Machine learning Techniques (KCS-055)

Unit-01: Introduction of machine learning

- Machine learning
- Types of machine learning
- Well-defined learning problems
- Designing a learning System
- History of machine learning
- Introduction of ML approaches –
  - Artificial Neural Network
  - Clustering
  - Reinforcement learning
  - Decision Tree learning
  - Bayesian Networks
  - Support Vector Machine
  - Genetic Algorithm
- Issues in machine learning
- Data Science Vs Machine learning

Unit-01

Machine learning :-

- Machine learning is the process in which a computer gets knowledge to do a task by itself without external programs
- Machine learning is the subset of AI, in which computer software applications become more accurate without explicitly programs.
- Machine learning enables a machine to automatically learn from data, improve performance from experience and predict things explicitly.

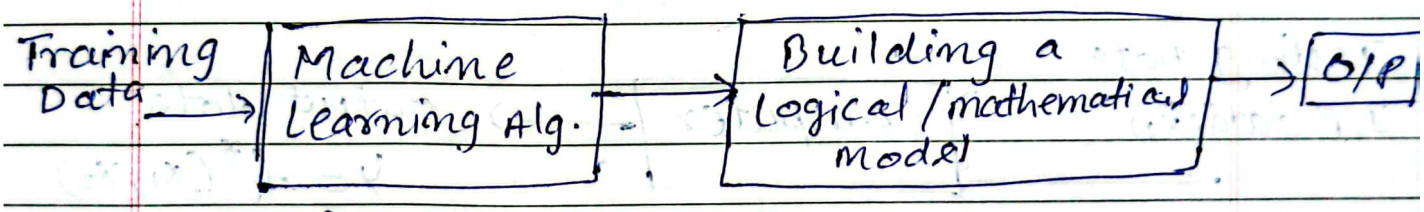
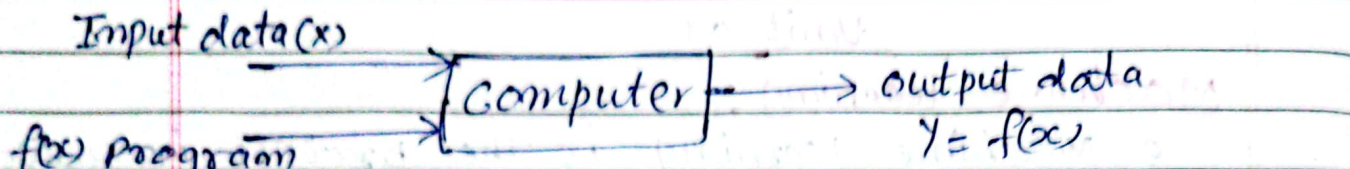


fig: Learning from Data.

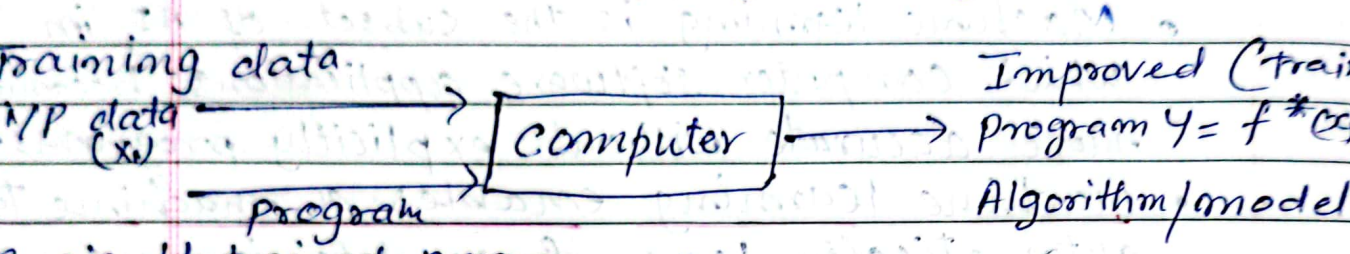
- Machine Learning algorithm learns from training data and then applies the learned knowledge on testing data.
- These learned model are used in real life problem solutions such as a business, e-commerce websites, social media platforms, stock market predictions or robotic driving car etc.

Traditional programming model and machine learning model

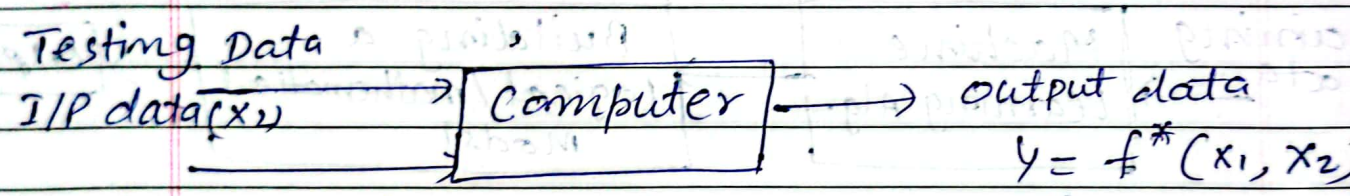
Traditional programming model and machine learning model as follows—



Traditional programming model



Basic Untrained program.
$f(x)$ (a) Training.



Final improved program
(Algorithm) $f^*(x_1)$
(b) Testing
fig: Machine learning model.

Well defined learning problems :-

- Problem that is well formulated and has clear objectives, input, data, output data.
- A well defined problem is necessary for effective algorithm development and evaluation.

Characteristics of well defined learning prob

- **Clear objectives** - The problem should hav clear objectives and goals that are well defined & measurable.
- **Input data** - The problem should have i/p data that is relevant to the problem, well structured and easily accessible.
- **Output data** - The problem should have o/p data that is well defined and related to objectives of the problem
- **Evaluation metrics** - The problem should have well defined evaluation matrics that can be used to measure the performance of the algorithm.

Example -

- Image classification
- Sentiment analysis
- Recommendation Systems
- checkers (Chess) game learning Problem
- Robort car Driving
- Handwriting recognition learning problem.
- Spoken words recognition learning probl
- New Astronomical structural learing problem.

Types of machine learning:—

- Supervised learning
- Un-supervised learning
- Reinforcement learning

Supervised learning :—

- Supervised learning is a type of machine learning, in which a computer algorithm learns to make predictions based on labeled data.
- Labeled data is made up of previously known input variables (features) and output variables (also known as labels).
- By analyzing patterns and relationship between input and out variables in labeled data, the algorithm learns to make decisions or predictions.

Examples:-

- Image and speech recognition
- Recommendation systems.
- Fraud Detection

Types of Supervised learning

- Classification tasks
- Regression tasks.

Difference between classification & Regression

Parameter	Classification Task	Regression Task
Discrete	classification is task of prediction a discrete ~~value~~ class label.	Regression is the task of Predicting a
Continuous		Continuous as a lite

Parameter	classification Task	Regression Task
2) Example	To classify an E-mail is spam or not	To estimate the price of a stock market
3) Multiclass or multivariable	A classification problem with two or more classes called multiclasses classification	A regression problem with multiple variables called multivariate regression problem

Un-Supervised learning :—

- Unsupervised learning in which, only i/p data is provided to the model.
- In unsupervised learning, the models are trained using unlabelled data.
- The goal of unsupervised learning is to find the hidden patterns in input data.
- Unsupervised learning methods are suitable when output variable (labelled data) are not provided.

Example –

- clustering task of customer segmentation
- product recommendation
- Friend suggestions on social media platform

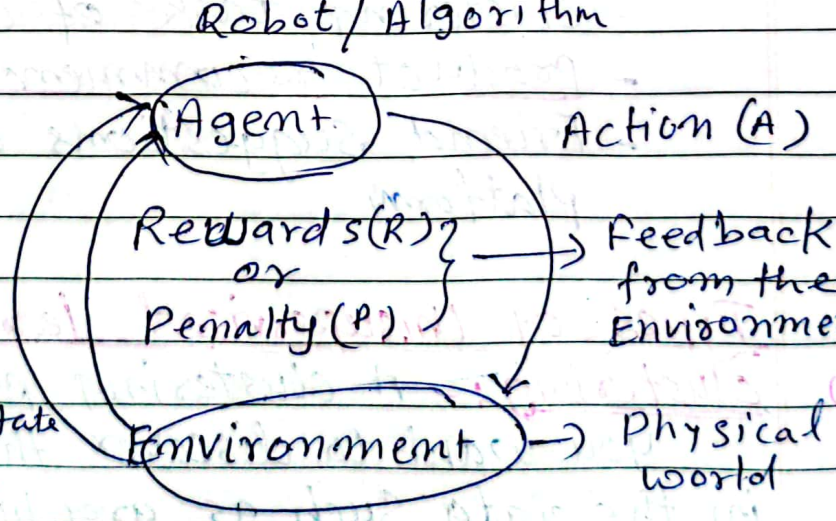
Types of Unsupervised learning :—

(1) Clustering :- A clustering problem is where you want to discover the inherent grouping in the data such as grouping customer by purchasing behavior.

(2) **Association** :— An association rule learning problem is where you want to discover rules that describe large portion of data such as – people that buy X, also tends to buy Y.

Reinforcement learning :—

- Reinforcement learning is a type of machine learning method in which the desired behaviour is rewarded and undesired behaviour is punished.
- In this learning, a software agent or hardware agent is used. This agent interacts with external environment by taking some actions.
- Depending upon some actions, some reward or penalty is given to the agent.
- Then agent learns from its rewards and penalty and improves its next action.



State (S)
Present Situation

- Policy = method to map agent's state to action

value f^n = future Reward.

fig: Reinforcement learning

Examples of Reinforcement :—

- Robotic Movement control
- Marketing and Advertisement
- Trading and Finance
- Industrial Automation.

Importance of machine learning :—

- The machine learning can be used to reduce the costs, improve quality of product/services detecting cyber security attacks, self driving car
- Machine learning can also be used widely in e-commerce website for products/services reccommedations like Amazon, Flipcart, Snapdeal
- Machine learning is also used to analyze hidden patterns in historical data.

Operation of machine learning Process :→

(i) choose and prepare a training set data.
(ii) Select an algorithm to apply training dataset.
(iii) Train the algorithm to build a new model.
(iv) Evaluate/Test model.
(v) Improve the model by training its parameters
(vi) Finally, used the improved model in real life applications.

Components of machine learning process :—

The learning process of a human or a machine can be divided into four components

(1) Data Storage :— Data is stored in hard disk memories
(2) Abstraction :→ The process of extracting Knowledge from stored data is called abstraction.

(3) Generalization :— Utilizing knowledge into useful action.

(4) Evaluation — Feedback to improve performa

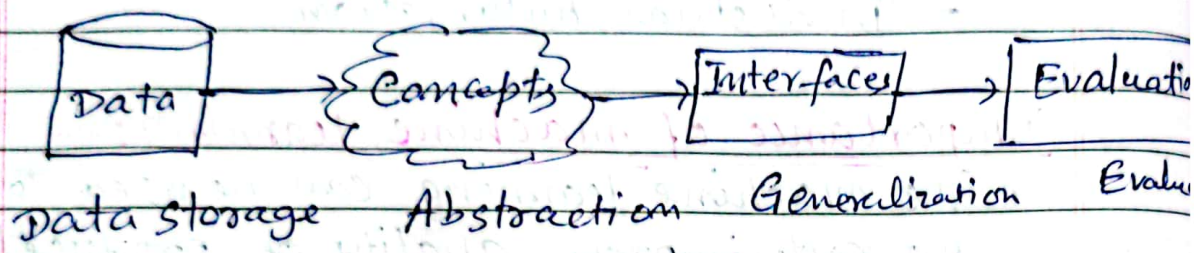


fig: Basic Components of learning Proces

Steps in machine learning process or steps in making ML model :—

(1) Gathering data / data collection.
(2) Preparing data / data preparation
(3) Choose model (algorithm)
(4) Training of algorithm (Model Training)
(5) Evaluation of model (Testing)
(6) Hyper Parameter tuning (Model improve
(7) Prediction (Use model in real life)

Application of machine learning :—

- Self driving car
- Web search result
- stock market Price
- Banking fraud Transaction detection
- Pattern recognition

Objective / Goal of machine learning:-

The primary purpose of machine learning is to discover patterns in data. After Knowing Pattern, the task is make predictions to solve the business problems.

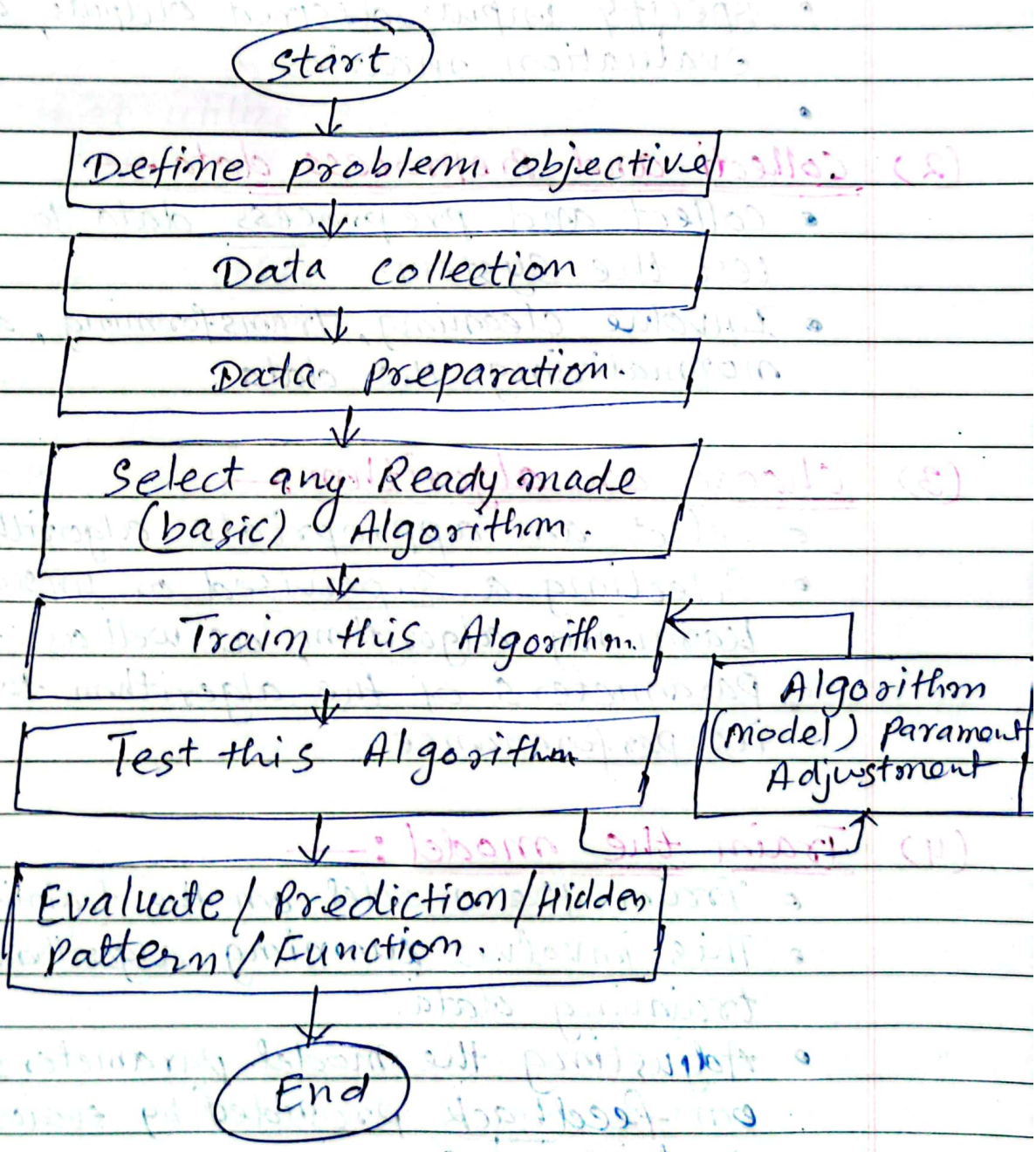


fig: Machine learning process workflow (operation / step of ML)

Designing a learning System :→

Designing a learning system involves several steps, including -

(1) **Defining the problem :-**

- Define the problem you want to solve.
- Specify inputs, desired outputs, and the evaluation matrics.

(2) **Collect and preprocess data -**

- Collect and preprocess data to train and test the system.
- Involve cleaning, transforming, and normalizing the data.

(3) **Choose an algorithm :-**

- Select an appropriate algorithm.
- Selecting a supervised or unsupervised learning algorithm, as well as tuning parameters of the algorithm to optimize its performance

(4) **Train the model :-**

- Train the model on the training data
- This involves running algorithm on the training data.
- Adjusting the model parameters based on feedback provided by evaluation matrics. ↓(performance measurement)

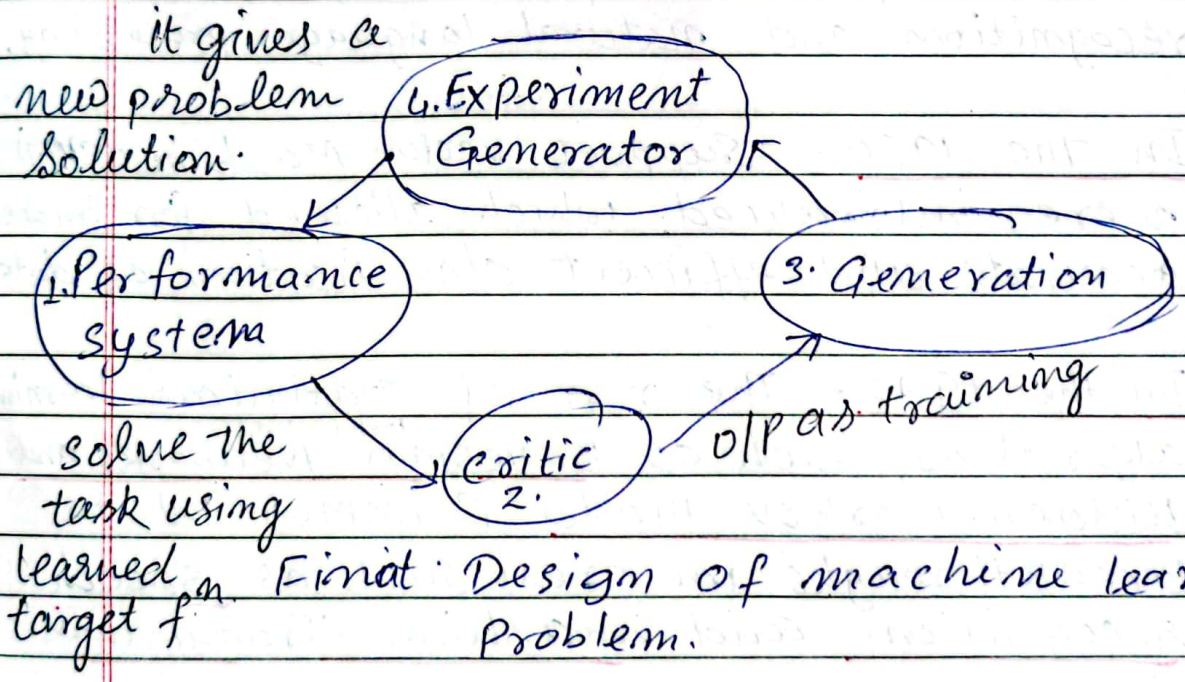
(5) **Test the model :-**

- Test to evaluate its performance

- Identify any issues or limitations with the model and suggest area for improvement.

(6) Deploy the model :—

- Deploy in a production environment
- Involve integrating the model into existing system or developing a new application that utilize the model.



Final Design of machine learning Problem.

1. Performance system —
It is the module which solves the given performance tasks by using learned target function.

2. Critic — It takes input as history of each step. The critic gives the output as training example and produce output as training example of target fn

3. Generalize — take i/p as training example of the target function.

4. Experiment Generator — takes the current input as current hypothesis and gives o/p a new problem to expl-

History of machine learning –

1. 1950 – 1960 – First machine learning algorithms were developed. And were based on ideas from psychology and statistics.

2. In the 1970 – Machine learning began to be used for practical applications, such as image recognition and natural language processing.

3. In the 1980 – Support vector Machines (SVM) were introduced, which allowed for more accurate and efficient classification of data

4. In the 1990 – The rise of statistical learning algorithms, such as Bayesian Probabilistic Networks and Hidden Markov Model (HMMs), led to breakthroughs in areas such as speech recognition and machine Translation.

(AI)

5. In 1997 – The IBM computer Deep blue, which was a chess playing computer beat the world chess champion.

6. In 2000 – The advent of deep learning algorithms such as convolutional neural Networks (CNNs) and Recurrent neural Networks (RNNs) allowed for more complex and sophisticated machine learning applications.

Ex. Virtual Assistant
Automation of car.

7. The 21st Century — Today machine learning is used in a wide range of fields.
a) AlphaGo.
b) Image recognition and object classification detection: Self Driving cars, Security Camera
c) Speech Recognition — such as Siri, Google Assistant, Alexa
d) Recommender Systems — Such as Netflix and Amazon.
e) Medical diagonis and prediction — Ex. cancer Detection.
f) Fraud Detection
g) Natural language Processing — Sentimental Analysis, Chatbots.

(SL)

Introduction to ML approaches:—

(1) Artificial Neural Network :→

- Inspired by structure and function of the human brain.
- Computational model that consists of interconnected nodes, called neurons, which are organized into layers
- Each neuron receives input from other neurons, processes that input and then sends an output signal to other neurons in the networks.

- The input layer receives input data which is then passed through one or more hidden layers, where that data is transformed and processed.
- The final output is produced by output layer, which provides the result of the computation.

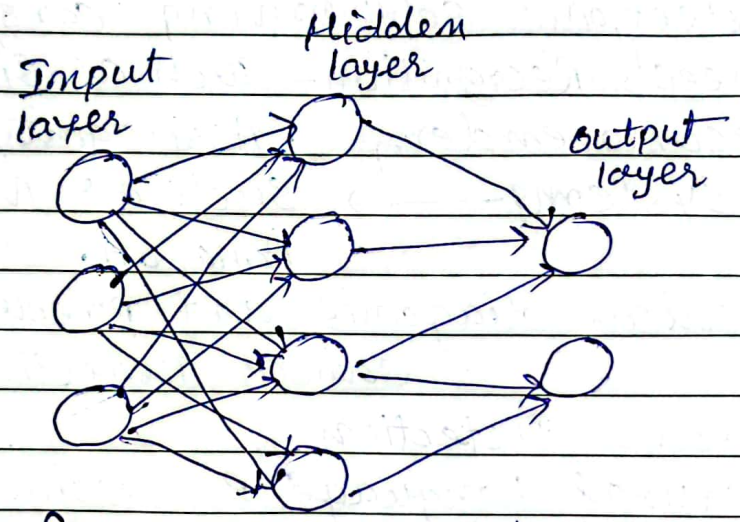
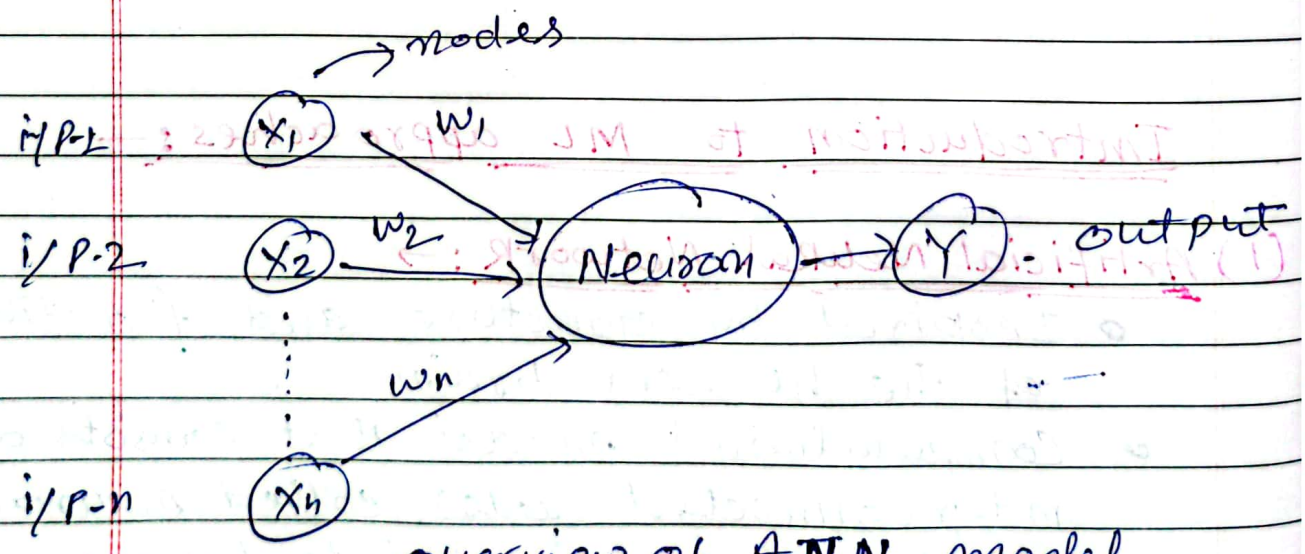


fig: ANN model



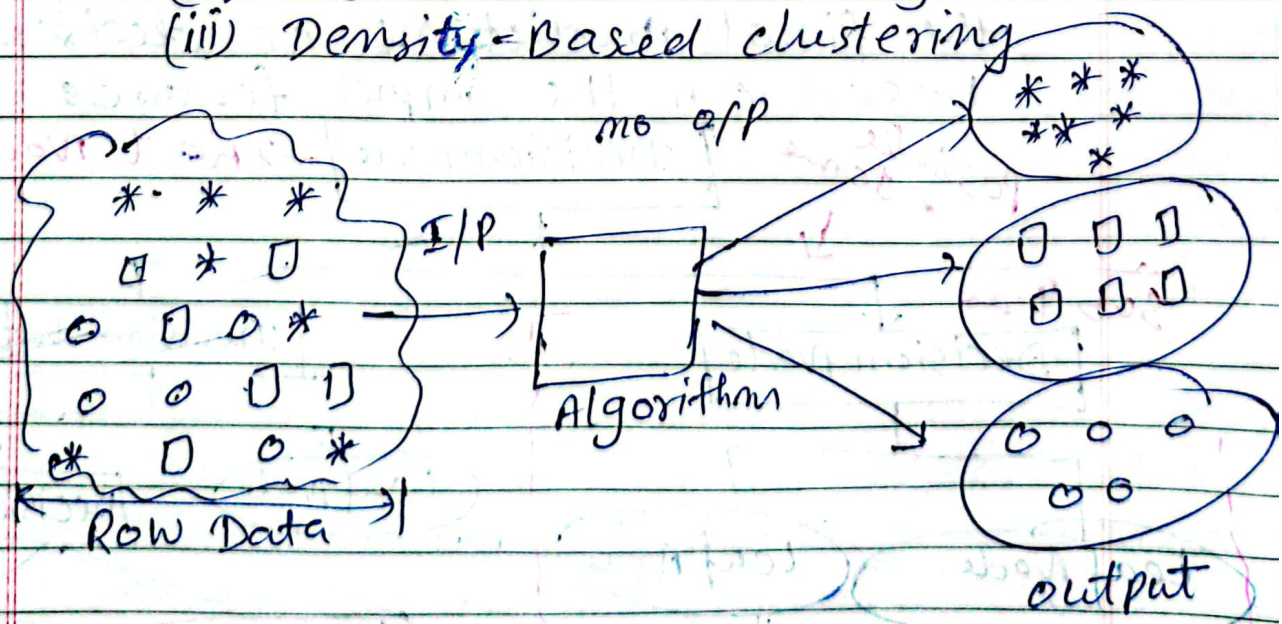
overview of ANN model.

- ANNs can be used for a wide range of machine learning tasks such as -

- Image and speech recognition
- Natural language Processing
- Time Series prediction.

(2) Clustering :—

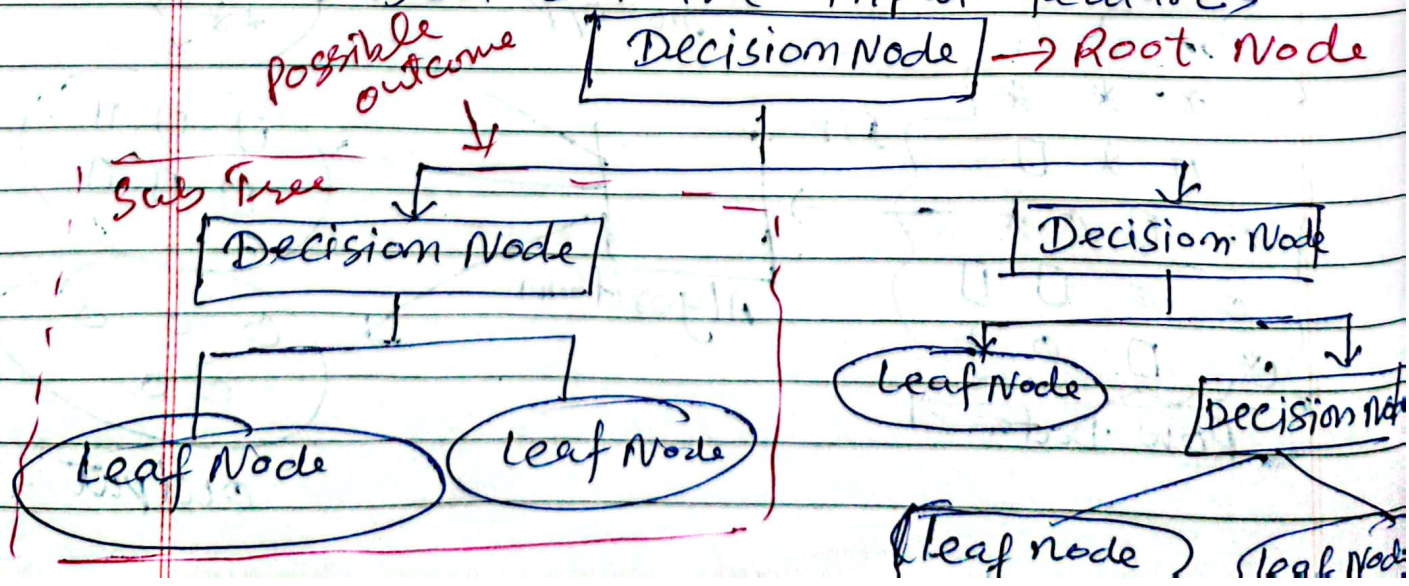
- Type of unsupervised machine learning algorithm.
- Group with similar data points together based on their similarity or distance to one another.
- The goal is to identify patterns in data without any prior knowledge of the structure of the data or labels associated with data.
- There are many different clustering algorithms, each with its own strengths and weaknesses.
- Most common clustering Algorithms include —
  - (I) K-mean clustering
  - (II) Hierarchical clustering
  - (III) Density-Based clustering



- clustering has many applications in machine learning and data analysis such as —
  - ➢ clustomer Segmentation ✓
  - ➢ Image and Text analysis
  - ➢ Anomaly detection or Error Detection

(3) Decision Tree learning :—

- Type of Supervised machine learning that is used for classification and regression tasks.
- Works by constructing a decision tree that represents a sequence of decisions and their possible consequences
- Each node in the decision tree represents a decision based on the value of an input feature and each branch represents the possible outcomes of that decision.
- The leaf nodes of the tree represent the final predictions or decision based on the input features



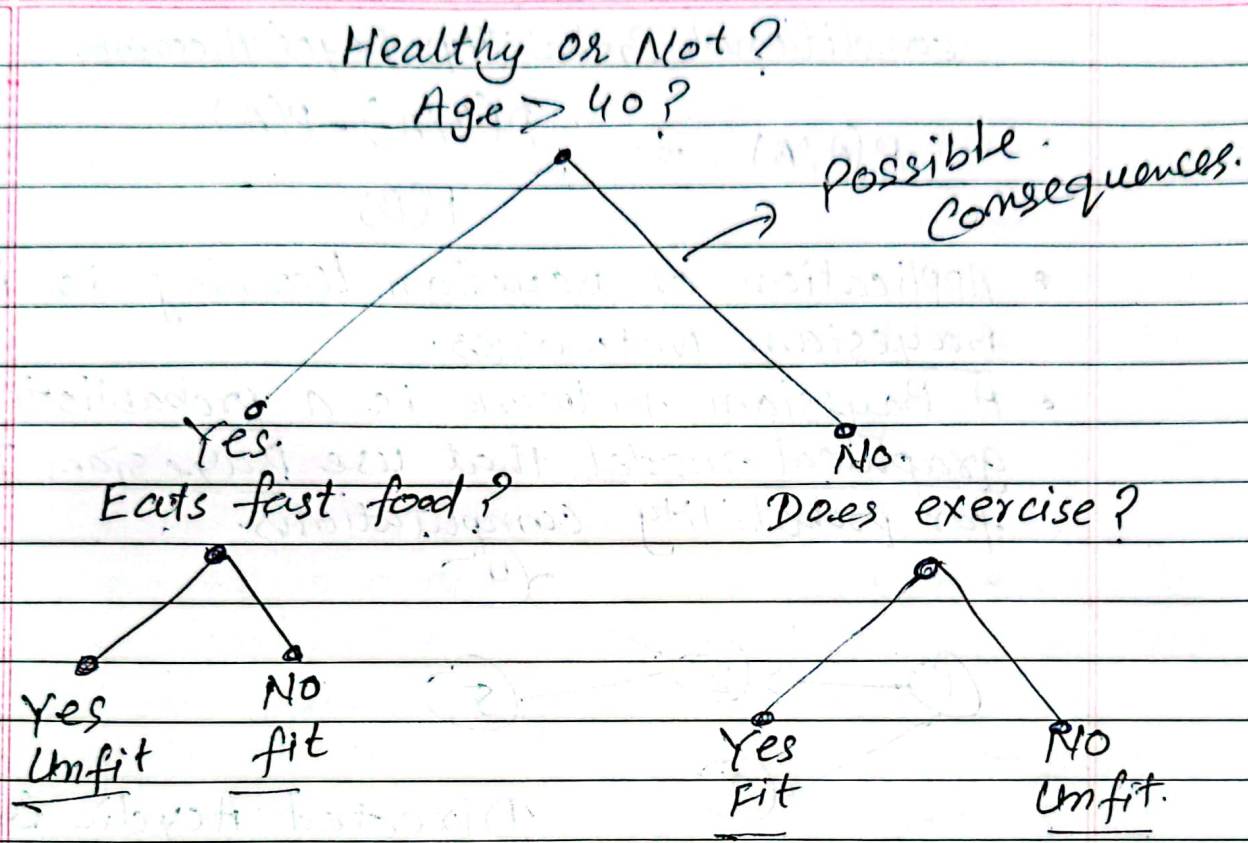


fig: Decision tree learning

- Decision trees have many advantages, such as being easy to interpret and visualize and requiring minimal data preparation.
- There are many different algorithms and variation of decision tree learning, such as CART (Classification and Regression Tree) ID3 (Iterative Dichotomiser-3) and C4.5

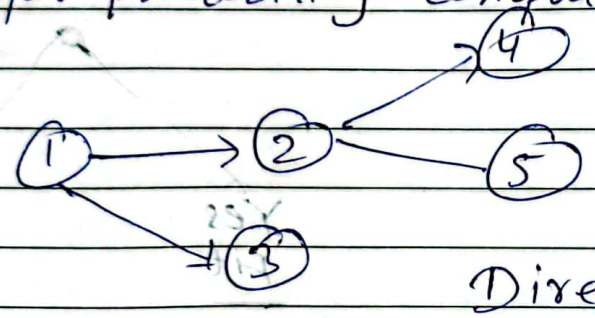
(4) Bayesian Networks:—

- Based on Bayes' Theorem which is mathematical formula that describes how to update the probability of a hypothesis given new edivence.

Conditional Probability: Bayes' Theorem.

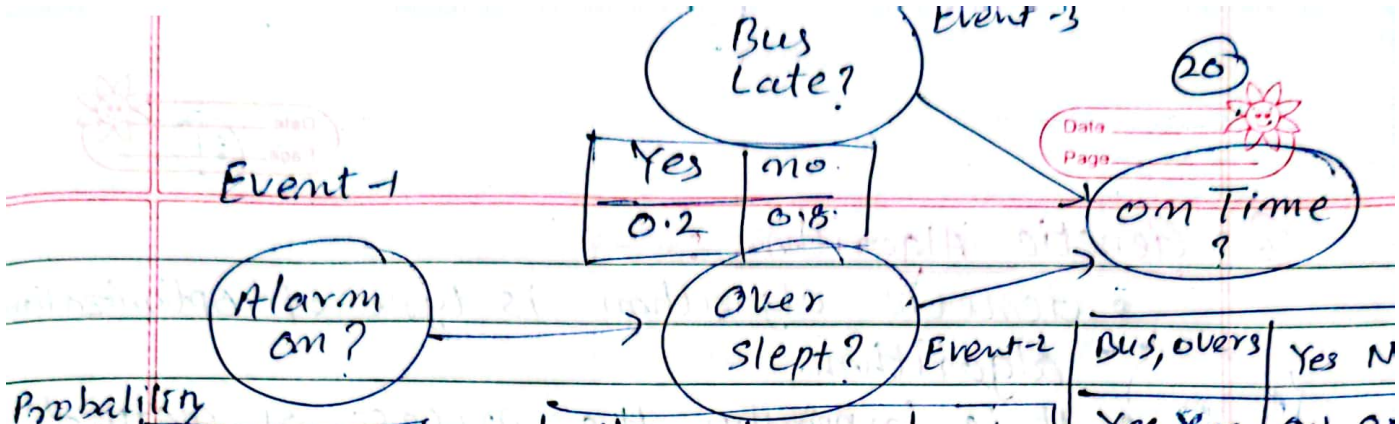
$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

- Application of Bayesian learning is in Bayesian Networks:
- A Bayesian network is a probablistic graphical model that use Bayesian Infere for probability computations.



Directed Acyclic Graph (DAG)

- Bayesian networks are ideal for taking an event that occured and predicting the likehood that any one of several possible known causes was contributing factor.
- Bayesian networks can be used for modelin complex systems and making predictions in various domains. (wether forecasting and Brest cancer Detection)
- It is often in various applications —
  - Natural processing language
  - Computer Vision and Robotic



Yes	no
0.2	0.8

Probability given.

Yes.	no.
0.9	0.1

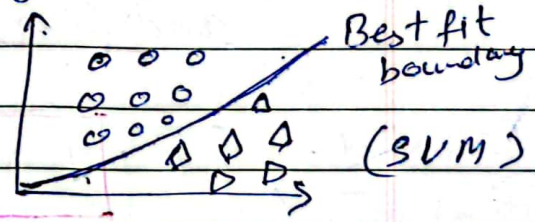
Event-2

Alarm	Yes	no
Yes.	0.1	0.9
No.	0.9	0.1

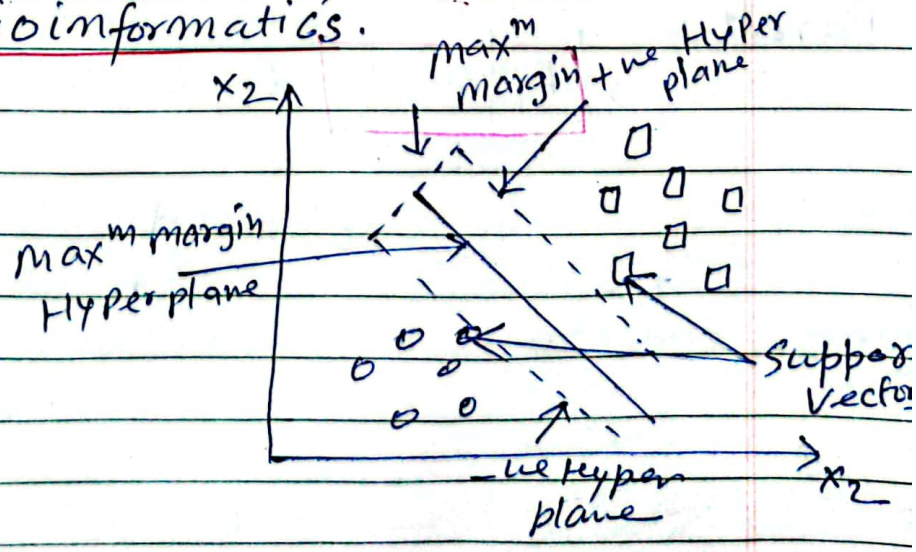
Bus, overs	Yes	N
Yes, Yes	0.1	0.
Yes, No	0.2	0.
no, Yes	0.3	0.
no, no	0.9	0.

(5) Support Vector Machine (SVM) —

- SVM is a type of supervised machine learning algorithm that can be used for classification or regression tasks.
- SVM tries to find the best possible boundary between the different classes in data by finding a hyper plane that maximizes the margin between two classes
- The data points are represented as vector in a high Dimensional space

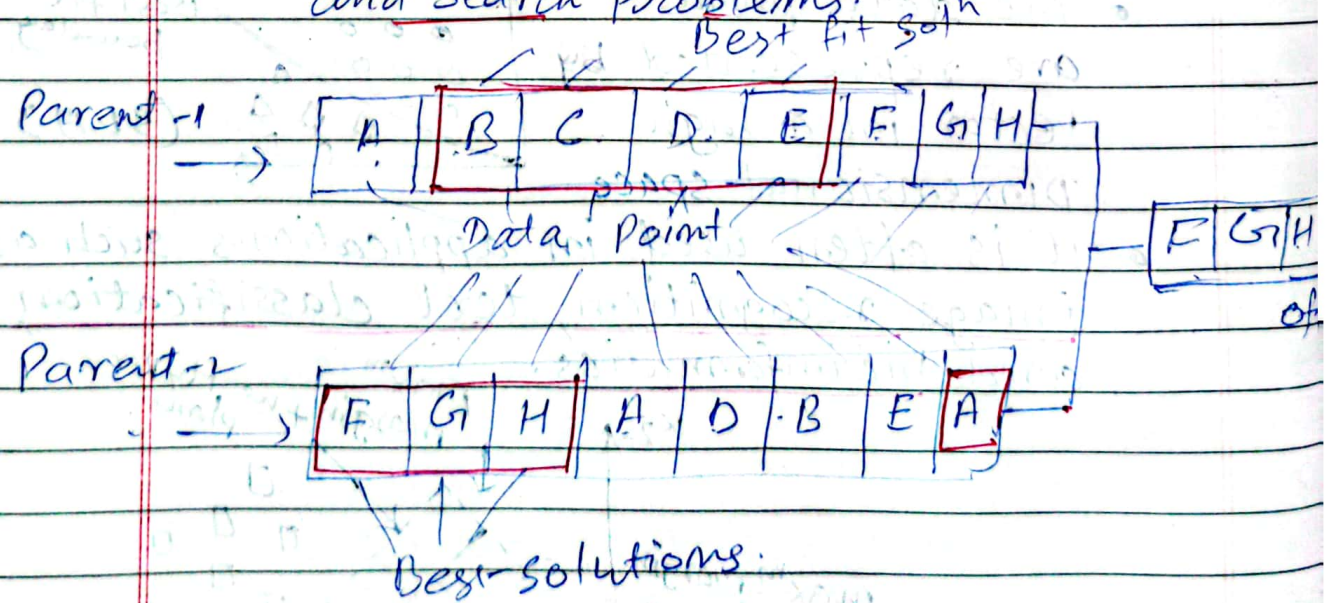


- It is often used in applications such as image recognition, text classification and bioinformatics.



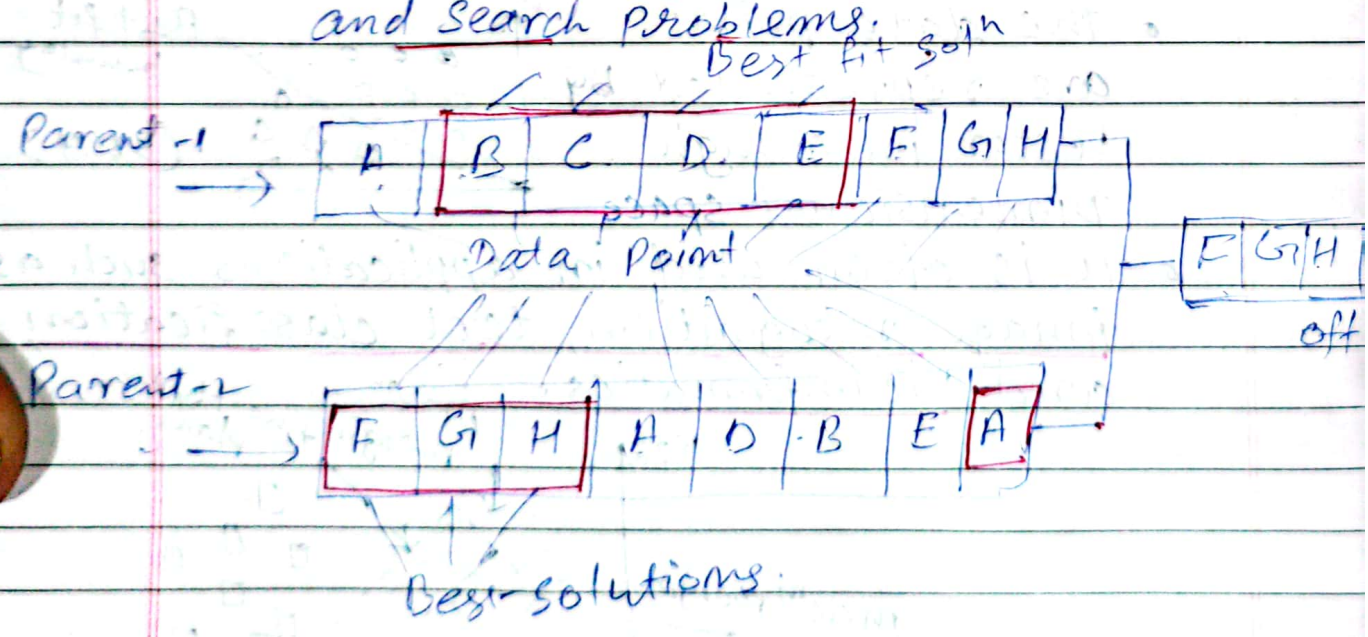
⑥ Genetic Algorithm :—

- Genetic algorithm is type of optimizat algorithm.
- It is inspiredby the process of natural Selection in biological evolution.
- It starts by generating a b:anch o random solutions and then testy (calculate fitness) them to see how well they perform.
- The best solutions are then combine to create new solutions, just like how a baby inherits traits from its parer
- This process is repeated until the bes solution is found.
- Commonly used to generate high-qua solutions for optimization problems and search problems.



6 Genetic Algorithm :-

- Genetic algorithm is type of optimization algorithm.
- It is inspired by the process of natural selection in biological evolution.
- It starts by generating a branch of random solutions and then tests (calculate fitness) them to see how well they perform.
- The best solutions are then combined to create new solutions, just like how a baby inherits traits from its parents.
- This process is repeated until the best solution is found.
- Commonly used to generate high-qual solutions for optimization problems and search problems.



How genetic Algorithm works:

Initialization

Fitness Assignment

Selection

Reproduction

Termination

Issues in machine learning:

There are many issues and challenges in machine learning and data science—

(1) Data Quality and bias:—

Machine learning algorithms are only as good as data they are trained on. if data is poor quality then contains errors or inconsistencies, or is biased in any way, the resulting model will also be flowed

B	C	D	E	A

spring (solutions)

(2) Model interpretability:—

As machine learning models become more complex, it can be difficult to understand how they arrived at their predictions or decisions.

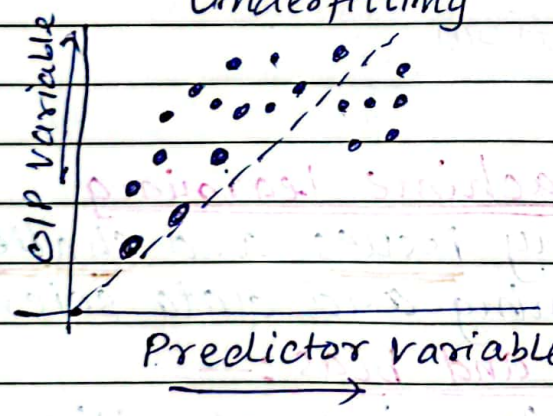
(3) Privacy and security:—

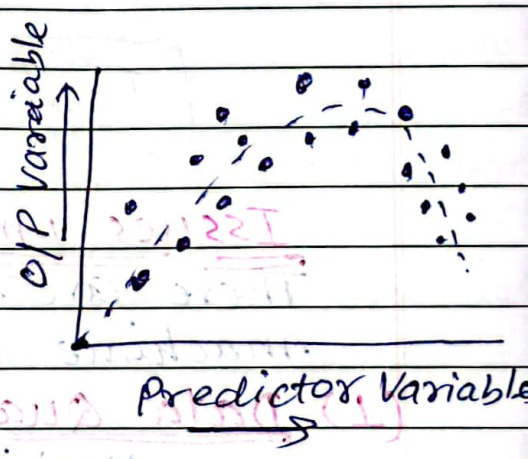
Machine learning algorithms often require access to large amount of sensitive data such as personal data, medical records or financial data.

(4) Overfitting and underfitting :—

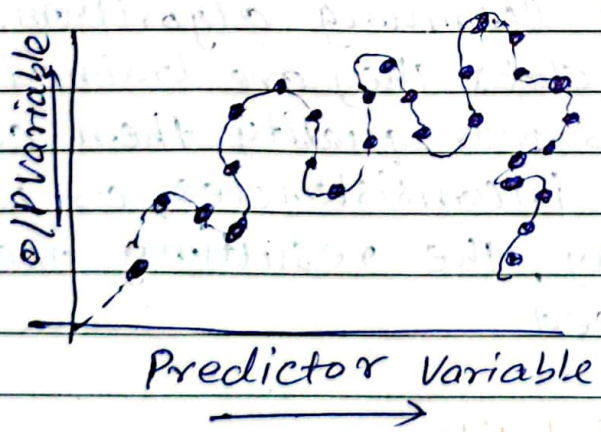
Machine learning model can suffer from overfitting and underfitting, which can result in poor performance on new data

Underfitting





overfitting



(5) Data Scarity :— In some domains such as medical research or rare event prediction, there may be limited data available for training machine learning model.

(6) computational resource :— Machine learning

algorithms can be computationally intensive and may require significant amounts of processing power or specialized hardware to efficiently.

Data Science Vs Machine learning:—

Data Science	Machine Learning
Field that determines the process, systemses and tools needed to transform data into insight to be applied in various fields.	Machine learning is the part of data science. Its algorithms train on data discovered by Data Science to learn.
Skill Required—	Skill Required—
• Statistics	• Programming skill
• Data Visualization	• Data Modeling
• Coding skill (Python)	• Prototyping
• SQL/MYSQL	• Comfortable working with data.