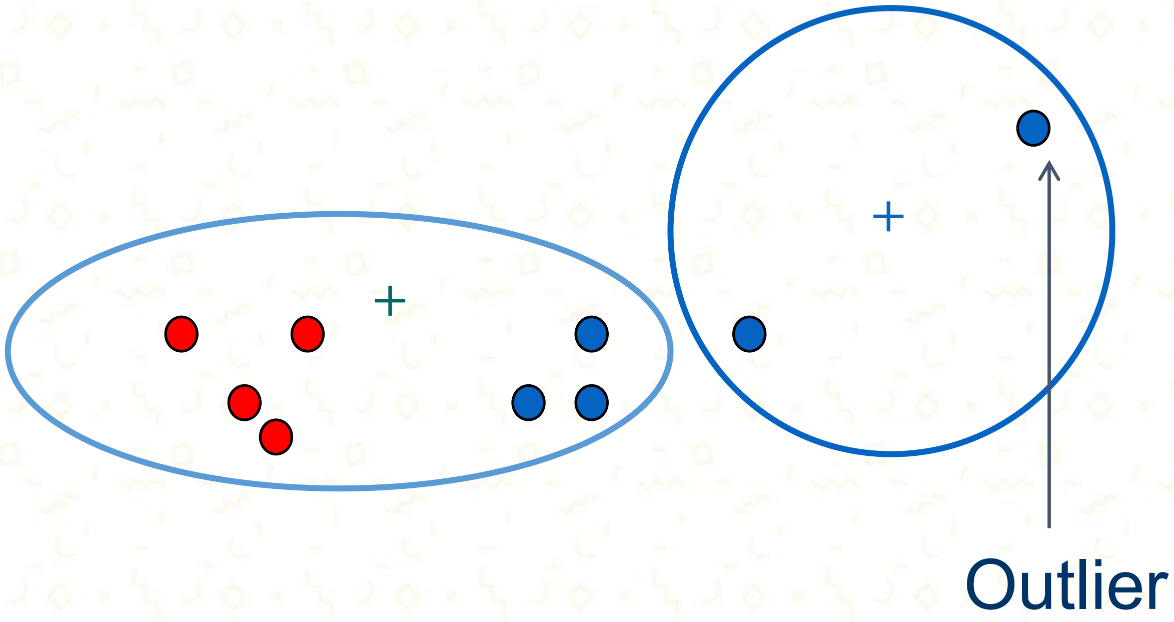
- **Advantages of K-means**
  - It is very simple to implement.
  - It is scalable to a huge data set and also faster to large datasets.
  - It adapts the new examples very frequently.
  - Generalization of clusters for different shapes and sizes.

- **Disadvantages of K-means**
  - It is sensitive to the outliers.
  - Choosing the k values manually is a tough job.
  - As the number of dimensions increases its scalability decreases.

https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering/

A Problem of K-Means

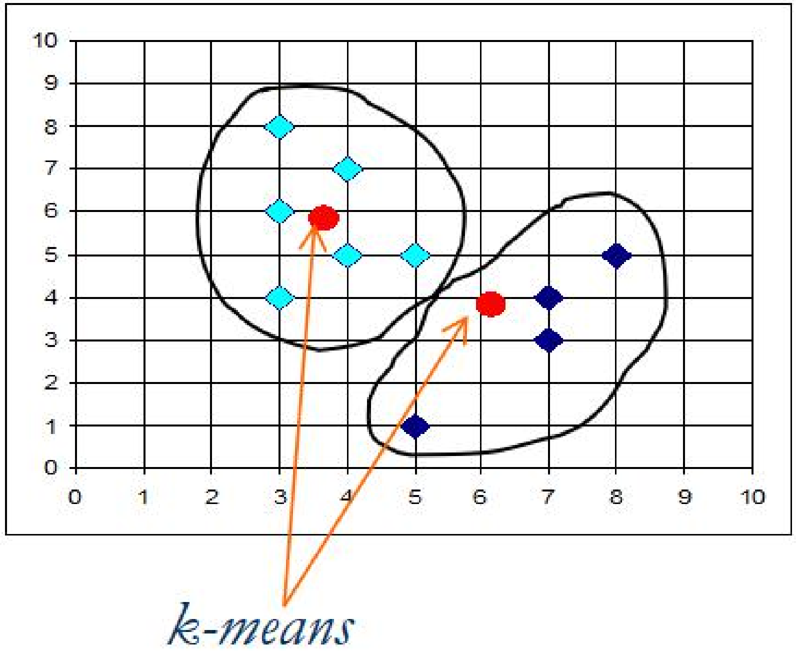
- Sensitive to outliers
- Outlier: objects with extremely large (or small) values

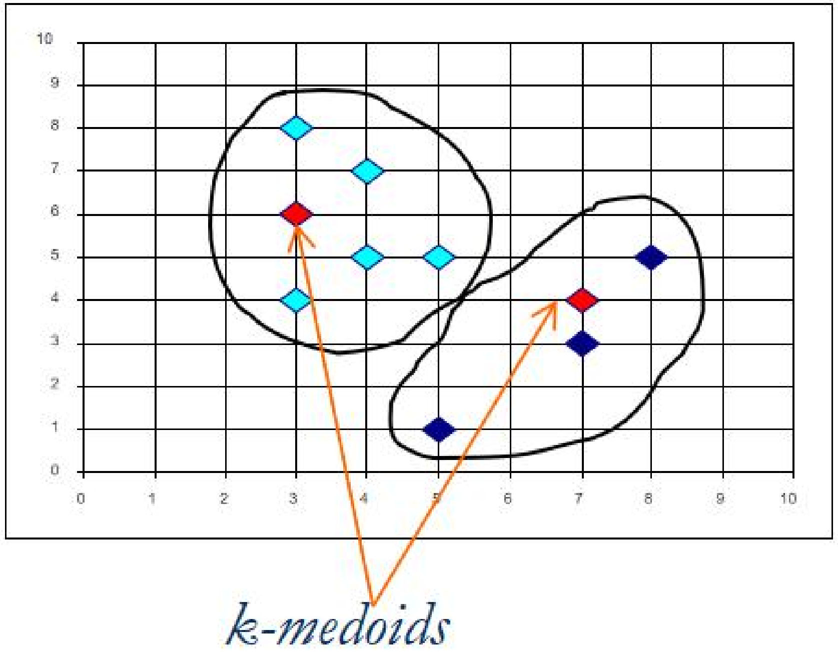


K-Medoids

K-Medoids

- K-medoids: Find k representative objects, called medoids.
  - While K-Means tries to minimize the within cluster sum-of-squares,
  - K-Medoids tries to minimize the sum of distances between each point and the medoid of its cluster.





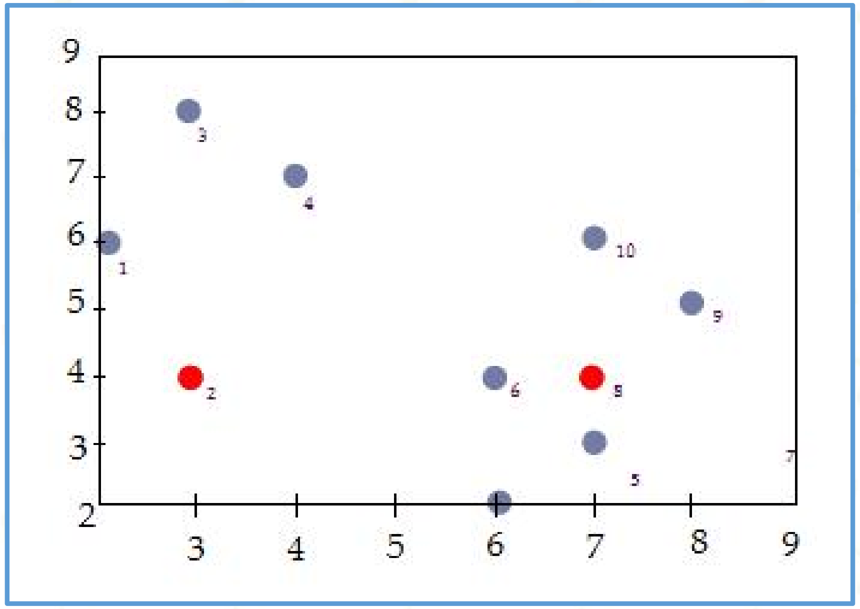
How K-Medoids (PAM) works?

- Partitioning Around Medoids (PAM)

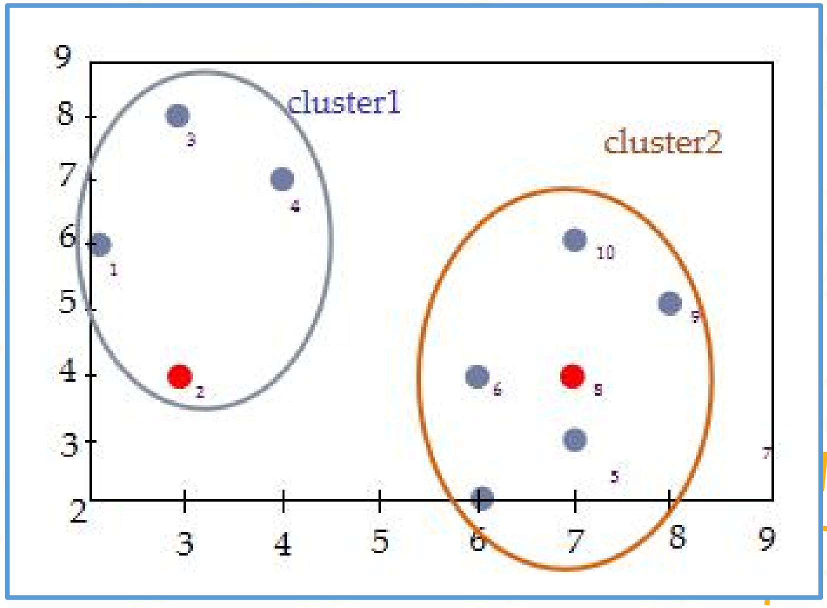
- 1. Initialize: select k random points out of the n data points as the medoids.
- 2. Repeat:
  - Assign each point to the cluster with the closest medoid *m*.
  - Randomly select a non-representative object *oi*
  - Compute the total cost of swapping S, the medoid *m* with *oi*
  - If S < 0:
    - Swap *m* with oi to form new set of medoids.
- Stop when convergence criteria is meet.

How K-Medoids works?

- Pick K random medoids
- Shown here for K=2



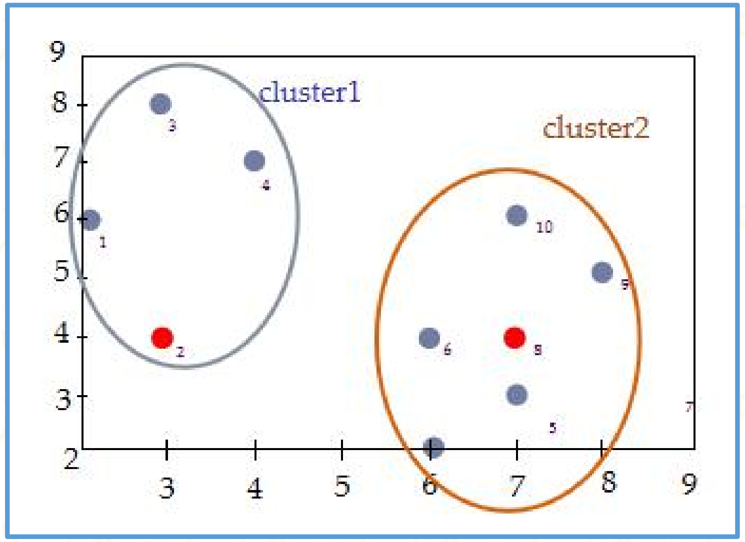
- • Assign data points to closest cluster center



How K-Medoids works?

Data Objects

	A_1	A_2
O_1	2	6
O_2	3	4
O_3	3	8
O_4	4	7
O_5	6	2
O_6	6	4
O_7	7	3
O_8	7	4
O_9	8	5
O_{10}	7	6



Compute the absolute error criterion [for the set of Medoids (O2,O8)]

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - O_i| = (|O_1 - O_2| + |O_3 - O_2| + |O_4 - O_2|) +$$

$$(|O_5 - O_8| + |O_6 - O_8| + |O_7 - O_8| + |O_9 - O_8| + |O_{10} - O_8|)$$

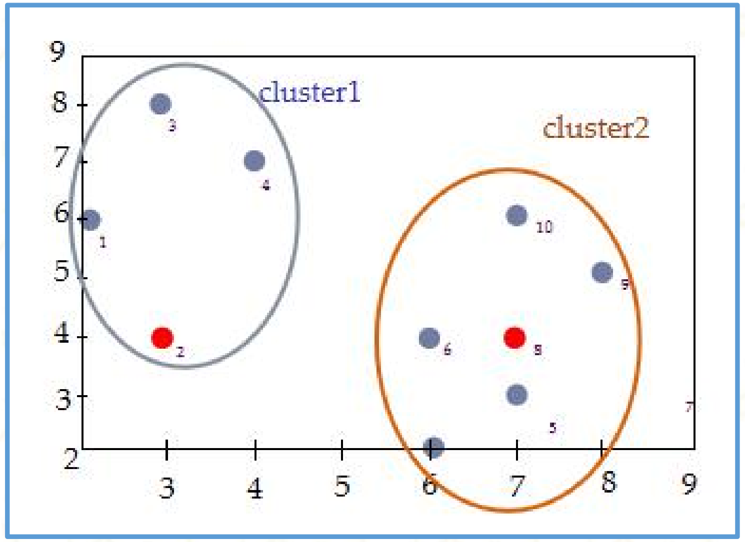
The absolute error criterion [for the set of Medoids (O_2,O_8)]

E = (3+4+4)+(3+1+1+2+2) = 20

How K-Medoids works?

Data Objects

	A_1	A_2
0_1	2	6
0_2	3	4
0_3	3	8
0_4	4	7
0_5	6	2
0_6	6	4
0_7	7	3
0_8	7	4
0_9	8	5
0_{10}	7	6



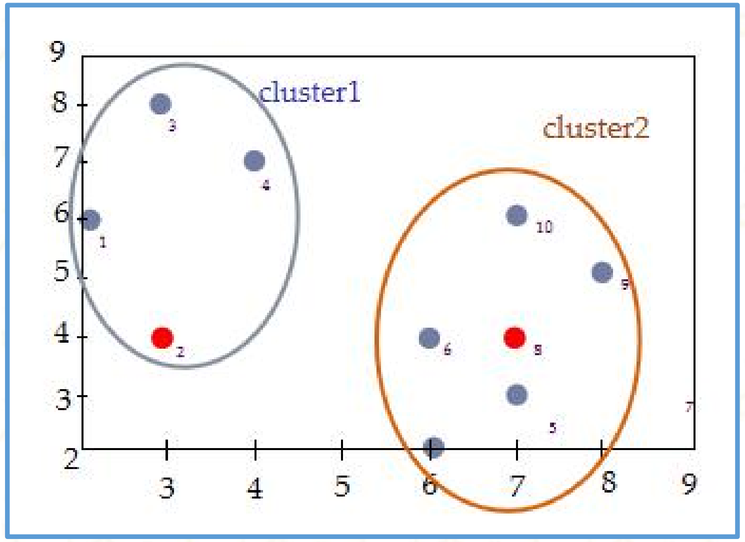
- Choose a random object 0_7
- Swap 0_8 and 0_7
- Compute the absolute error criterion [for the set of Medoids $(0_2, 0_7)$

$E = (3+4+4)+(2+2+1+3+3) = 22$

How K-Medoids works?

Data Objects

	A_1	A_2
0_1	2	6
0_2	3	4
0_3	3	8
0_4	4	7
0_5	6	2
0_6	6	4
0_7	7	3
0_8	7	4
0_9	8	5
0_{10}	7	6



→Compute the cost function

Absolute error [0_2 ,0_7] - Absolute error [for 0_2 ,0_8]

S=22-20

S> 0 => It is a bad idea to replace 0_8 by 0_7

Discussion on the K-medoids

- **Advantages:**
  - It is simple to understand and easy to implement.
  - K-Medoid Algorithm is fast and converges in a fixed number of steps.
  - PAM is less sensitive to outliers than other partitioning algorithms.

- **Disadvantages:**
  - It may obtain different results for different runs on the same dataset because the first k medoids are chosen randomly.
  - PAM algorithm for K-medoid clustering works well for dataset but cannot scale well for large data set due to high computational overhead.

https://www.geeksforgeeks.org/ml-k-medoids-clustering-with-example/

Lets Practice!

DigitalSkola
Uncover The World Of Digital Skills With Us

Thank YOU