04 December 2018
18:53

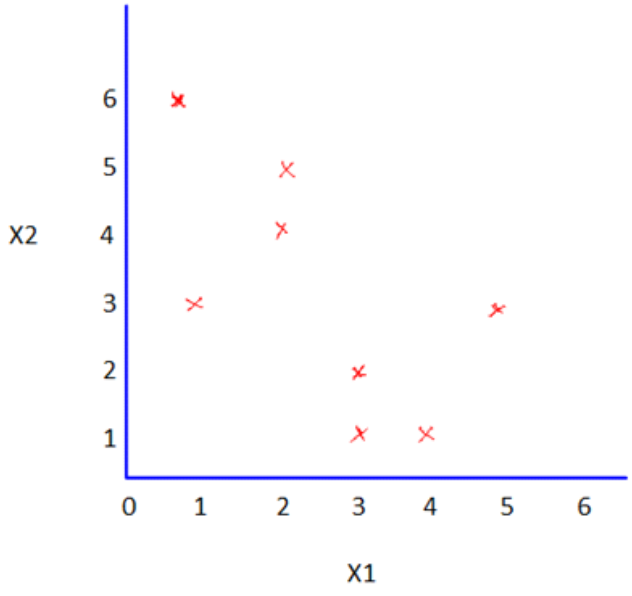
- CLUSTER ANALSYIS

- Suppose we have some dataset which contains of huge number of observations and there is no Y variable.
- So, How you are going to work on those scenarios?

- To over come this situation, we need to create a 'Y' variable
- First we need to make partition the dataset
- How it is partitioned?
- Using Clustering as a technique we are going to partitioned the data in to two or more sets based on the how data was collected.

- Example:

X1	X2
1	6
3	1
5	3
2	4
3	2
1	3
2	5
4	1



Clustering is a data segmentation technique that divides huge datasets into different groups on the basis of similarity in the data.

It is a statistical operation of grouping objects.

Hence, the resulting groups are called clusters.

Clusters have the following properties.

- We find them during the operation and their number is also not always fixed in advance.
- They are the combination of objects having similar characteristics.

Applications
Marketing
Retail
Medical Science

So, there will be no 'Y' variable here.

So it is a Unsupervised Learning We need to make data in to clusters and then create your 'Y' variable.

Hence it is also called as Semi-Supervised Learning.