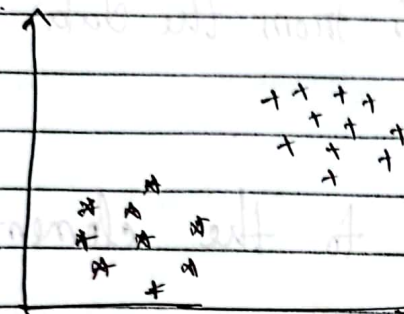


Unsupervised learning.

K-mean's clustering.

Elbow method
→ find the right
value of "K"

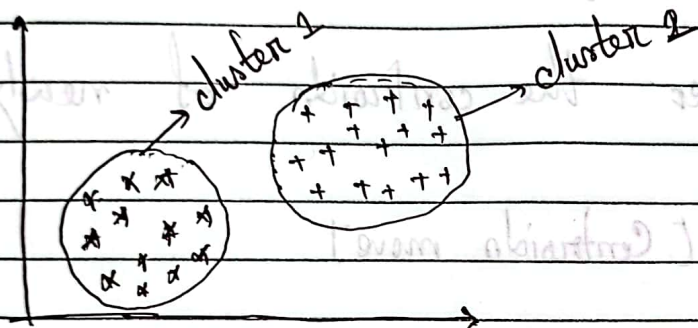


→ Manhattan Distance

→ Eucledian Distance

→ Elbow method

Select "K" value



① K-value → Centroid

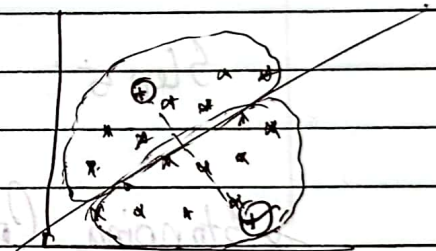
② Initialized the centroid Randomly.

③ Select the group (based on nearest distance)
and find the mean.

④ Update the centroid and move the position.

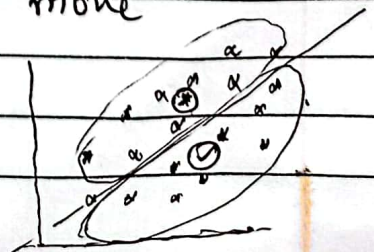
(based on distance)

iterate and untill any centroid is not move



let's consider
K=2 centroid.

[shakil]



Step 1: Choose the number of clusters "K"

Step 2: Select "K" Random points from the data as 'Centroid'

Step 3: Assign the all points to the closest cluster center. [group]

Step 4: Re-compute the centroids of newly formed cluster.
[Centroids move]

Step 5: Repeat 3 and 4.

⇒ Stopping Criteria for K-mean's clustering

1. Centroids of newly formed cluster do not change.

2. Points remain the same cluster.

3. Maximum number of iteration Reached.

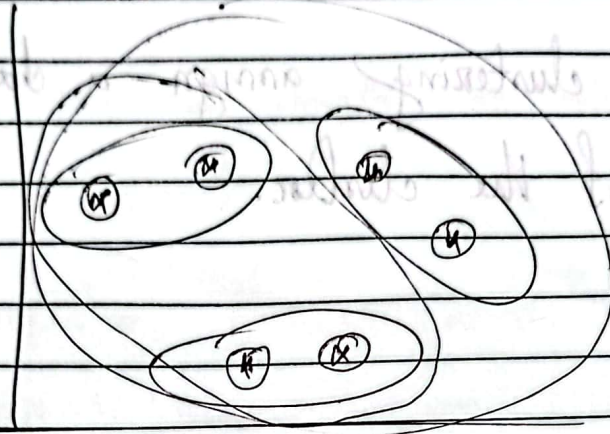
⇒ Types of cluster.

→ Hard cluster : Each data point either belongs to a cluster completely or not.

OR, A Hard clustering assign a data point to be in each of the cluster.

1. Assign all data point as a cluster or.

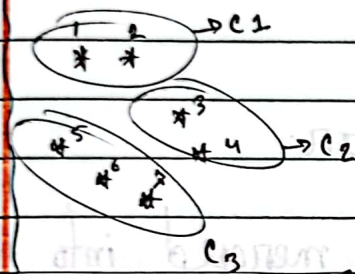
2. Two Nearest cluster are merge in to one cluster.



Clustering Type

→ **Hard cluster**: Each data point either belongs to a cluster completely or not.

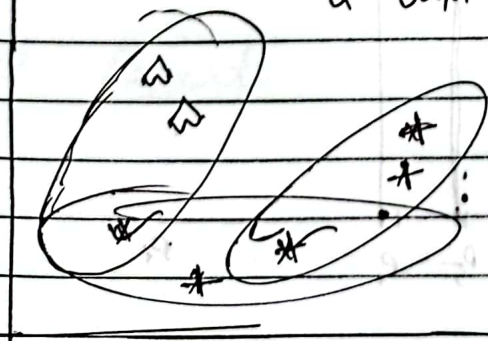
→ A Hard clustering assigns a data point in one and only one of the clusters.



Every node belongs to only one cluster.

→ **Soft cluster**: Instead of putting each data point into separate cluster, a probability or ... likelihood of that data point to be in those clusters is assigned.

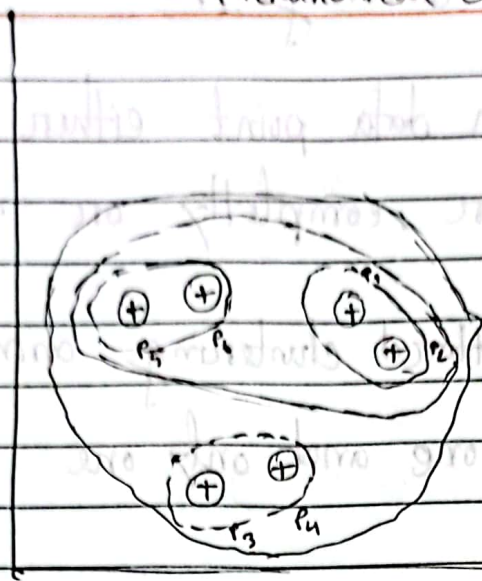
→ Soft clustering gives you the probability of a data point to be in each of the clusters.



Every node belongs to several clusters with a fractional Degree.

Best Explain ML A-2
course

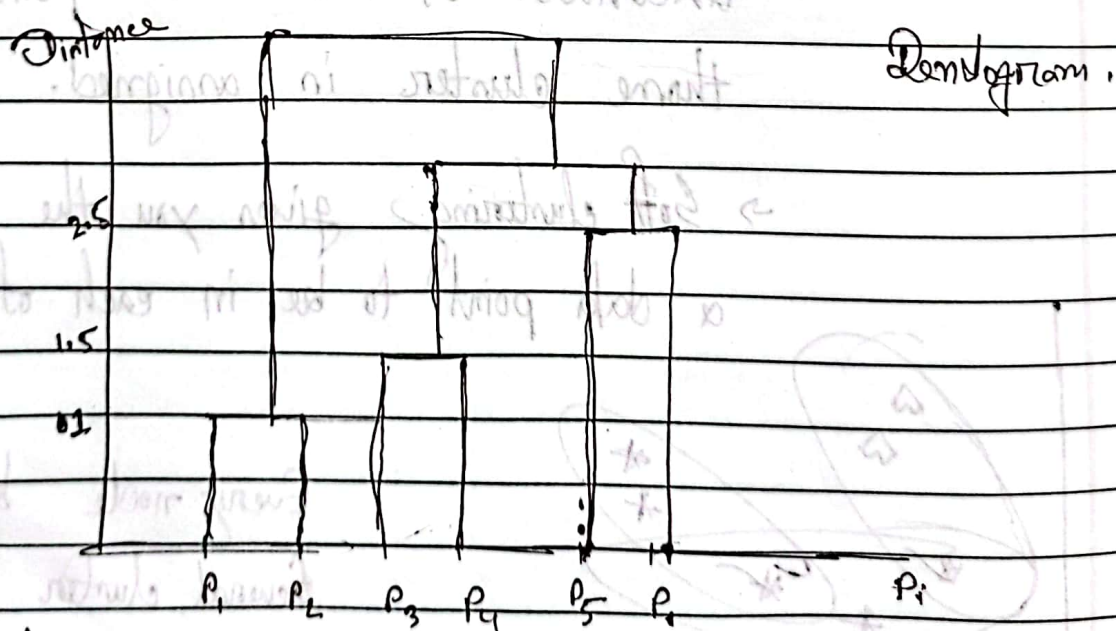
Hierarchical clustering



→ Assign each data point as a cluster.

→ Then two nearest cluster's are merged into the same cluster.

→ In the end, this algorithm terminates where there is only a single cluster left.



Kinda

Choose No. of cluster.

→ largest Distance →

No of vertical lines in the dendrogram cut by the horizontal line that can transverse the maximum distance vertically without intersecting a cluster.

→ Make each data point a single-point cluster - That forms N cluster.

→ Take the closest data points and make the one cluster.

→ Take the two cluster of closest and make them into one cluster

→ Repeat until there is only one cluster.

Curse of Dimensionality.

↓
Dimension / Feature / Attribute

Size | No. of bedroom

more information.

↓↓↓↓↓

? Price

M_1

M_2

M_3

M_4

2 feature

5 feature

10 feature

200 feature

Accuracy₁

< Accuracy₂

< Accuracy₃

< Accuracy₄

Threshold
value 200 feature

M_5

200 feature

M_6

4000 feature

Accuracy₅

M_7

10000 feature.

Accuracy₆

→ Than an increase the number of feature Accuracy is "less" than previous.

Accuracy₄ > Accuracy₅

or Accuracy₅ < Accuracy₄

⇒ if feature increase exponentially, our model being Confused.

that is curse of dimensionality.

Accuracy → → → → → After some time Accuracy
← ← ← ← ←
Decreasing.