

Introduction to Unsupervised Machine Learning :

We have a dataset given for Suppose, the data set containing only input label (ie there is no output label) then we have to do Unsupervised Machine learning.

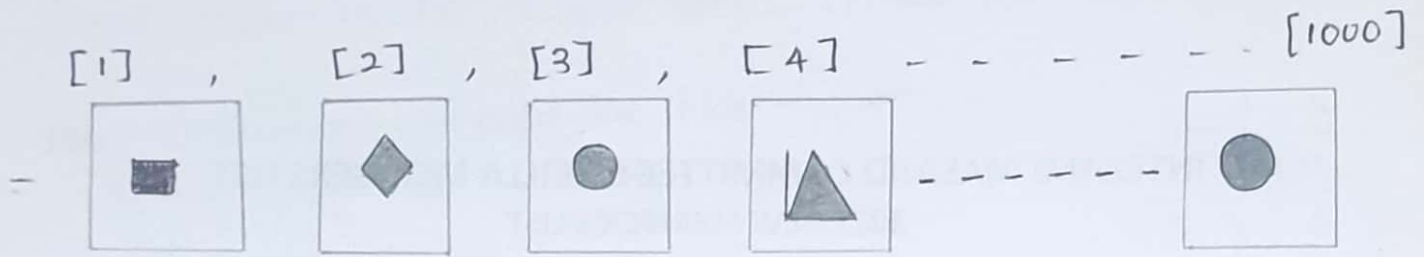
In previously we are using Supervised machine learning techniques, where we have input and output label present on it. if the Output datapoint containing numerical or real valued feature we are going to do regression task, if it containing Categorical then we go to classification tasks, in which we are predict the Output on Unseen or future data point. but it is not in case of Unsupervised machine learning technique.

Ex : (x_1, x_2, x_3, y) ; (x_1, x_2, x_3)

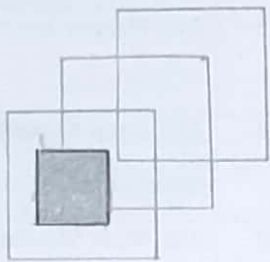
Supervised machine learning ; Unsupervised machine learning.
where label data associated with ; where label data not associated
input variables with input variable.

Task :- In Supervise technique Our task is to predict the Output on future data points., but in Unsupervice. Our task is to group. or collect the similar kind of data points generally called as cluster, the process is called as Clustering.

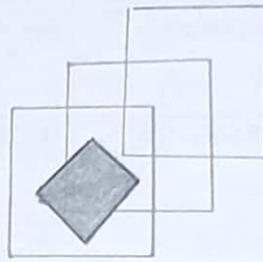
Ex: Let Consider a deck of cards, where we have different kind of symbols present on it, if we use a technique called as unsupervised machine learning, it generally grouped all the similar kind of cards into one group. we called it as a cluster. as shown in figure -



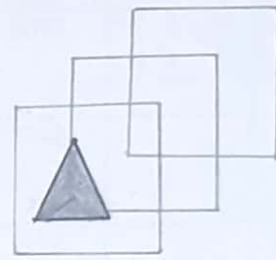
Task



cluster - I



cluster - II



cluster - III

Trying to recognising the pattern and grouped them on similar data point, ie forming a "cluster".

So In Supervise technique we are predicting., In Unsuper-

-vise we are doing "Data Mining".

In Supervise - $D_n = \{ (x_i, y_i) \mid x_i \in \mathbb{R}^d, y \in \mathbb{R} \}$

↪ Regression Task

$D_n = \{ (x_i, y_i) \mid x_i \in \mathbb{R}^d, y \in \{-1, +1\} \}$ ↪ classification task

Similarly in Unsupervised Technique —

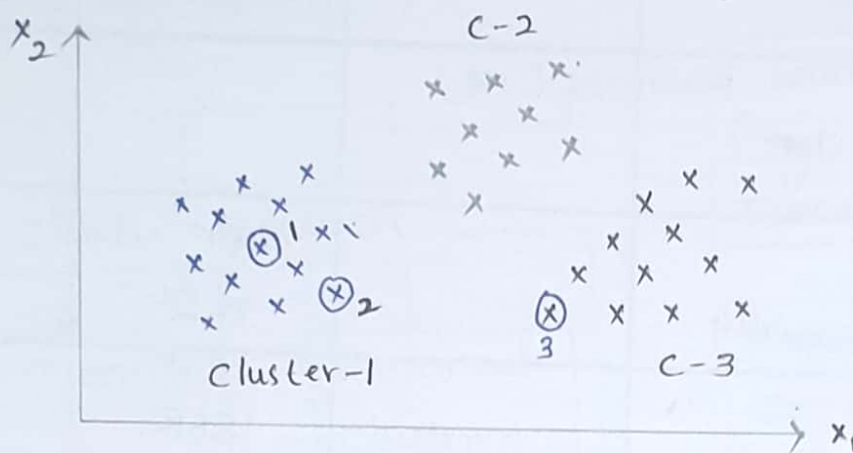
$$D_n = \{ (x_i) \mid x_i \in \mathbb{R}^d \}$$

Dimensions (columns) \rightarrow

Input variable \leftarrow

Real value \rightarrow

for Suppose, if we consider only '2' columns i.e. $x_i \in \mathbb{R}^2$ that should be rows containing having '2' features which are input, there is no output label. if we plot a diagram (scatter) the resultant should be like figure —



| | 1 | 2 |
|---|-------|-------|
| | x_1 | x_2 |
| 1 | 1 | 6 |
| 2 | 2 | 7 |
| 3 | 3 | 8 |
| 4 | 4 | 9 |
| 5 | 5 | 0 |

Row \leftarrow
Column \leftarrow

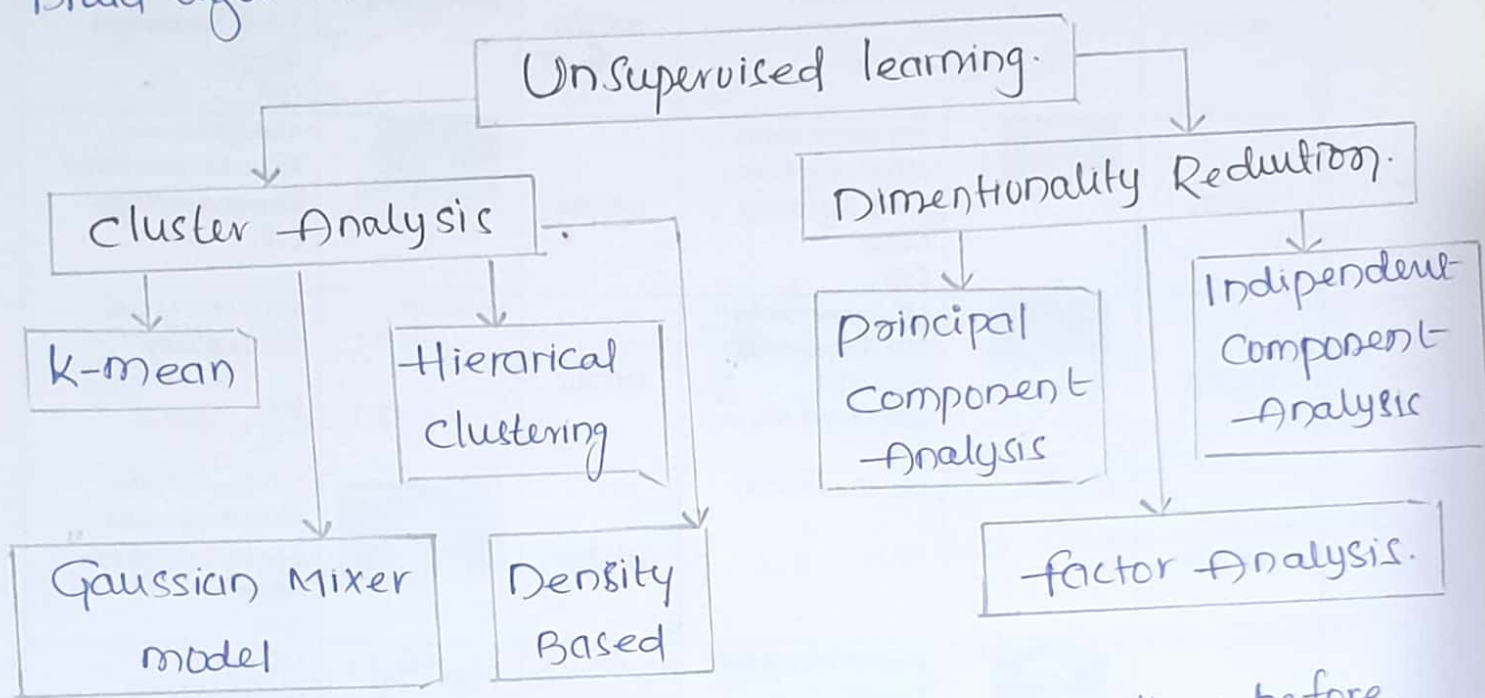
if we observe the scatterplot, we can say that there is forming '3' clusters nothing but similar kind of datapoints groups are formed, so here we can say this data points in clusters are similar, for that we have to calculate distance b/w them. Consider, Euclidean distance if we observe the distance between point 1 and point 2 they are much closer to each other, as if we consider point (1) to (3) or point (2) to (3) in both cases there is large distance b/w them. by this we can say this are from different cluster or dissimilar.

data points. and few more things we have to consider.

- ① points to be in a cluster should be close together.
- ② points in different cluster should be far away as possible.

Algorithms in Unsupervised machine learning. :-

We mainly classify them into two categories, then we have broad algorithms in each category. are mentioned below.



So here we discuss clustering and its algorithms before dive deep into this we discuss some applications of it.

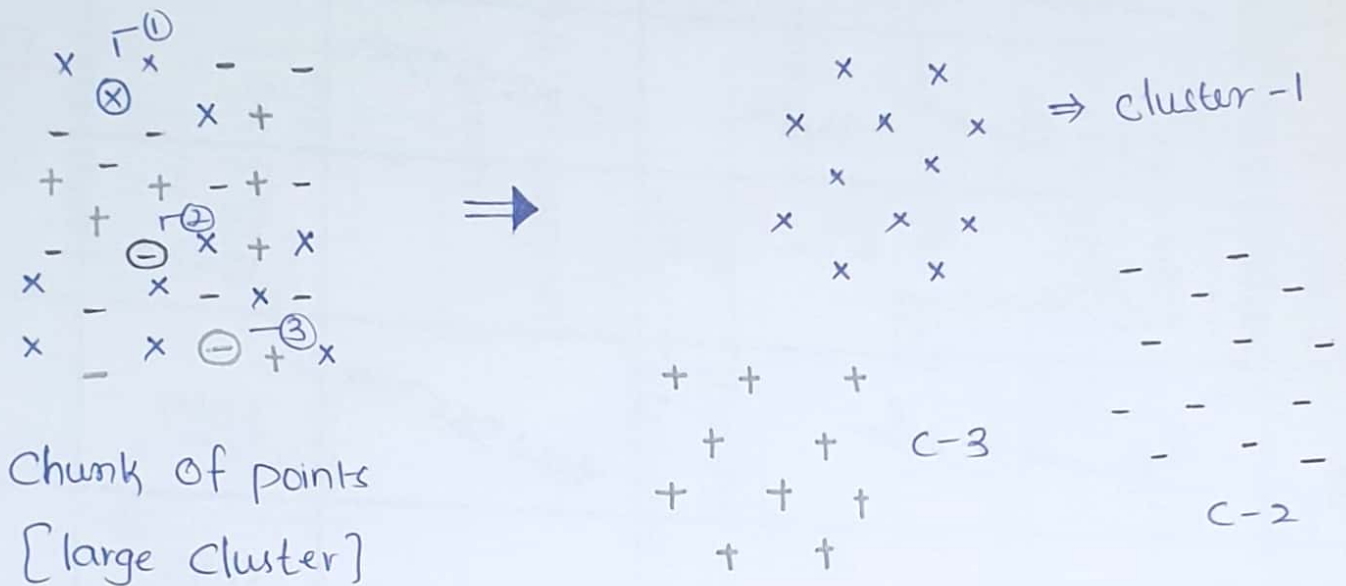
Applications of clustering :-

- ① The main Application of it is used in E-commerce.

In E-commerce - Task is to group similar clusters based on their purchasing behavior. Customer behavior can include -

- 1) How much money they spent.
- 2) Kind of Credit Card.
- 3) Kind of products they buy.
- 4) Geographical Area etc. like Consider.

E-commerce companies like "Amazon, Flipkart, Swiggi," etc.



So In E-commerce generally we divided large customers i.e. population, based on their purchase behavior we divided them into groups, where similar kind of data points are presented points come into one cluster, based on it we form three clusters.

② Image Segmentation. Is done by clustering where we are grouping of similar pixels, after grouping we typically, applying ML Techniques to perform object detection.

Regression

