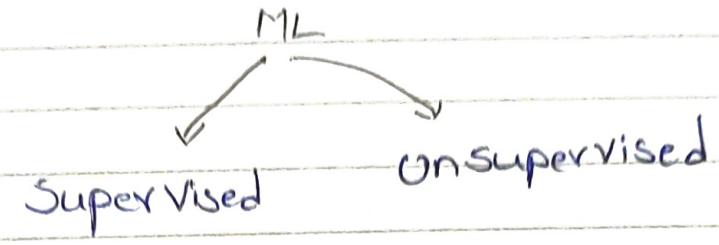
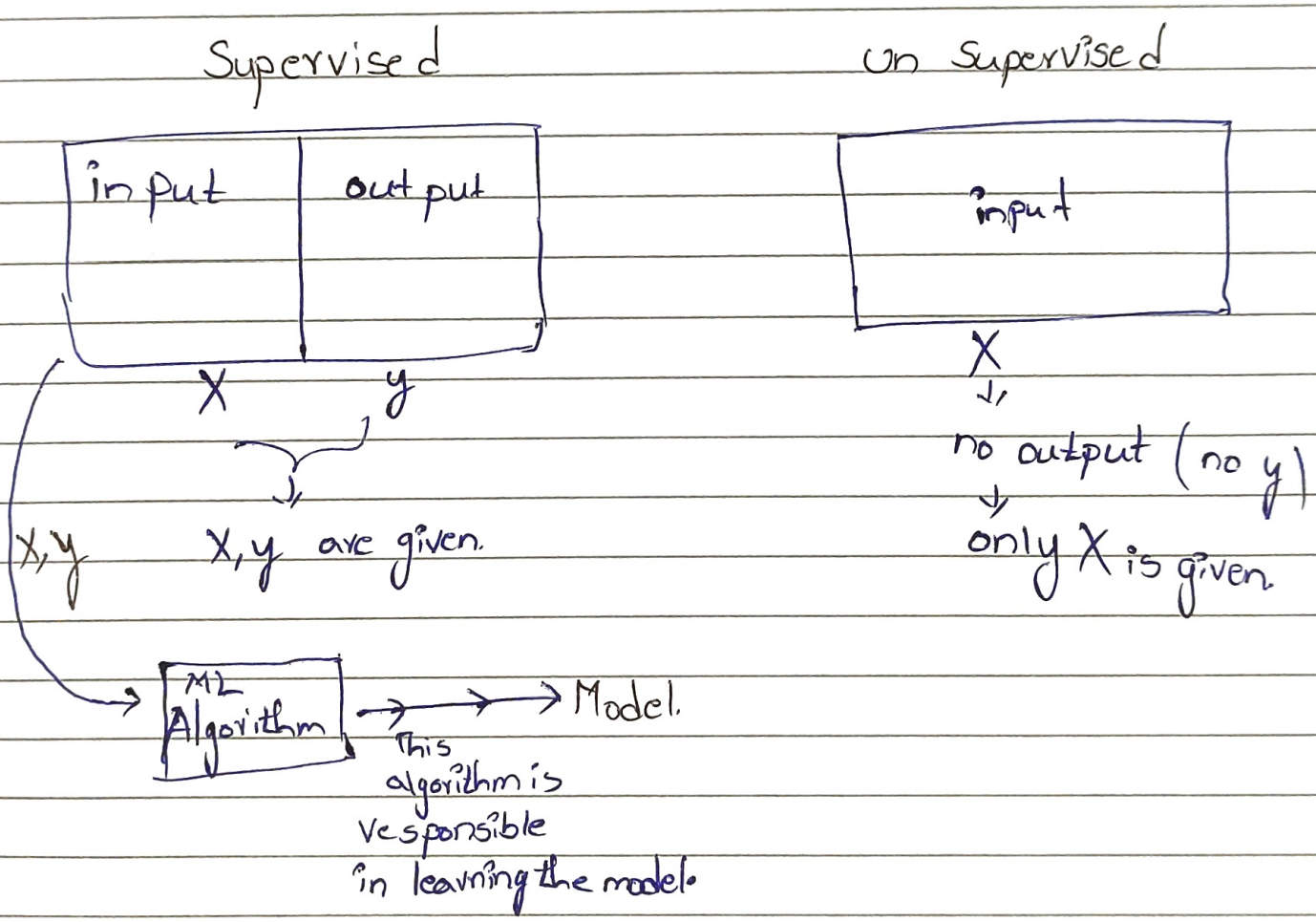


Q) How to identify whether a problem is unsupervised learning?

Ans) There is no output label associated with each input

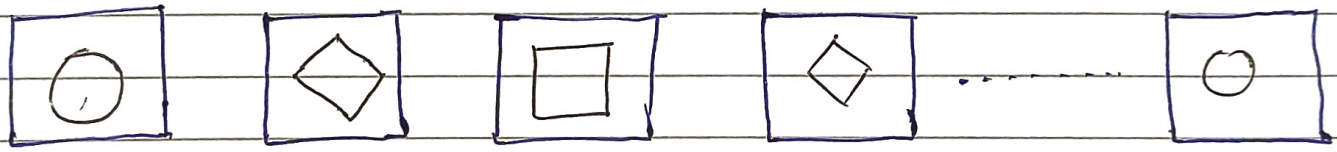


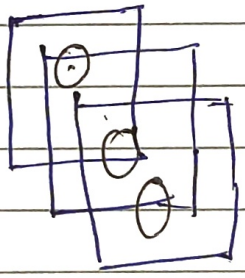
↳ This model will predict Future/unseen data.

→ In classification, the output column will tells us whether we need to perform Classification / Regression.

→ If output column is not there then we need to go with unsupervised learning.

→ Iam having 1000 Cards which consists of different objects So, we will group similar cards.





← Trying to group similar datapoints together

↳ There is nothing like prediction in unsupervised learning. we are trying to recognize the pattern & we will group it to together.

Data mining:→

↳ Collecting the data & recognizing the pattern

Ex:- Extract gold (in mining) → Looking for gold

↳ Similarly extract some useful information from the data.

↳ Looking for pattern

Dataset Definition:→

$D = \{ (x_i, y_i) \mid x_i \in \mathbb{R}^d, \; y_i \in \mathbb{R} \}$

- x_i → Input
- y_i → output (or) target
- $x_i \in \mathbb{R}^d$ → Input is having "d" no. of columns
- $y_i \in \mathbb{R}$ → "$\mathbb{R}$" → Real Valued ~~feature~~ → Continuous Value → Regression

Dataset

$D = \{ (x_i, y_i) \mid x_i \in \mathbb{R}^d, \; y_i \in \{-1, +1\} \}$

- x_i → Input
- y_i → output (or) Target
- $x_i \in \mathbb{R}^d$ → Input is having "d" no. of Columns.
- $y_i \in \{-1, +1\}$ → y_i is having either −1 (or) +1 → Classification.

$D_n = \{(x_i) \mid x_i \in \mathbb{R}^d\}$ → For un supervised learning.

→ In data set each row is a Vector

→ So, here x_i is a Vector

→ x_i is having 'd' no. of Components ("d" dimensional like 2-D, 3-D, --)

→ x_i is Real Value

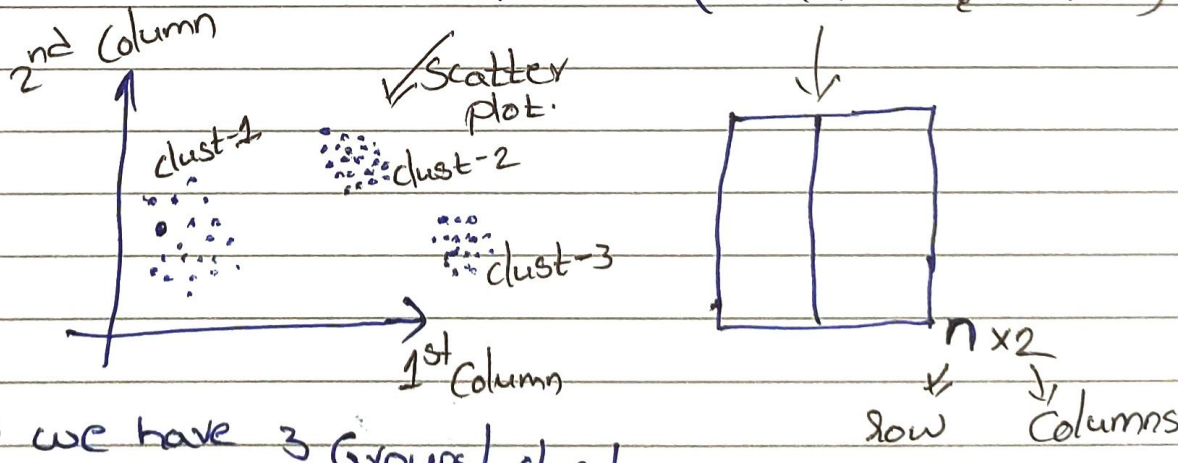
Ex!→ I want to store a single Value in "a" then

$a = 5$

→ If I want to store two values in 'a' then

$a = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ → Vector. → Its 2-D

↳ $x_i \in \mathbb{R}^2$ (2 columns & n rows)



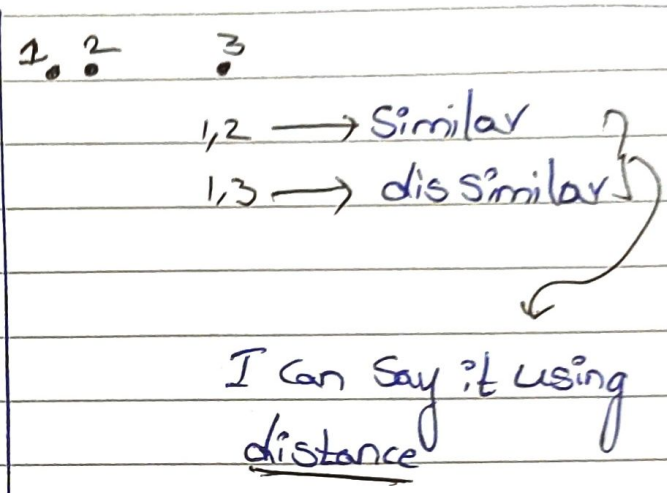
Observation ⇒ we have 3 Groups/ cluster

Cluster:→ Group of Similar datapoints.
↓
distance

*) points in a cluster should be cluster close together
↳ (Intra cluster distance → small)

*) points in two different clusters should be far as possible

otherwise two clusters will combine.
↳ Intercluster distance → high



Clustering is UnSupervised learning

Intra → With in
Inter → Between

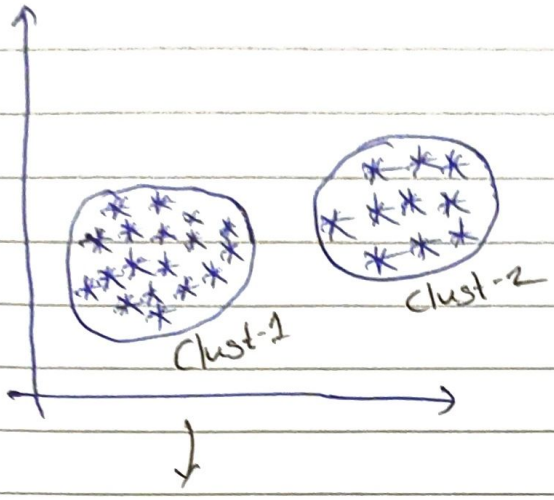
↓
K-means, K-means++, Hierarchical Clustering, DBSCAN.

classification
↳ Task → seperate different classes.

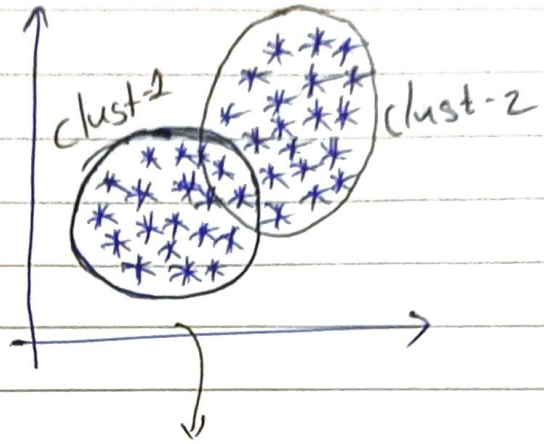
Regression
↳ Task → ~~try to~~ Best fits the data.

Q) what if, if a customer/datapoint belongs to all the clusters.

Ans. It means clusters are overlapping clusters are not properly defined.



clusters are properly defined



clusters are not properly defined. (It is overlapping)

↓ In this area

We need to do more Feature Engineering

↓

So that this clusters stops overlapping eachother.

In Real world → We will see many atimes that clusters are overlapping

↓

In that case there are many other algorithmns we can look at GMM

↓ └) probablistic approach → Gaussian Mixture Model approach

↓

It will assumes Normal Distribution of all the clusters

Applications of unsupervised Learning!

1. e-Commerce
   ↳ jio, airtel → Some customers are not happy with the services provided by Airtel so they will move to other (Jio)

   ↓

   When we do unsupervised learning on top of this data then we know this group of customers are not happy, happy....

   ↓

   So Rather than providing offers everybody, if they give discount to unhappy customers then it is good for them

   ↓

   Customers won't switch to Jio.

   (Anyway happy customers will use Airtel (won't "" "") eventhough you give offers/Not)

   ↓

   That's why we use to get messages in whatsapp. stating 50Rs discount,....

2. Image Segmentation.

↳ Grouping of similar pixels

↳ after Grouping, we typically apply ML techniques to perform object Detection.

↳ Grouping of similar pixels

↓ in terms of ML

Group nearest data points

↳ Vector.

(Self Driving Car ↳ It has object Detection)

↓

It Came into picture in recent Years

→ But this Segmentation (self driving Car) has been used in 1980's by US Defense. But not Came into this world because of Computional power.

→ Because this Vectors ~~are not having~~ need high Computational power.

↳ Now only laptops, Computers are having high

Computational power (like SSD). and it was available to everyone easily.

→ Thats why now it is getting popular.