Date :

Machine Learning

What is Machine Learning ?

Machine Learning is a branch of artificial intelligence that enables computers to learn and improve from experiences without explicit Programming.

It involves developing algorithms that analyze data, identify patterns and make decisions or predictions.

Categories :

1. Supervised Learning
2. Unsupervised Learning
3. Semi Supervised Learning
4. Reinforcement learning.

1. Supervised Learning : uses labelled data to train models for classification
   eg :- Spam detection.

2. Unsupervised Learning : works with unlabelled data to uncover hidden patterns. Such as clustering customers by behaviour.

3. Semi - Supervised Learning : Combines small amount of labelled data and large amount of unlabelled datasets to improve learning.

4. Reinforcement Learning : Trains agents to make sequential decisions through rewards and penalties.
   - playing games or Controlling robots

Steps in ML:

1. Collect and Preprocess data.
2. Split data into training and testing sets.
3. Train the model on the training set.
4. Evaluate and refine model.
5. Deploy the model for real-time predictions.

2 Applications Areas of Machine Learning:

- Spam Filtering: Identifies and blocks unwanted mails.
- Smart assistants: Powering virtual assistants like Alexa, Siri and google Assistant to perform tasks using voice interaction.
- Social Media: Friend recommendations and Face Recognition.
- Navigation: Google Maps uses ML to analyze traffic and suggest optimal routes.
- Self Driving Cars: Helps vehicles to perceive their surroundings and make driving decisions.
- Recommender System: Suggests movies, products or music based on user preferences.
- Cyber Security: Detects unusual activities in network traffic for early warnings.

Date : _____

3 Data pre-processing :

Data pre-processing is the essential step for preparing raw data for machine learning models.

It improves the quality of data making it suitable for analysis.

1. Data Cleaning : Handling Missing values, removing duplicates and correcting errors.

2. Data Transformation : Converting data into usable format, such as scaling numerical values or encoding categorical variables.

3. Feature Selection : Choosing relevant features to reduce complexity and improve model performance.

4. Data Splitting : Dividing the dataset into training and testing sets.

Date :

4 Training and Choosing Predictive Models:

1. Training the model:
The model learns from a training dataset by identifying patterns and relationships between input features and output labelles

2. Choosing predictive Models:
Different algorithms (eg: Decision Tree, SVM, Neural Networks) are evaluated to find the best fit for the problem. The Choice depends on the factors like datatype, size and Complexity.

3. Optimization:
Adjust hyperparameters to improve the models performance.

4. Testing:
Test the model on unseen data to ensure it can make accurate predictions in real-world Scenerios.

Date :

Unit-I

1. Machine Learning.

Machine learning is a branch of artificial intelligence where computers learn to perform tasks by analysing patterns in data instead of being explicitly programmed to follow specific rules.

How it works :

1. Data is fed into the machine
2. The machine finds patterns or relationships in the data.
3. It uses those patterns to make predictions or decisions without being told exactly what to do.

Example : To teach Machine to recognize Cats, by providing labeled images of Cats and Non-cats. The machine learns patterns from these images and uses them to identify Cats in new pictures.

Categories :

1. Supervised
2. Unsupervised
3. Semi-Supervised
4. Reinforcement

1. Supervised :

In Supervised learning, the machine is trained using labeled data. We provide input-output pairs, and machine learns the relationship between them.

- Example : Predicting house prices based on features like size, location etc.

Date :

2. Unsupervised Learning :

Here, the data is not labeled. The machine tries to find hidden patterns or structures in the data without any guidance.

• Example : Grouping customers based on their purchase behaviour (clustering).

3. Semi-Supervised Learning :

A mix of both Supervised and Unsupervised learning. The machine is given a small amount of labeled data and a larger amount of unlabeled data

• Example : A machine learning model trained to classify images with a small number of labeled photos and many unlabelled ones.

4. Self-Supervised Learning :

In this type, the machine generates its own labels from the data. It can predict one part of the data using another part of the same data.

• Example : Predicting the next word in a sentence during natural language processing.

5. Reinforcement Learning :

The machine learns by interacting with an environment and receiving feedback in the form of rewards or penalties.

• Example : A robot learning to navigate a maze by receiving rewards for reaching the destination.

Date :

2 Applications of Machine Learning.

Machine Learning has a wide range of applications across various industries.

1. Health Care
2. Finance
3. Retail and E-Commerce.
4. Autonomous Vehicles.
5. Natural language Processing
6. Image and vidio Analysis.
7. Robotics
8. Cybersecurity
9. Gaming
10. Marketing and Advertising.

1. Health Care:

- disease prediction and diagnosis
- Personalized treatment recommendations.
- detecting tumors in X-rays.

2. Finance :

- Fraud detection
- Stock Market Predictions.
- Credit Scoring and risk Management.

3. Retail and E-Commerce:

- Product Recommendations (Amazon, Netflix)
- Customer behaviour Analysis.

4. Autonomous Vehicles :

- Self-driving Cars.
- Traffic pattern Analysis.

Date :

5. Natural Language Processing :
- Speech Recognition (Siri, Alexa)
- Language translation (Google Translate)
- Chatbots and virtual assistants

6 Image and Vidio Analysis :
- Facial Recognition (unlocking phone)
- Object detection in images and videos

7. Robotics :
- Industrial Automation and robot Control
- Reinforcement learning in robot movment and task performance

8. Cyber Security :
- Detecting and Preventing Cyberattacks.
- Spam filtering and malware detection.

9. Gaming :
- AI opponents in games (chess)
- Game world Simulation and player interaction.

10. Marketing and ~~Advert~~ Advertising :
- Target Advertising.
- Ad performance optimization
- Predictive analytics for sales and conversions.

Date : _____

3 Data Pre-Processing

Data Preprocessing in machine learning is the step where raw data is cleaned up and prepared to be used by models.

Tasks :

- Handling Missing Data (filling in or removing incomplete entries)
- Normalizing or Scaling data (putting it on the same scale)
- Encoding Categorical Data (turning text labels into numbers)
- Removing noise or irrelevant data.

4 Choosing Predictive Models

Choosing a predictive model in machine learning depends on the type of problem and the data.

1. Regression Model (like Linear Regression) are used when predicting continuous values (eg. house prices)
2. Classification Models (like Decision Tree or Logistic Regression) are used when predicting categories (eg: whether an email is spam or not)
3. Clustering models (like K-means) are used to group similar data points without predefined labels.

Date :

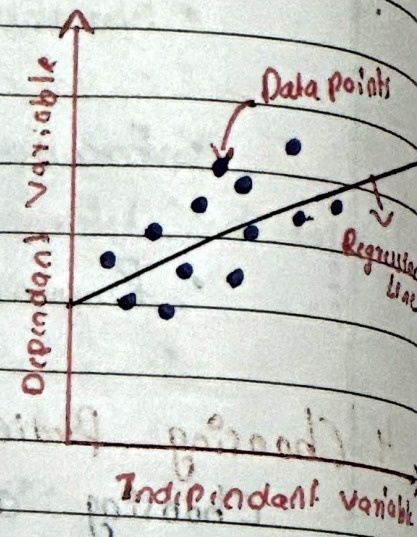
Unit - II

1 Linear Regression

Linear Regression is a machine learning algorithm used to predict a continuous value based on the relationship between input features and the output. It finds the best-fitting straight line (regression line) through the data points by minimizing the differences between the predicted values and the actual values.

Eg:

- weight of the car and miles per gallon
- Area of the house and price of the house.



Linear Model

- They make the prediction using a linear function of the input features.
- For Linear Regression, the general prediction formula is

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Date :

2 Qn1) Glucose level and Age of Six Subjects are given in table below. Predict the glucose level of a person of age 55, using Linear Regression:

Subject	Age	GLevel
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

X	Y	X^2	XY
43	99	1849	4257
21	65	441	1365
25	79	625	1975
42	75	1764	3150
57	87	3249	4959
59	81	3481	4779
247	486	11409	20485

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{486 \times 11409 - 247 \times 20485}{6(11409) - (247)^2}$$

$$= \frac{5,544774 - 5,059,795}{68454 - 61009}$$

$$= \frac{484979}{7445}$$

$$b_0 = 65.142//$$

Date :

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{6(20485) - (947)(486)}{6(11409) - (947)^2}$$

$$= \frac{9868}{7445} = 0.$$

$b_1 = 0.3952$ //

$y = b_0 + b_1 x$

$= 65.102 + 0.3952(55)$

$y =$.

Date :

3 Advantage of Linear Regression.

1. Easy to Implement : It is easy to implimint and maintain because it doesn't need much effort or complex engineering.

2. Scalability : Since its not Computationally expensive, it can be used for situations that need Scalling like big data applications.

3. Interpretability : This easy to understand and quicks to train. Unlike deep learning. It doesn't need as much data or time to train well

4. Applicability in Real-time : Because it's easy to train and doesn't need much Computional power, it can be quickly retrained with new data, making it useful for real time Predictions.

4 Dis-advantages of Linear Regression:

- It assumes a linear relationship between the depandant and indipendant variables, which is rarely seen in real-world data.

- It assumes a stright line relationship between variable which isn't the case.

- It is prone to noise and overfitting.

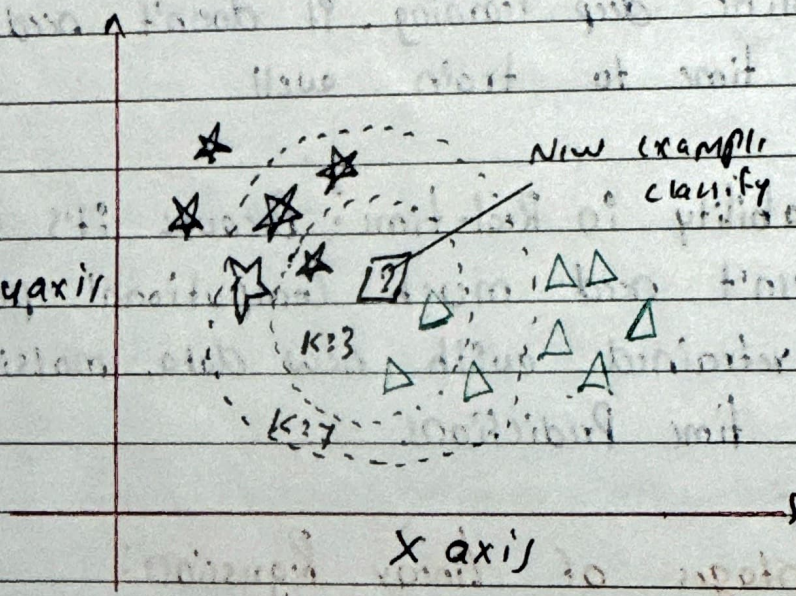
- It's not good choice for data sets with fewer observations than attributes, as it can cause overfitting.

Date :

KNN - K-Nearest Neighbour

- KNN is a lazy learning ; non-parametric algorithm uses data with several classes to predict the classification of the new sample point.

- This called a lazy learning algorithm or Lazy learner because it doesn't perform any training when you supply the training data. Instead, it just stores the data during the training time and doesn't perform any calculations.



formula : $\sqrt{(x-a)^2+(y-b)^2}$

Date :

	Brightness	Saturation	class
d1	40	20	Red
	50	50	Blue
	60	90	Blue
	10	25	Red
	70	70	Blue
	60	10	Red
	25	80	Blue

Brightness	Saturation	class
20	35	?
x	y	

Formula $\sqrt{(x-a)^2+(y-b)^2}$

$d_1 = \sqrt{(20-40)^2+(35-20)^2}$

$d_1 = \sqrt{(20)^2+(15)^2}$

$d_1 = \sqrt{400+225}$

$d_1 = \sqrt{625} = 25$

$d_2 = \sqrt{(20-50)^2+(35-50)^2}$

$= \sqrt{(900)+(225)}$

$= 33.54$

$d_3 = \sqrt{(20-60)^2+(35-90)^2}$

$= \sqrt{(40)^2+(55)^2}$

$= \sqrt{1600+3025}$

$= 68.007$

$d_4 = \sqrt{(20-10)^2+(35-25)^2}$

$= \sqrt{100+100}$

$= 14.14$

$d_5 = \sqrt{(20-70)^2+(35-70)^2}$

$= \sqrt{(2500)+(1225)}$

$= 61.03$

Date :

$d_6 = \sqrt{(20-60)^2 + (35-10)^2} = 47.17$

$d_7 = \sqrt{(20-21)^2 + (35-80)^2} = 45$

Brightness	Saturation	Class	Distance	Rank
40	20	Red	25	1
50	50	Blue	33.54	2
60	90	Blue	68.01	3
10	25	Red	10	4
70	70	Blue	61.03	5
60	10	Red	47.17	6
25	80	Blue	45	7

The new entry is classified as Red.

Steps involved in KNN algorithm.

- Calculate the distance between test data and each row of training data with the help of any of any the method.

  The most common method to calculate distance is euclidean.

- Now based on the distance value, sort them in acending order

- Now, it will choose the top K rows from the sorted array.

- Now, it will asign a class to the test point based on most frequent class of these rows.

Date :

5 Naive Bayes Classifiers (probability based).

Posterior Prob. $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$ →

$P(A|B)$ = Postirior probability
$P(B|A)$ = Likelihood
$P(A)$ = Class Prior
$P(B)$ = Prediction Prior.

Distance
Red
Red
Blue
Blue
Red.

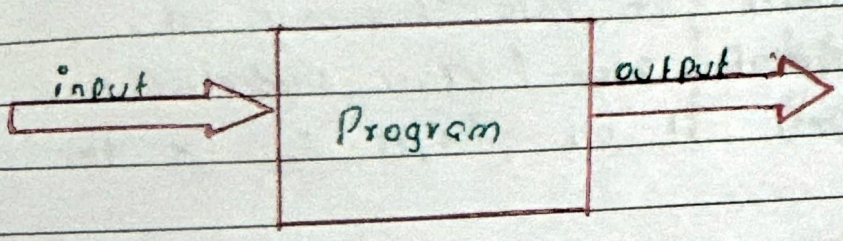
① Collect the data
② Convert it into frequency table
③ Calculate prior / class predictions
④ Apply it on Bayes to fix the predicted class.

Date :

6 Traditional Programming vs Machine Learning.

Traditional Programing.

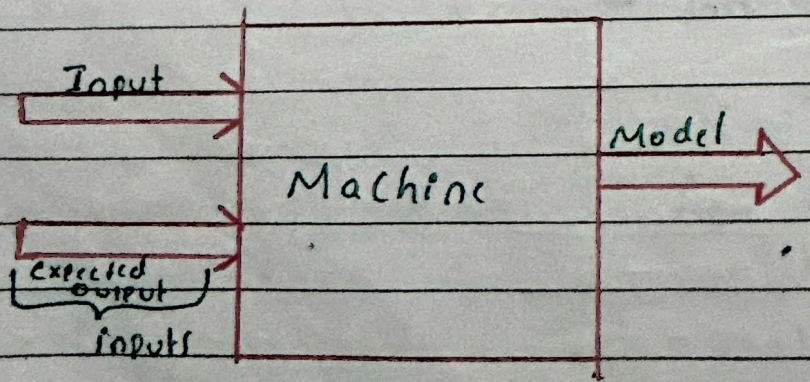
- In traditional programming, humans write explicit rules to solve a problem.
- humans give the computer inputs, and define the exact steps.



- The programs written once will not be changed until and unless they are updated manually by somebody.

Machine Learning.

- In machine learning, human don't explicitly write rules. Instead, the computer learns the rules on its own by analyzing data.



Date : _____

- These Model are nothing but Programs generated by algorithms without being explicitly programmed by human being.

- These models Continuously evove into better models automatically as the amount of quality data increases

Date :

Unit-II

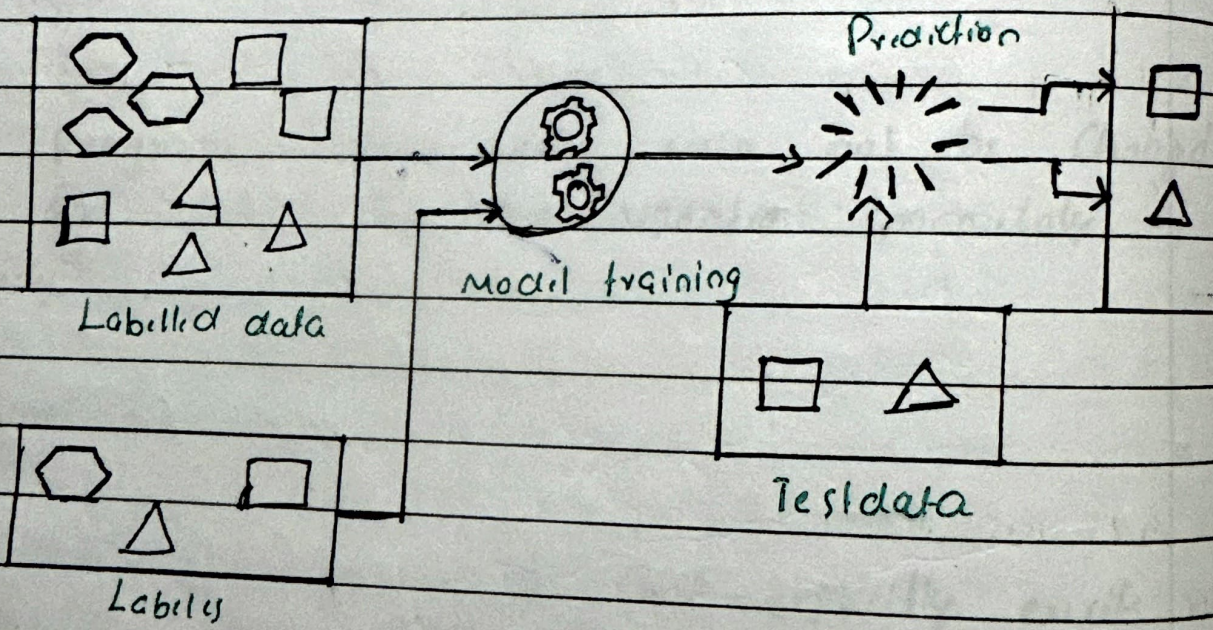
1. Supervised Learning :

Supervised learning is a machine learning approach where models are trained using labelled data.

Each input has corresponding output, allowing the algorithm to learn the relationship between th

It uses this learned relationship to make predictions on new, unseen data.

eg :- identifying if an email is spam or not



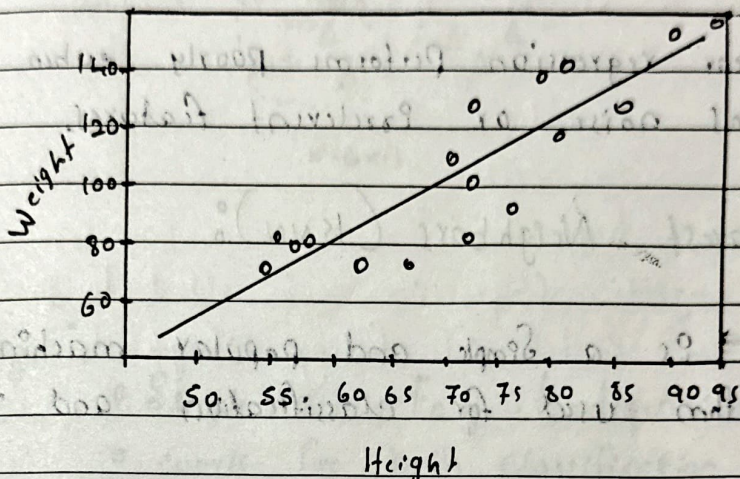
Date :

2 Linear Regression :

Linear Regression is a Supervised learning technique used to predict continuous numerical value by finding the relationship between inputs and outputs.

Formula : $y = m*x + c$.

where m is slope and c is the intercept



eg:
- weight of the car and miles per Gallon
- Area of the house and Price of the house.

Advantage :
- Linear regression is easy to understand and implement
- The model clearly shows the relationship between inputs and ouputs.
- It requires minimal computational resources, making it efficient for real-time applications.
- Can handle large datasets.
- Useful in fields like finance, healthcare and marketing.
- The model can trained quickly with new data

Dis. Advantages:

• It assumes straight line relationships between dependent and independent variables which is rare in real world.

• It is prone to noise and overfitting / underfitting

• Linear regression performs poorly when the data contains noise or irrelevant features.

3 K-Nearest Neighbors (KNN):

• KNN is a simple and popular machine learning algorithm used for classification and regression.

• It predicts results based on the K-nearest data points from the training datasets.

• It's called a lazy learning or Lazy learner because it doesn't perform any training when we supply the data, instead it just stores the data during the time.

How it works:

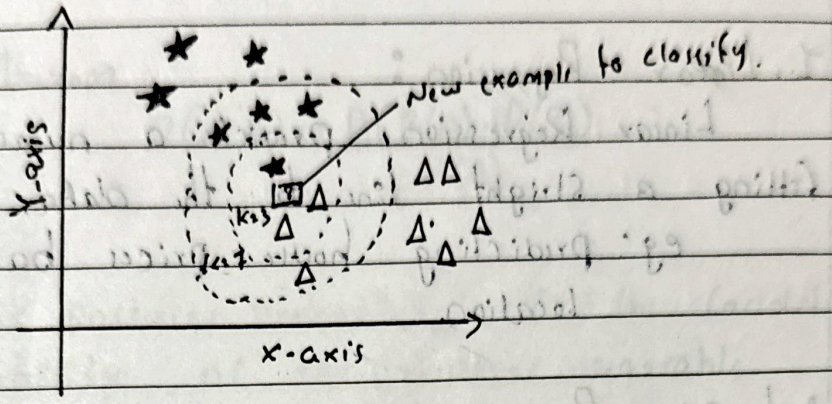
1. Distance Calculation: Compute the distance between the new data point and training point.

2. Sort the Neighbors: Sort the distance in ascending order.

3. Select the K-Neighbors: Picks the top K-nearest neighbors

Date :

4. Make Predictions :
   - For classification : Assign the most common class value
   - For regression : Take the average value.



Advantage :
- Simple and Easy to learn.
- works for both classification and regression
- Handles multi-class problems efficiently.
- No need for training model.

Dis-advantage :
- Computationally expensive
- struggles with large datasets
- Requires proper datasets for good accuracy

Example :
Predict Persons T-shirt size based on their height and weight by calculating distances.

4 Linear Models in Machine Learning:

Linear models a type of machine learning algorithm that make predictions based on a linear relationship between input features and the target variable

1. Linear Regression:

Linear Regression predicts a numerical value by fitting a stright line to the data.

eg: predicting house prices based on size and location.

2. Logistic Regression:

Logistic Regression is used to predict binary outcomes instead of fitting stright line

eg: predicting whether tumour is benign or malignant

Linear Regression	Logistic Regression.
• Used to predict the continuous dependent variables	• Used to predict the categorical dependant variable.
• Used for Solving Regression Problem	• Used for Solving classification Problem
• Predict the value of continuous variables	• Predict the value of categorical variables.
• Least square estimation method is used	• maximum likelihood estimation is used
• Output must be continuous	• Output must be categorical
• It is required to have linear relationship between dependant and independant variable	• It is not required to have linear relationship

Date :

5 Naive Bayes Classifier:

Naive Bayes assumes that all features in the dataset are conditionally independent within each class. It classifies data using Bayes theorem.

1. Bayes Theorem.

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

- P(A|B) or posterior Probability: is the conditional Probability of the response variable
- P(A): or Class prior: is the prior probability of the response variable
- P(B|A) or likelihood: is basically the reverse of the posterior probability.
- P(B) or Marginal Probability: is the evidence or the probability of training data.

How it works:

Step 1: Convert raw data into frequency ~~variables~~ table,
Step 2: Calculate prior, likelihood. and posterior probab
Step 3: Classify based on highest posterior probabi

Example:

Predicting if someone play tennis based on weather conditions like "Rainy", "Hot", "High humidity" and "No wind".