Machine learning

8k. Sandhani
Cse-ai

Unit-1

1. What is Machine learning? Explain types of Machine learning?

A) **Machine learning :-**

⇒ Introduced by Arthur Samuel in 1959

⇒ It is subfield of AI. It develop by algorithms which allow a computer (Machine) to learn from the data and past experience by it own.

⇒ The output of Machine learning is Model

⇒ In Machine learning 20% data for test and 80% for train.

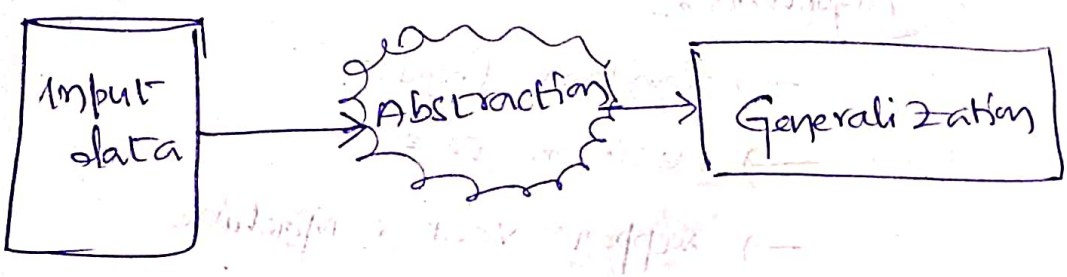
⇒ Machine learning process

1. Data input :- past data is used for future decision

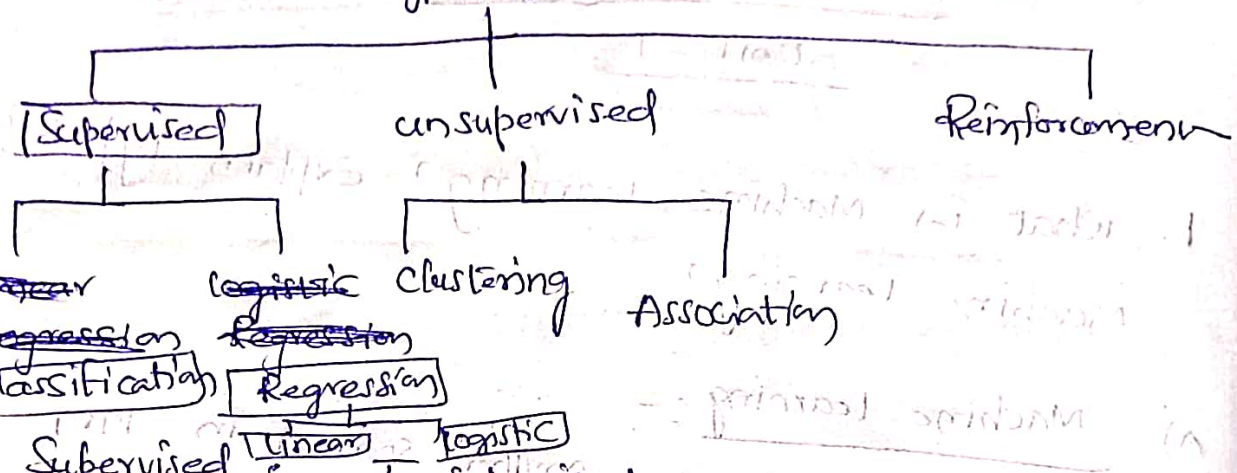
2. Abstraction :- it is preprocessing the data which means all duplicate, unorganized will removed and makes perfect data.

3. Generalization :- it is the combination of ML algorithm + user expectation + accuracy + Risk Management

ML Representation :-



Types of ML



1. Supervised :- In supervised learning, the machine has to train by labelled data.

The labelled data describes all feature of anything (person, object, place, whether, etc.)

⇒ The Main goal of supervised learning technique is to map input variable (x) with the output variable (y).

i. :- classification :-

classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "Yes" or "No", Male (or) Female, Red (or) Blue etc.,

⇒ It predicts the categories present in data set.

Ex :- Spam detection,
Email filtering etc.,

Algorithms :-

→ Random forest
→ Decision tree
→ Support-vector Machine
→ Logistic Regression

Regression :- It is used to solve the Regression problems in which there is an linear Relation between input and output.

⇒ These are used to predict continous output variables, such as Market trends, weather predictions etc,

Linear Regression :-

⇒ Linear Regression is $y = mx + b$

y = dependent variable,

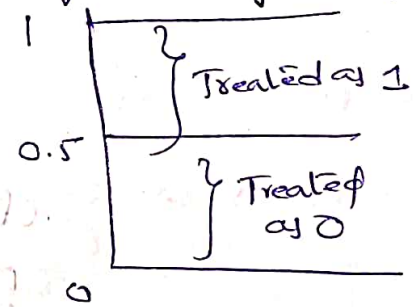
m = slope, x = independent, b = intercept

⇒ We cannot predict more accurate values in linear

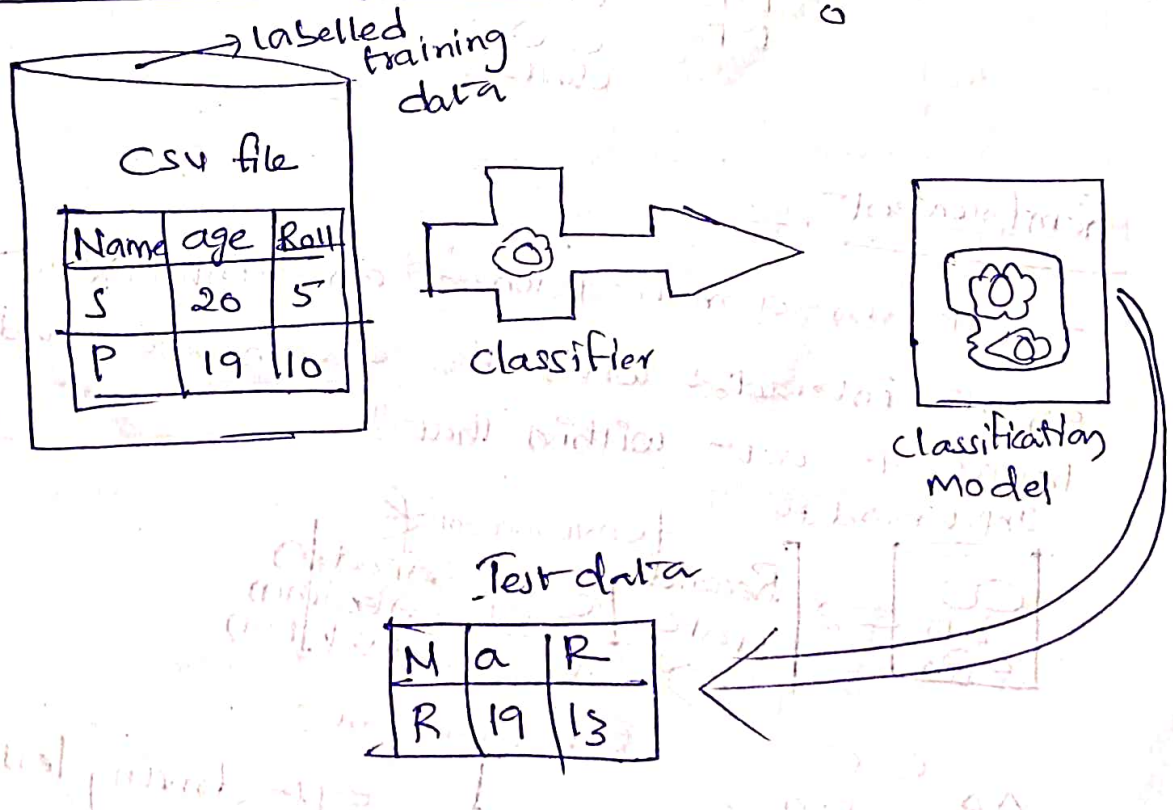
logistic Regression :-

⇒ From linear we find logistic by using sigmoid function

$$f(x) = \frac{1}{1+e^{-y}}$$



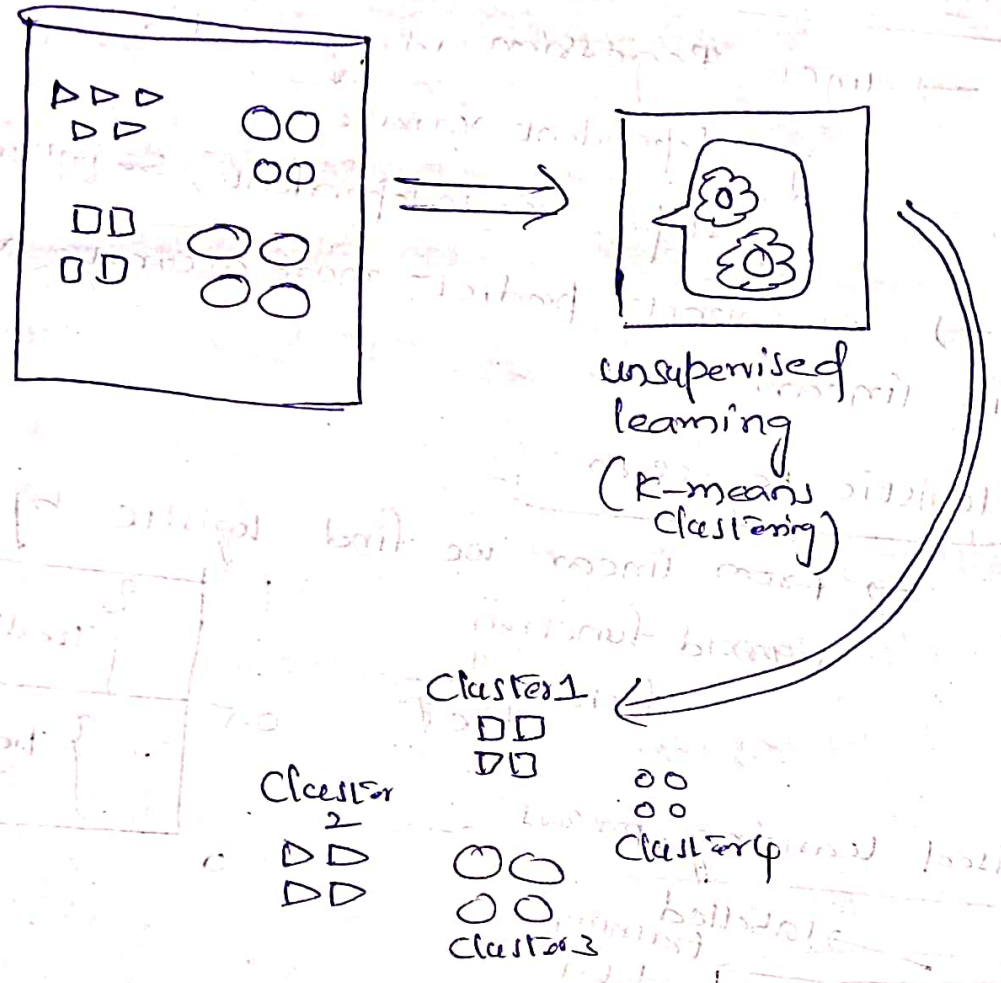
Supervised learning process :-



Un Supervised Learning :-

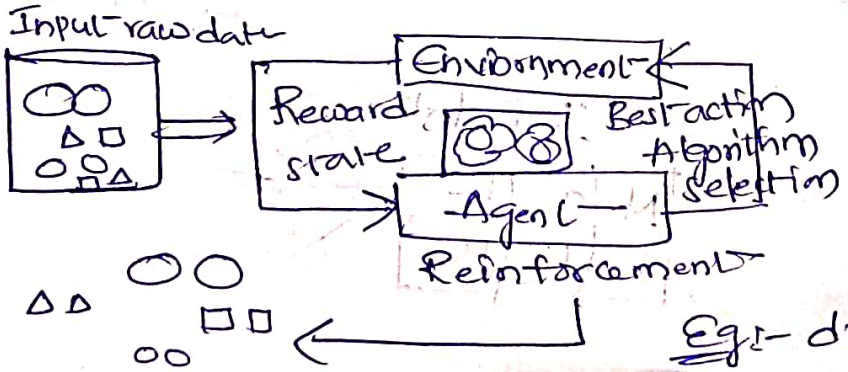
⇒ There is no labelled training data to learn from and no prediction to made

⇒ The Main objective is dataset as input and try to find natural groupings (or) patterns within the data elements (or) records.



Reinforcement :-

⇒ It is a method where an "intelligent agent interacts with the environment, and learns to act within that"



Eg:- driving less car

2) What are the problem not to be solved using Machine Learning?

A) problems not to be solved by using Machine learning are :-

1. Reasoning power
2. Contextual limitation
3. Scalability
4. Regulatory Restriction for Data in ML
5. Internal working of Deep learning

⇒ Machine Learning should not be applied to tasks in which humans are very effective (or frequent human intervention is needed.

Eg :- air traffic control is very complex task

⇒ In very simple tasks which can be implemented using traditional programming paradigms, there is no sense of using Machine learning.

Eg :- Simple rule driver / formula based application like price calculator engine

1. Reasoning power :-

⇒ The ML has not mastered sucessfully is reasoning power

⇒ Algorithms available today are Mainly

focused on specific use-cases and are narrowed down to an application

⇒ They cannot think as to why a particular method is happening that way.

Eg :- An image Recognition algorithm identifies apples and Organes in given scenario.

But it cannot say if the apple is good (or) bad, or why is that fruit an apple (or) orange

⇒ Mathematically all of this learning process can be explained by the trainer

2. Contextual limitation :-

→ If we consider the area of Natural language processing (NLP), text and speech information are there to understand language by NLP algorithms.

⇒ They may learn letters, words, sentences (or) even the syntax, but, the algorithms don't understand the context of the language used

⇒ ML does not have an overall idea of the situation

3. Scalability :-

⟹ ML algorithms are depens on data as well as it scalability

⟹ Data is growing at an enormous rate, and has many forms which largely affects the Scalability of an ML project.

⟹ Algorithms cannot do much about this, unless they are updated constantly for new changes to handle data.

⟹ There is where ML regularly requires human interventions in terms of scalability and remains unsolved mostly

4. Regulatory Restriction for Data in ML :-

⟹ ML usally needs considerable amounts of data in stages such as training, cross-validation

⟹ Sometimes, data includes private (sensitive) as well as general information

⟹ Collecting and maintaining these types of data are complicated

⟹ Because, the risk of the wrong usage of data, espically in critical areas such as medical Research, health insurence etc.

⟹ If we restrict the amount of data they there is chance to lost required data.

5. Internal working of Deep learning -

→ Deep learning (DL) now powers applications such as voice recognition, image recognition and etc., through artificial neural networks

→ But the internal working of DL is still unknown and yet to be solved.

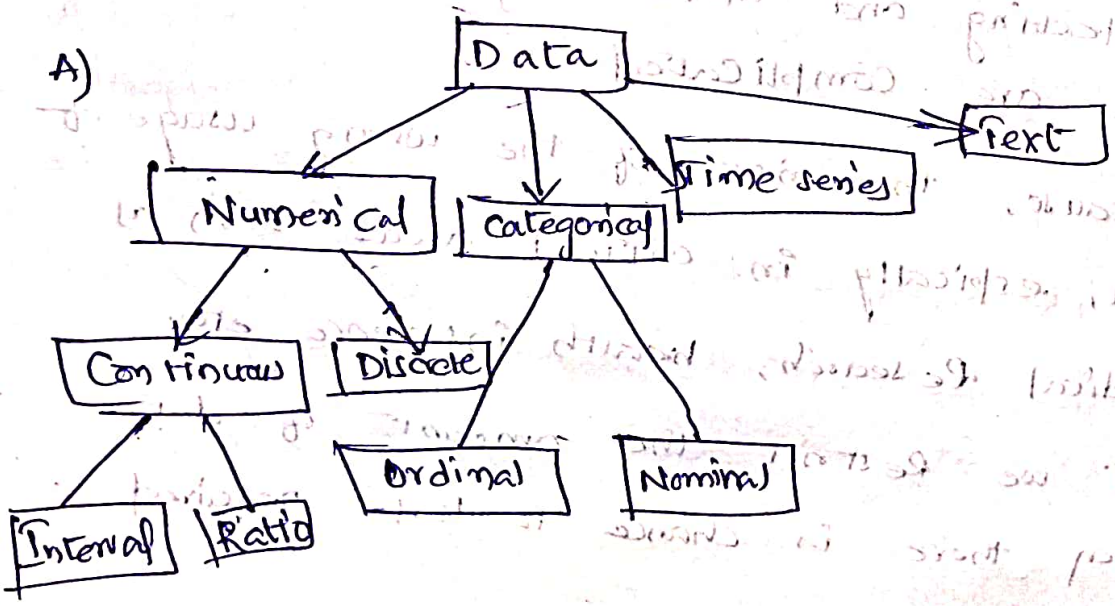
→ Millions of neurons that form the neural networks in DL increase abstraction at every level, which cannot be understood at all.

→ Advanced DL algorithms still confuse the researchers in terms of its working and efficiency

→ This is why deep learning is dubbed a 'black box' since its internal agenda is unknown

3) What are the basic types of data used in Machine learning?

A)



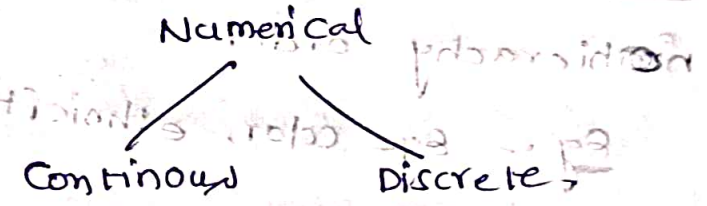
Data can come from in many forms, but Machine learning models rely on four primary data types.

They are

1. numerical data
2. categorical data
3. time series
4. Text data.

1. Numerical data :- Numerical data is mean anything represented by numbers (floating point (or integer)

Eg :- 0, 1, 1.0000, 20,



1.1 Continous :- It is an infinite number

Eg :- Height, weight, salary, temparature, interest rates

1.2 Discrete :- It is an finite optim

Eg :- shoe size, number of students

2. Categorical Data :- It is an textCorg Mat of words (or) labelled data.

Eg:- Eye colour, gender, blood type, ethnicity

Categorical Data

Ordinal | Nominal

2.1 Ordinal :- The data which is hierarchy called Ordinal

Eg:- Mood, Satisfication rating, pain severity, job grade, age group

2.2 Nominal :- The data which is no hierachy order

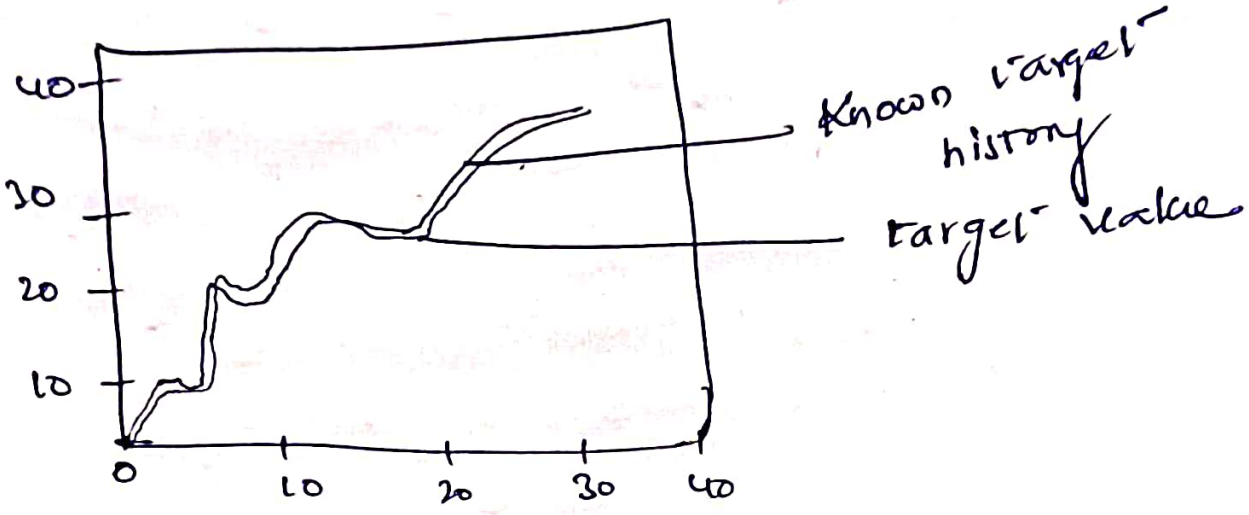
Eg :- Eye color, ethnicity

3. Time Series :-

Time Series is a sequence of data points indexed in time order.

⇒ It is a technique that forecasts target value based solely on a known history of target values

⇒ It is specalised form of regression.



Time series prediction

Text-data :-

Textual data comprise of speech and text-databases, text-corpora., and other metadata-added textual resources used for language and linguistic research.

⇒ Some Copora uses publishing dictionaries, grammer books, teaching materials, usage guides.