Combining multipl

UNIT-III

SUPERVISED LEARNING

Introduction to Machine Learning:-

ML Learn from historical data, build the Prediction models and whenever it receives new data, Predicts the output for it.

The accuracy of predicted output depend upon the amount of data, as the huge amount of data helps to build a better model which Predict the output more accurately.

i/p past data → Training → ML algorithm → Building Logical Models → output

↑ New data.

Features:-

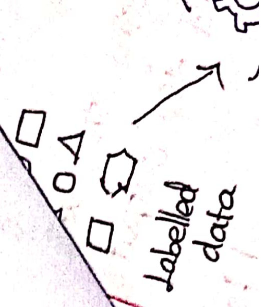
- → ML uses data to detect various patterns in a given dataset.
- → It learn from past data and improve automatically.
- → It is a data driven technology

Importance of Machine Learning.

- → Rapid increment in the production of data.
- → Solving Complex problem, which are difficult for a human.
- → Decision Making in various sectors including finance
- → Finding hidden patterns and extracting useful information from data.

Types of ML:-

* Supervised Learning
* unsupervised learning.
* Reinforcement learning.



Supervised Learning:-

Training models on labeled data to predict outcomes or classify data points.

Unsupervised Learning:-

Discovering patterns and ~~x~~ Structures in unlabeled data.

Reinforcement Learning:-

Training agents to make decision in an environment to maximize a reward.

ML

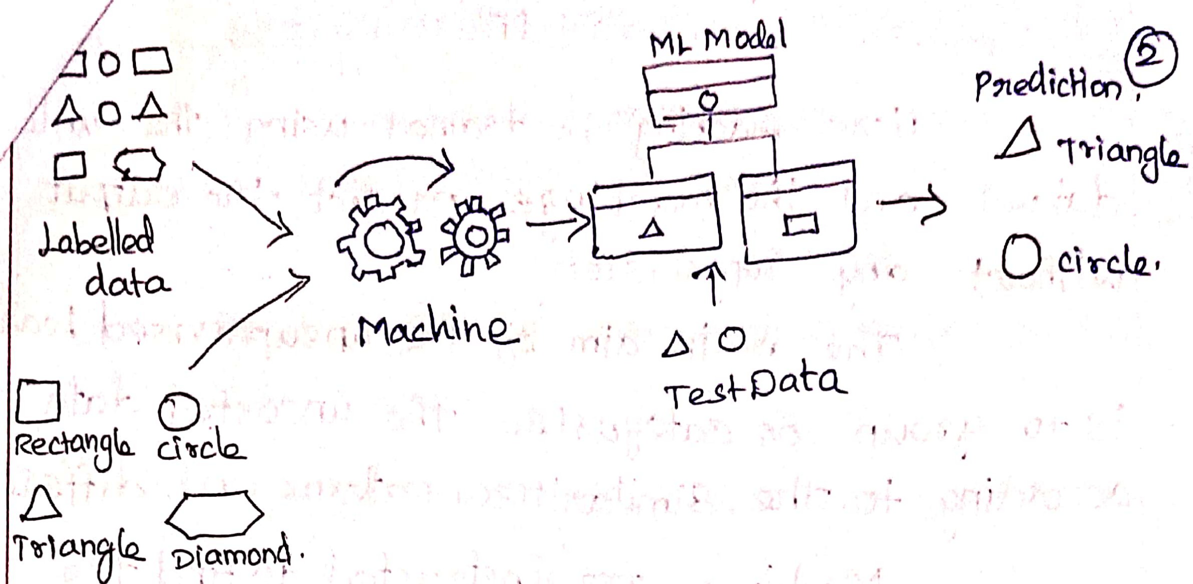
- Supervised Learning
  - Classification
  - Regression.
- unsupervised Learning..
  - clustering.
  - Association analysis.
- semisupervised learning.
- Reinforcement learning.
  - Active
  - Passive.

Supervised ML:-

It train the machine using the labelled dataset and based on the training, the machine Predicts the output.

The labelled data specifies that some of the input are already mapped to the output.

Train the machine with the input and corresponding output, and then the machine will Predict the output using the test dataset.



Types :
i) Classification
ii) Regression.

i) Classification:-

↦ It used to solve the classification problem in which the output variable is categorical.

↦ Such as 'yes' or 'No', female or Male

↦ The classification algorithm predict the categorical present in the dataset.

↦ Some popular classification algorithm are Random forest, Decistion tree, logistic regression, Support vector Machine.

Regression:-

↦ It is used to solve Regression Problem in which the time is a linear Relationship b/w i/p and o/p value

↦ This are used to predict continuous output variable such as market friend, weather prediction, mark of students etc.

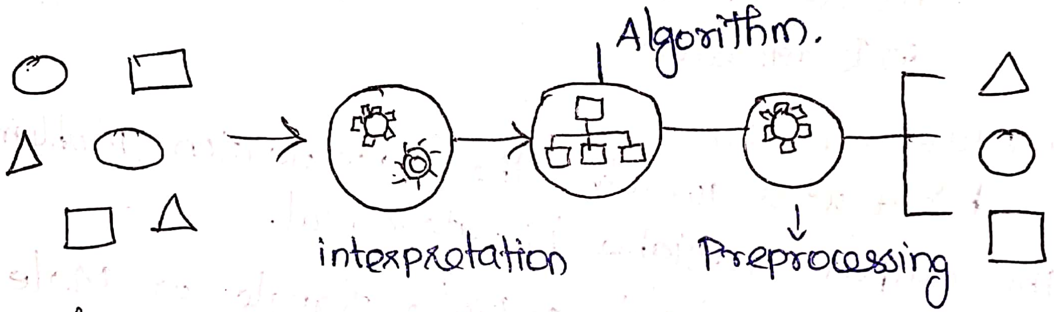
Some popular Regression Algorithm are, Simple linear Regression, Multivariate Regression, Decision Tree, Lasso Regression.

Unsupervised Maching Learning:-

Here maching is trained using the dataset, and the machinge predict the output without any supervision.

The main aim of the unsupervised learning is to group or categories the unsorted dataset according to the similarities, patterns and differences.

Machines are instructed to find the hidden patterns from the i/p dataset.



Types: of unsupervised ML.

i) clustering
(ii) Association.

Clustering:-

It is used to find the inherent groups from the data.

The object in the group are most similar to each other and no similarities with the objects to other groups.

An example of the clustering algorithm is grouping the customers by their purchasing behavious.

ed Learning.

Popular clustering Algorithms:-

Multiple learners: ...

③

i) k-Means clustering.
ii) Mean shift Algorithm.
iii) Principal Component Analysis
iv) Independent component Analysis.

Association:-

↳ It find interesting relation among variables within a large dataset.

↳ It is used to find the dependency of one data items on another data item.

↳ Based of dependency it maps those variable so that it can generate maximum.

Semi Supervised Learning:-

↳ To overcome the drawbacks of supervised learning and unsupervised learning algorithm, the Concept of Semi-Supervised is introduced.

↳ Hence it uses the combination of labelled and unlabelled dataset during training Period.

Reinforcement Learning:-

↳ Reinforcement Learning works on a feedback based process - Learning from experiences and improve its Performance.

↳ In Reinforcement Learning, there is no labelled data like supervised learning and agents learn from their experiences only.

Types of Reinforcement Learning:-

1) Active Reinforcement Learning
2) passive Reinforcement Learning.

Passive

→ The agent policy (sequence of actions) is fixed which means that it is told what to do

→ The goal of passive RL agent is to execute a fixed policy and evaluate it.

Active

↳ An Agent need to decide what to do as there's no fixed policy that it can act on.

↳ An active RL agent is to act an learn an optimal policy.

Linear Regression Model:-

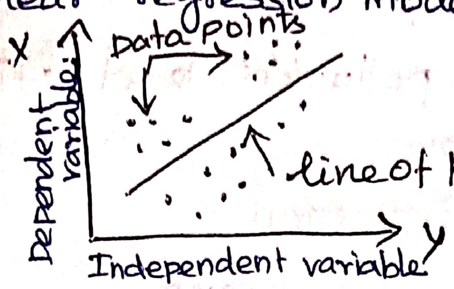
Linear Regression:-

↳ linear Regression is an algorithm that Provides a linear relationship b/w an independent variable and a dependent variable to predict the outcome of future events.

↳ Here the dependent variable is also Called the response variable and independent Variable also called a explanatory or predictor Variables.

↳ Thus linear regression is a Superrised Learning algorithm that Simulates a mathematical relationship b/w Variables and make predictions for continuous or numeric variables such as Sales salary, age, etc.

Here A Sloped Straight line represent the linear regression model.



Least Squares:

In statistics, Linear regression is a linear approach to model the relationship b/w a scalar response (dependent variable) - say x, and one or more explanatory variable (independent variable) - say y.

The linear regression model provides a sloped. Straight line representing b/w the variables Mathematically, we can represent a linear regression as

$$y = a_0 + a_1 x + \epsilon$$

y – dependent variable (Target variable)

x – Independent variable (Predictor variable)

a_0 – Intercept of the line

a_1 – linear regression coefficient / ϵ – random error

Types Of Linear Regression:-

1) Simple linear Regression:-

If a single independent variable is used to Predict the value of a numerical dependent variable, then such a linear regression algorithm is called Simple linear regression.

Simple linear regression reveals the correlation b/w a dependent variable (i/p) and an independent variable (o/p).

It is relationship strength b/w the ... of the dependent variable is based on the ... of the independent variable.

Ex! The value of pollution level data at a specific temperature.

(2) Multiple linear Regression:-

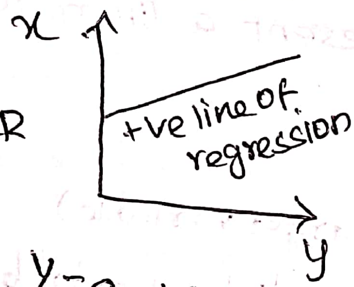
1) positive linear regression

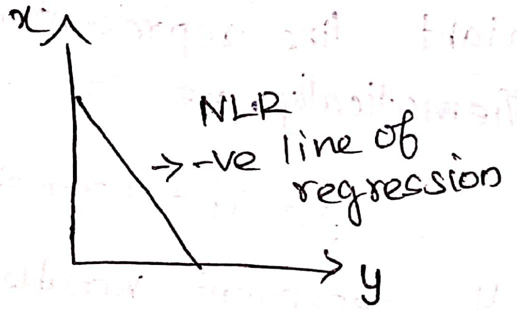
2) Negative linear regression.

1) Positive linear Relationship:-

If the dependent variable increase on the y-axis and independent variable increase on the x-axis, then such a relationship is termed as PLR.

PLR





$y = a_0 + a_1 x$

$y = a_0 + a_1 x$

$y = -a_0 + a_1 x.$

(ii) Negative linear Relationship:-

If the dependent variable decreases on the y-axis and independent variable increase on the x-axis then, such a relationship is called NLR.

Simple linear Regression Problem:-

A Dataset containing information about the no of hours students spend studying and their corresponding scores on a test. Your task is to perform simple linear regression to predict test scores based on the no of hours studies using following dataset.

S No of hours studied (x)	Test scores (y)	x^2	xy
2	75	4	150
8	82	9	246
4	93	16	372
5	89	25	445
6	98	36	588

$\bar{x} = \frac{20}{5} = 4$ $\quad \bar{y} = \frac{437}{5} = 87.4$ $\quad \overline{x^2} = \frac{90}{5} = 18$ $\quad \overline{xy} = \frac{1801}{5} = 360.2$

$$a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \qquad a_0 = \bar{y} - a_1\bar{x}$$

$$\text{Mean} = \frac{\text{sum of all obs}}{\text{no of obs}}$$

Simple linear Reg $y = a_0 + a_1 x \rightarrow (1)$

$$a_1 = \frac{360.2 - (4)(87.4)}{18 - (4)^2} = \frac{360.2 - 349.6}{18 - 16}$$

$$a_1 = 5.3 = \frac{10.6}{2} = 5.3 \rightarrow (2)$$

$$a_0 = \bar{y} - a_1\bar{x} \Rightarrow 87.4 - (5.3)(4)$$

$$87.4 - 21.2$$

$$a_0 = 66.2 \rightarrow (3)$$

Sub eqn (2) & (3) in eqn (1)

$$y = 66.2 + 5.3x$$

Scanned with OKEN Scanner

Ass. Sum:- A dataset consider 4 Instances of Salary and expenditure data is given as shown table Apply linear Regression tech to Predict the expenditure with Salary 3 lakh.

Salary (x)	
	6
5	
1	
2	
7	7
3	8
	?

Multiple linear Regression Problem.

$$y = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

$a_1, b_1, b_2 \cdots$ — coefficient.

x_1, x_2 . — Independent variable

y — o/p, dependent variable.

Age x_1	year x_2	Salary y
35	1	20
25	1	10
40	2	35
50	4	?

$$a = \bar{y} - b_1\bar{x_1} - b_2\bar{x_2}$$

$$y = a + b_1x_1 + b_2x_2$$

$$b_1 = \frac{(x_2^2)(x_1y) - (x_1x_2)(x_2y)}{(x_1)^2 \; (x_2^2) - (x_1x_2)^2}$$

$$b_2 = \frac{(x_1^2)(x_2y) - (x_1x_2)(x_1y)}{(x_1^2)(x_2^2) - (x_1x_2)^2}.$$

x_1	x_2	y	x_2^2	x_1^2	x_1x_2	x_1y	x_2y.
35	1	20	1	1225	35	700	20
25	1	10	1	625	25	250	10
40	2	35	4	1600	80	1400.	70
x_1= 100	x_2= 4	y= 65	x_2^2= 6	x_1^2= 3450	x_1x_2= 140	x_1y= 2350	x_2y= 100.

$$b_1 = \frac{(x_2^2)(x_1y) - (x_1x_2)(x_2y)}{(x_1^2)(x_2^2) - (x_1x_2)^2}$$

Scanned with OKEN Scanner

⑥

$$b_1 = \frac{(6)(2350)-(140)(100)}{(3450)(6)-(140)^2}$$

$$= \frac{14100 - 14000}{20700 - 19600} = \frac{100}{1100} = 0.09.$$

$$b_2 = \frac{(x_1^2)(x_2y) - (x_1x_2)(x_1y)}{(x_1^2)(x_2^2) - (x_1x_2)^2}$$

$$= \frac{(3450)(100) - (140)(2350)}{(3450)(6) - (140)^2}$$

$$= \frac{345000 - 329000}{20700 - 19600.} = \frac{16000}{1100} = 14.54$$

Mean $\bar{y} = \frac{65}{3} = 21.6$

$\bar{x_1} = \frac{100}{3} = 33.3$

$\bar{x_2} = 4/3 = 1.33$

Sub b_1 & b_2 & mean value in a.

$a = \bar{y} - b_1\bar{x_1} + b_2\bar{x_2}$

$= 21.6 - (0.09)(33.3) + (14.5)(1.33)$

$= 21.6 - 2.99 + 19.28$

$= 37.89.$

$y = a + b_1x_1 + b_2x_2$ (x_1, x_2) value from Table.

$= 37.89 + (0.09)(50) + (14.5)(4)$

$= 37.89 + 4.5 + 58.$

$= 100.36$

Solve a sum by Linear Regression

Least Square method :-

Equation of linear regression.

$y = a + bx$

$$b = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$\bar{y} = a + b\bar{x}$

x	y	$x-\bar{x}$	$y-\bar{y}$	$x-\bar{x} * y-\bar{y}$	$(x-\bar{x})^2$
2	3	-3	-3.25	9.75	9
4	7	-1	0.75	-0.75	1
6	5	1	-1.25	-1.25	1
8	10	3	3.75	11.25	9
$\bar{x}=5$	$\bar{y}=6.25$			19	20

eqn of linear Regression $y = a + bx$.

Here x is indedependent variable

y is dependent variable

a is intercept, b is slop or Coefficient.

$$b = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$\bar{y} = a + b\bar{x}$

$b = \frac{19}{20}$ $\boxed{b = 0.95}$

$\bar{y} = a + b\bar{x}$

$a = \bar{y} - b\bar{x}$

$= 6.25 - (0.95)(5)$

$= 6.25 - 4.75$

$a = 1.5$

$y = a + bx$.

$y = 15 + 0.95x$

Predict the final exam grade of a student who received an 86 in the mid term exam.

$y = 15 + 0.95(86) = 15 + 81.7 = 96.7$

Assignment Problem:-

The following table shows the mid term and final exam grades obtained by the students in database course. Use the method of least squares to find the eqn for the prediction of students final exam grade based on the students midterm grade in the course. predict the final exam grade of student who received as 36 in the mid term exam.

x Mid Term	y Final exam.
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90.

BAYESIAN LINEAR REGRESSION:-

Bayesian linear Regression is a Statistical technique that integrates linear regression with Baysian inference.

It accounts for uncertainity in model parameter by considering them as random variable and assigning probability distribution to them.

Bayesian linear Regression is particularly valuable in machine learning (ML) for several reason especially when dealing with incertainly incorporating prior knowledge and improving model interpretability.

Bayesian models provide more interpretable results because they produce distributions over parameters.

A traditional linear regression model give you a single predicted price, but it tell how confident the model is in that prediction.

Bayesian linear regression, on the other hand would provide a distribution of possible prices, reflecting the uncertainity in the maket. This probabilistic prediction is far more useful in making informed decision, such as risk Management.

Prior / Parameters belief

Linear Regression

* To predict an outcome y (eg: income) based on an i/p x (eg: years of experience)
* The relationship is modeled as $y = B_0 + B_1 x + \epsilon$
* Where B_0 is the interrupt B_1 is the slope and ϵ is the error term.

Baysian Approach.

* Bayesian approach find a distribution of possible Parameters.
* Start with a prior belief about the Parameter and update this belief using the data to get a Posterior distribution.

Posterior = (likelihood * Prior)/Normalization.

Model Specification:-

Assume a linear relationship b/w the i/p. & the o/p y; $y = B_0 + B_1 x + \epsilon$

where

B_0 is intercept

B_1 is slope

ϵ is error term, typically assumed to be normally distributed. $\epsilon \sim N(0, \sigma^2)$

(8)

(2) Prior Distribution:-

Assign Prior distribution to the model Parameters B_0 & B_1. These represent your initial belief about the parameters before seeing the data.

$$B_0 \sim N(\mu_0, \sigma_0^2)$$

$$B_1 \sim N(\mu_1, \sigma_1^2)$$

where μ_0, μ_1 are the mean & σ_0^2, σ_1^2 are the variation of the period.

(3) Likelyihood:

The likelihood function represents the Probability of observing the data given the Parameter assuming normally distributed errors.

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2).$$

This is the likelihood of observing y_i given x_i & the parameter B_0 & B_1

Posterior Distribution:-

Using Bayes theorem, combine the Prior and the likelihood to obtain the posterior distribution of the parameters.

$$P(B_0, B_1 \mid data) \propto P(data (B_0, B_1) \times P(B_0) \times P(B_1).$$

Prior $P(B_0)$ and $P(B_1)$

likelihood $P(data \mid B_0, B_1)$

Posterior $P(B_0, B_1 \mid data)$

This formula provides the updated belief about the parameters after considering the observed data.

Prediction:-

To Predict the output y^* for a new x^* use the posterior distribution of the Parameter.

$$y^* \sim N(\hat{B}_0 + \hat{B}_1 x, \sigma^2_{Predictive}).$$

Posterior means.

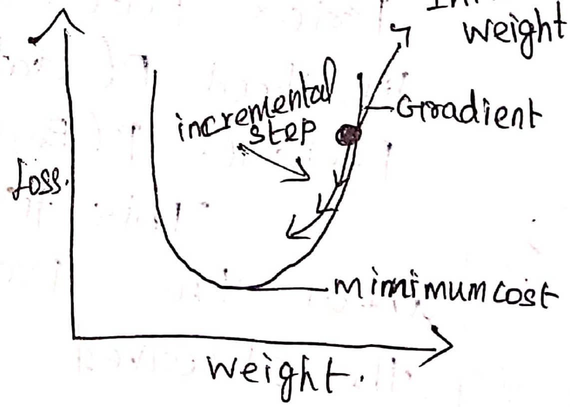
GRADIENT DESCENT:-

It is defined as one of the most Commonly used iterative optimization algorithm of ML to train the ML and DL models. It helps in finding the local minimum of a function. (Reduce Cost) or (loss function)

If we move towards a negative gradient or away from the gradient of the function at the Current point, it will give the local minimum of the function.

Whenever we move towards a Positive gradient or towards the gradient of the function at the Current point, we will get the Local maximum of the function.

The main objective of using a gradient descent algorithm is to minimize the cost function using iteration.



Predictive:- Accounts for both the uncertainity in the parameter and the noise in the data. This approach allows for the incorporation of prior knowledge and the estimation of uncertainity in the model parameter and prediction.

Adv

* Incorporates prior knowledge.
* Probabilistic Prediction
* Natural regularization
* Model Comparision
* Online Learning.

DisAdv

* Computational cost
* Prior specification
* Interpretation
* Expertise
* Sensitivity to outliers.

Linear class is ML algorithm used to classify data into categories by drawing a linear boundary. It uses a linear fn to model the relationship b/w i/p features & target o/p

LINEAR CLASSIFICATION MODELS:-

Discriminate function:-

→ A classification algorithm that make it classification based on a linear predictor function, combining a set of weight with the feature vector.

→ A linear classifier does classification decision based on the value of linear combination of characteristics.

→ Imagine that the linear classifier will merge into it's weights all the characteristics that define a particular class.

→ Linear classifiers can represent a lot of things but they can't represent everything.

→ The classic example of what they can't represent is the XOR function.

LDA is a supervised learning algo which means that it requires a labelled set of data points in order to learn the linear discriminant function.

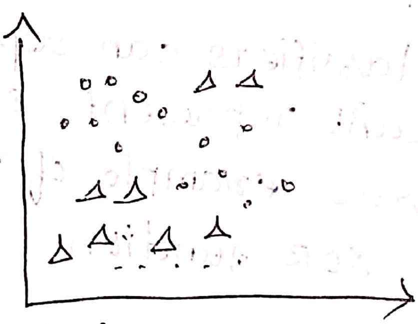
The main purpose of LDA is to find the line or plane that best seperates data points belonging to different classes.

The key idea behind LDA is that the decision boundary should be chosen such that it maximize the distance b/w the means of two classes while simultaneously minimizing the variance within each class data or within class scatter. This criterion is known as the fisher Criterian.

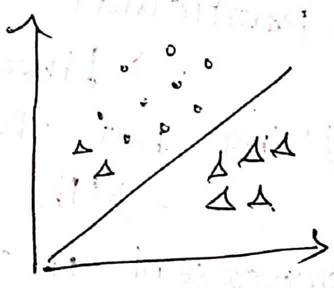
LDA is one of the most widely used ML algorithm due to its accuracy and flexibility

LDA can be used for a variety of task such as classification, dimensionality reduction and features selection.

Suppose we have two classes & we need to classify them efficiently, then using LDA, classes are divided as follow.



Before LDA



After LDA

Steps in LDA Algorithm:-

i) The first steps is to calculate the mean & Standard deviation of each feature.

ii) within class scatter matrix and b/w class scatter matrix is calculated.

(iii) These matrices are then used to calculate the eigenvectors and eigen values.

iv) LDA chooses the k eigenvectors with the largest eigenvalues to form a transformation matrix.

v) LDA uses this transformation matrix to transform the data into a new space with k dimension.

vi) once the transformation matrix transform the data into new Space with k dimension, LDA can then be used for classification or dimensionality reduction.

Probabilistic Discriminative Model:-

Discriminative model are a class of Supervised ML Models which make predictions by estimating conditional probability $P(y|x)$. In order to use a generative model, more unknowns should be solved. one has to estimate probability of each class and Probability of observation given class.

For two-class classification problem, the Posterior probability of class C_1 can be written as a logistic sigmoid action on a linear function of x.

$$P(C_1/x) = \sigma\left[\ln \frac{P(x|c_1)\,P(C_1)}{P(x|c_2)\,P(C_2)}\right] = \sigma\left(w^T x + w_0\right).$$

Logistic Regression:-

Logistic regression is a supervised algorithm that accomplishes binary classification tasks by predicting the probability of an outcome event or observation. The model delivers a binary outcome limited to two possible outcomes: yes/no, 0/1 etc...

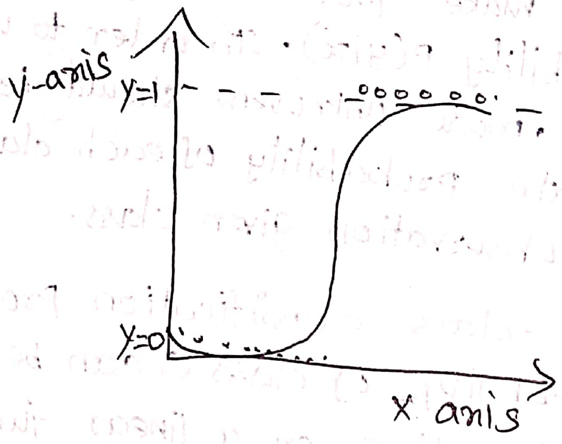
When considering the two class problem using a generative approach and under general assumption, the posterior probability of class C_1 is re-written:

: a logistic sigmoid on linear function of the feature vector $\phi = \phi(x)$:

$$P(C_1|\phi) = y(\phi) = \sigma(w^T\phi) \text{ with } P(C_2|\phi) = 1 - P(C_1|\phi).$$

→ The logistic sigmoid function is defined as

$$\sigma(a) = \frac{1}{1+\exp(-a)} \text{ with } a = \ln \frac{P(\phi|C_1)\,P(C_1)}{P(\phi|C_2)\,P(C_2)}.$$



Generative Model:-

(11)

A generative Model is a statistical model of the Joint probability distribution $p(x,y)$ on given observable variable x and target variable y.

Models with linear decision boundaries arise from assumption about the data.

In generative approach to classification we first model the class - conditional densities $P(x|C_k)$ and the class prior $p(C_k)$, and then we compute posterior probabilities $p(C_k|x)$ through Baye's theorem.

Naive Bayes:- (Refer unit-2)

i) Naive Bayes Algorithm
(ii) Baye's theorem
(iii) Working of Naive Baye's classifier
iv) Problem
v) App, Adv, Dis Adv.

Maximum Margin classifier:-

The maximal margin classifier is the optimal hyperplane defined in the (rare) case where two classes are linearly seperable.

Given an $n \times p$ data matrix X with a binary response variable defined as $y \in [-1, 1]$, it might be possible to define a p- dimensional hyperplane.

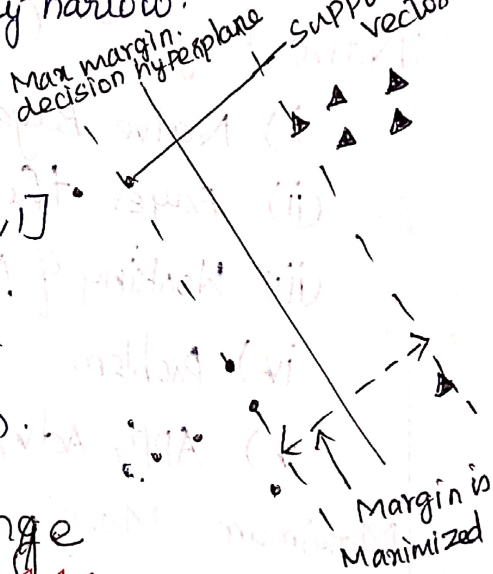
$$h(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$
$$= x_i^T \beta + \beta_0 = 0.$$

This seperating hyperplane has the property that if β is constrained to be a unit vector $||\beta|| = \sum \beta^2 = 1$, then the product of hyperplane response variable are positive perpendicular distance from the hyperplane, the smallest of which may be formed the hyperplane margin, M.

The maximal margin classifier is the hyperplane with the maximum margin max [M] subject to $||\beta|| =$

A seperating hyperplane rarely exist. In fast, even if a seperating hyperplane doest exist, its resulting margin is probably undesirably narrow.

Here is the maximal margin classifier the data set has 2 linearly seperable classes $y \in [-1, 1]$ described by 2 features $x_1 \& x_2$. This code is unimportant - Just trying to produce the visualization.

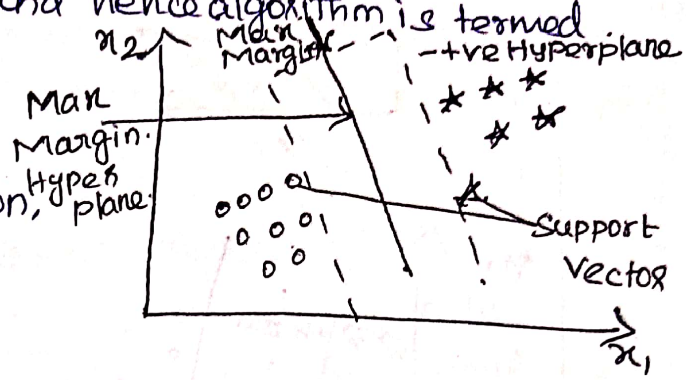


Support Vector Machine

SVM is one of the most popular supervised Learning algorithm, which is used for classification as well as regression problems. However, primarily it is used for classification problem in ML.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vector that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed Support vector Maching



* SVM algorithm can be used for face detection, image classification ~~fact~~ text categorization etc.

Types of SVM:-

1) Linear SVM:- linear svm is used for linearly seperable data, which means if a dataset can be classified into two classed by using a single straight line, then such data is termed as linearly seperable data and classifier is used called as linear SVM cdassifier.

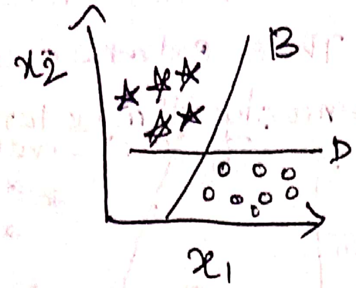
Non linear SVM:-
It is used for non linearly seperated data which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used called Non-linear SVM classifier.

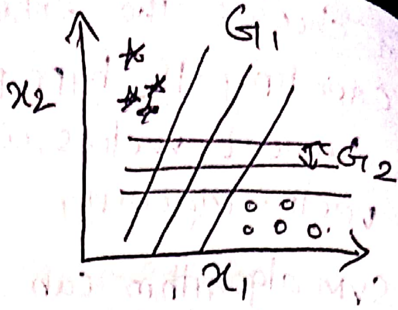
LINEAR SVM:-

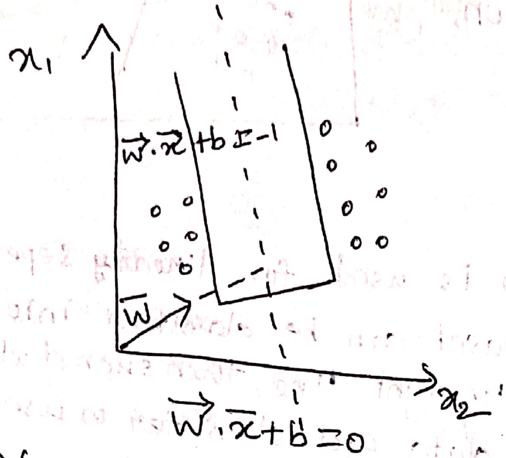
The Working of the SVM can be understood by using an example.

Suppose we have a dataset that has two tags (Green & blue) and the dataset has two features x_1 and x_2

We ~~cannot~~ want a classifier that can classify the pair $(x1, x2)$ of coordinates in either green or blue.







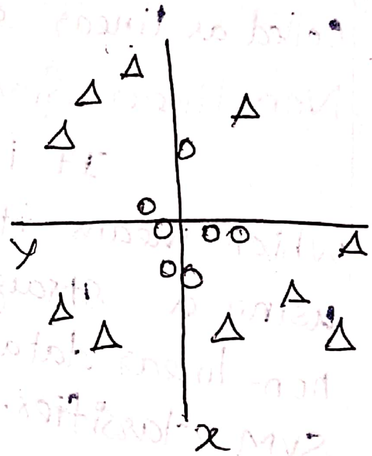
$$\max \frac{2}{\|W\|}$$

$(W \cdot x + b) \geq 1 \;\forall x$ of class 1

$(W \cdot x + b) \leq -1 \;\forall x$ of class 2

NON LINEAR SVM:-

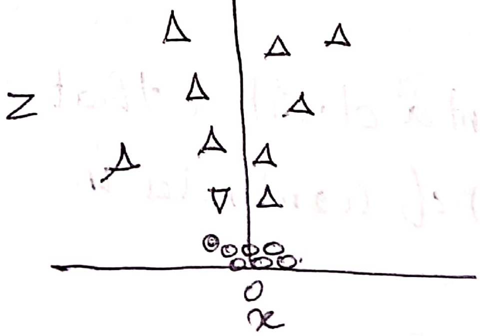
If data is linearly arranged then we can seperate it by using a straight line, but for non-linear data, we cannot draw a single straight line

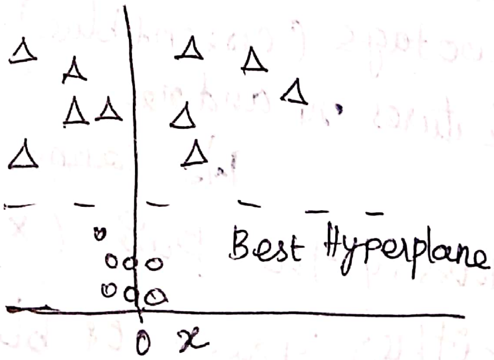


So to seperate these data Points, we need to add one more dimension.

For linear data, we have used two dimension x and y, so far non-linear data, we will add a third dimension z.

It can be calculated as $z = x^2 + y^2$

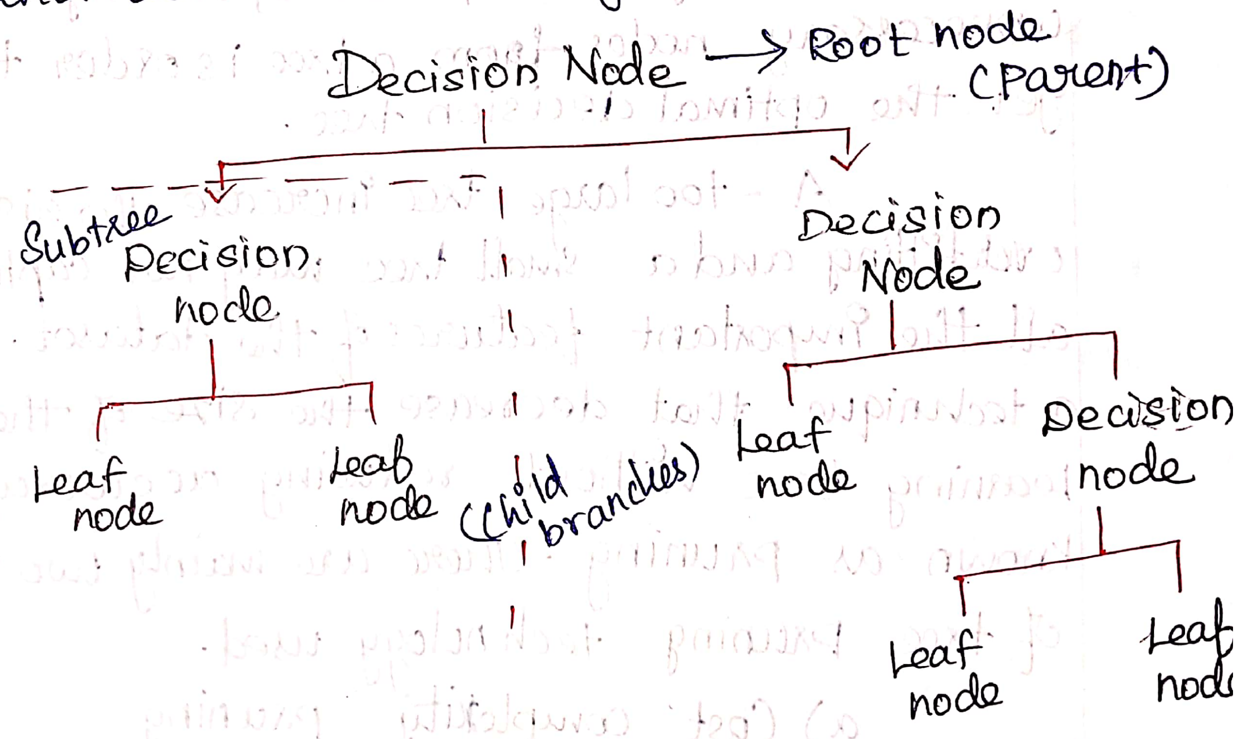




DECISION TREE:-

Decision Tree is a Supervised learning technique that can be used for both classification and regression problem, but mostly it is preferred for solving classification problems. It is a tree-Structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represent the outcome.

In a Decision tree, there are 2nodes which are the decision Node and leaf node. Decision nodes are used to make any decision and have multiple branches whereas leaf nodes are the output of those decision and do not contain any farther branches.



Decision Tree Terminologies:-

1) Root Node:- Root node is from where the decision tree starts. It represents the entire dataset which further gets divided into two or more homogeneous sets.

2) Leaf node:- Leaf nodes are the final output node, and the tree cannot be ~~seperated~~ Segregated further after getting a leaf node.

3) Splitting:- Splitting is the process of dividing the decision node/root node into sub-nodes according to the given condition.

Branch/Sub tree:- A tree formed by splitting the tree.

Pruning: Getting an optimal Decision Tree:-

Pruning is a process of deleting the unnecessary nodes from a tree is order to get the optimal decision tree.

A - too large tree increase the risk of overfitting and a small tree may not capture all the important features of the dataset. Therefore a technique that decrease the size of the learning tree without reducing accuracy is known as pruning. There are mainly two types of tree pruning technology used.

a) Cost complexity pruning.

b) Reduced Error Pruning.

Adv

* Easy to understand and Interrupt
* Handles Different Data types.
* Less Data preperation.
* Handles Non linear relationship.
* Feature Importance.

Disadv.

* overfitting.
* Instability
* Bias towards Dominant classes.
* Greedy approach
* Limited experiencess.

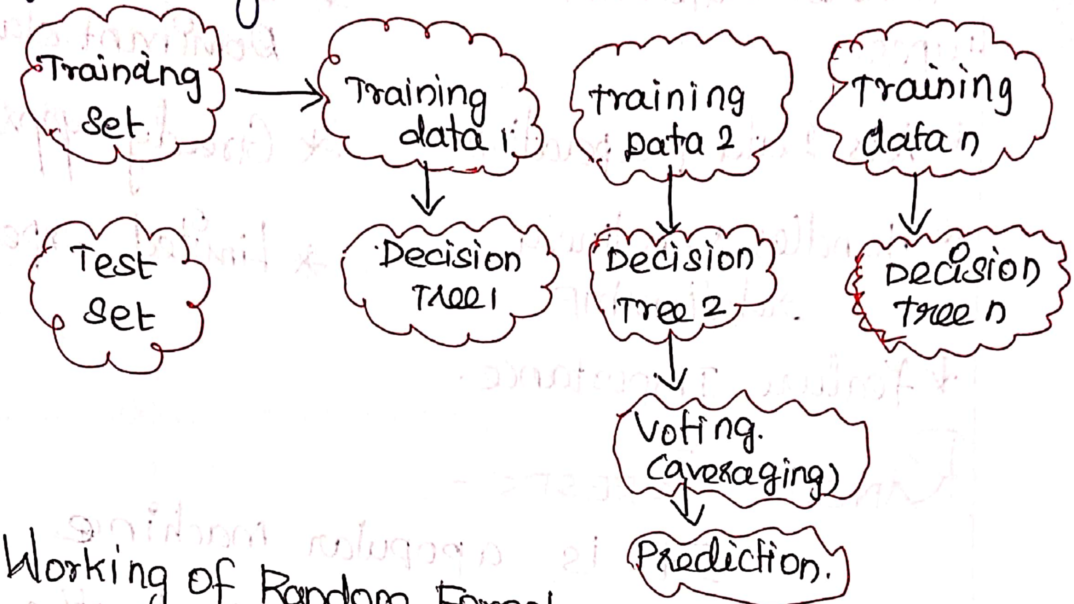
RANDOM FORESTS:-

It is a popular maching learning algorithm that belongs to the Supervised learning technique. It can be used for both classification and regression Problem in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classification to solve a complex problem and to improve the performance of the model.

Random forest is a classifier that contains a no of decision tree on various Subset of the given dataset and takes the average to improve the predictive accuracy of the dataset. Instead of relying on one decision tree, the random forest takes the

Prediction from each tree and based on the majo Votes of prediction, and it predicts the final outp

The greater number of trees in the forest leads to higher accuracy and Present the problem of overfitting.



Working of Random Forest:

Random forest work in two phase first is to Create the random forest by combining N Decision tree, and second is to make prediction for each tree created in the first phase.

The working process can be explained in the below steps and diagrams:-

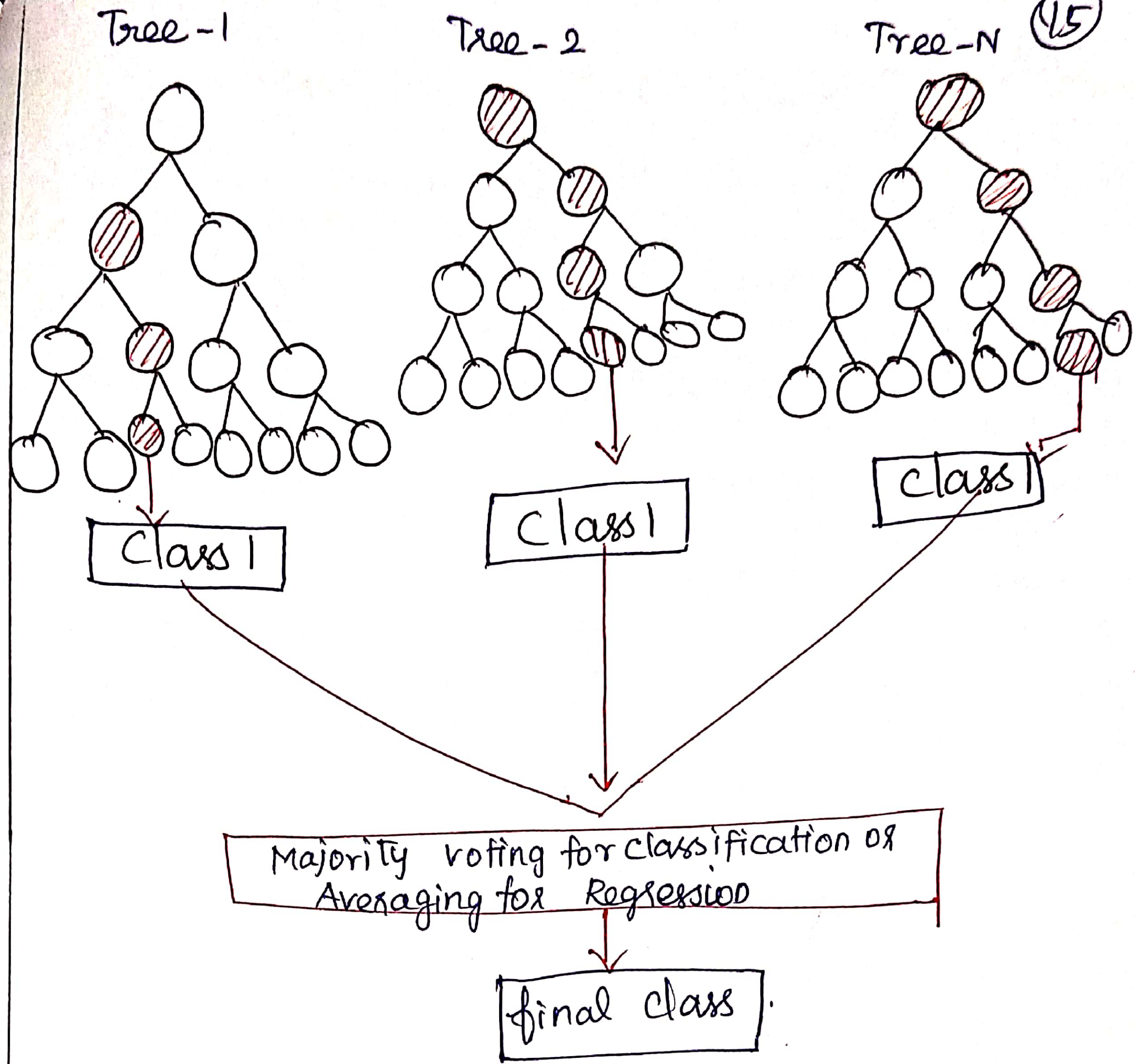
Step1: select random k data point from the training set.

Step2: Build the decision ~~task~~ tree associated with the selected data points.

Step3: choose the number N for decision trees that you want to build.

Step4:- Repeat Step1 & Step 2

Step5:- For new data points find the prediction of each decision tree and assign new data point to the category that win the majority votes.



Application: * Banking, * Medicine * Marketing.

Adv:-

* High Accuracy
* Robustness to overfitting
* Handle various data types
* Handle Missing values
* Feature importance.
* Scalability.

DisAdv

* Complexity & interpretability
* computationally intensive
* Resource requirements
* Prediction speed.
* May not perform well on linear relationship.