16-March-2023

Date: / / Page no:

Machine Learning

UNIT :- 1

Introduction to Machine Learning

Learning :-

* Learning is the process of acquiring new knowledge, behaviours, skills, values, attitude & preferences.

* Learning process happens when you observe a phenomenon & recognize a pattern.

* Learning process includes gaining of new symbolic knowledge & development of cognitive skills through instructions & practice. It is also discovery of new facts & theories through observation & experiments.

* Learning is a phenomenon & process which has manifestations of various aspects.

* All human learning is observing something, identifying a pattern, building a theory model to explain these pattern & testing these theory to check if its fits in most of all observation.

Date: / / Page no:

Observation ——→ Learning ——→ Skill

* Both human as well as machine learning generate knowledge one residing in the brain & other residing in the machine.

* Human learning process varies from person to person. Once a learning process is said into the mind of people, it is difficult to change.

Types of Human learning

Human learning takes place in following ways:-

1° Self learning:-

Human try many times after multiple attempts, some being unsuccesful.

2° Knowledge gained from expert:-

We build our motion indirectly based on what we have learned from expert in the past.

Date: / / Page no:

3. Learning directly from Expert :-
Either somebody who is an expert in the subject directly teaches us.

17-March-2023

Difference between Human & Machine learning :-

Human learning	Machine learning
Humans acquire knowledge through experience either directly or shared by others.	Machines acquire knowledge through experience shared in the form of past data
Model free & model based mechanism can be found in human learning.	Knowledge based learning in machine learning.
Observation → Learning → skill	Data → Machine learning → skill

Machine learning :-

→ Machine learning is growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithm for building mathematical model & making predictions using historical data or information.

→ Currently it is being used for various task such as image recognization, speech recognization, email filtering, facebook autotagging and many more.

18-March-2023

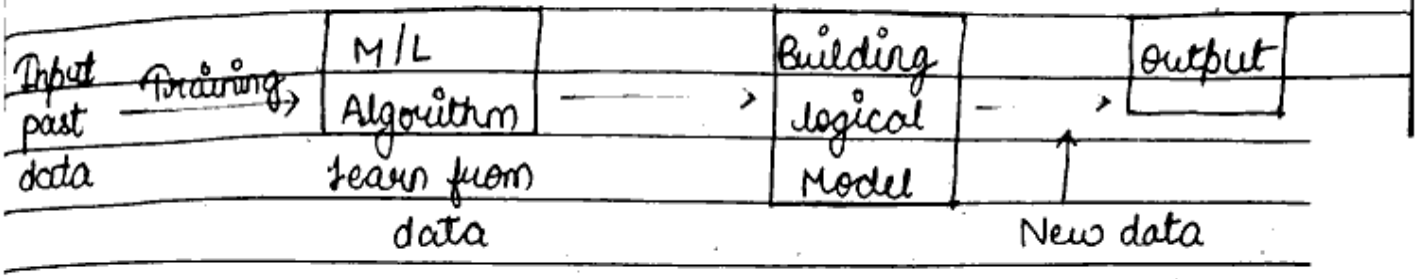
Machine learning working :-

→ A machine learning system learn from ~~model & whenever it receives a new data~~ historical data, builds the predication model & whenever it receives new data, predicts the output for it.

→ The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model, which predicts the o/p more accurately.

20 - March - 2023

Date: / / Page no:



Classification of Machine learning :-

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

21 - March - 2023

1. Supervised learning :-

It is a type of machine learning method in which we provide sample, labelled data to the machine learning system in order to train it, & on that basis, it predicts the output.

The system creates a model using labelled data to understand the data sets & learn about each data, once the training & processing are done then we test the model by provide a sample data to check whether it is predicting the exact output or not. The goal of supervised learning is to map the input data with the output data.

Date: / / Page no:

The supervised learning is based on supervision. The example of supervised learning is spam filtering.
Supervised learning can be grouped further in two categories of algorithm.

- Classification
- Regression

2. Unsupervised learning :-
It is a learning method in which a machine learn without any supervision. The training is provided to the machine with the set of data that has not been labelled, classified or categorised, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns (features).
Unsupervised learning categories :-

- Clustering
- Association

22 - March - 2023

Date: / / Page no: ____

3. Reinforcement learning :-

It is a feedback based learning method in which a learning agent gets a reward for each right action & gets up punishment / penalty for each wrong action. The agent learn automatically with these feedback & improve its performance. In reinforcement learning the agent interacts with the environment & explores it. The goal of an agent is to get the most reward points & hence it improves its performance.

The example of reinforcement learning is self-driving cars.

Elements of Reinforcement learning :-

- Agent
- Environment
- Reward
- State
- Action

29- March - 2023

Date / / Page no

Difference b/w supervised, Unsupervised & Reinforcement learning.

Supervised	Unsupervised	Reinforcement
Supervised learning requires that the target variable is well-defined & that the sufficient no's of its values are given.	For unsupervised learning typically either the target variable is unknown or has only been recorded for too small a no. of faces.	Reinforcement learning is learning what to do & how to map situations to actions the learner is not told which action to take.
Supervised learning deals with two main tasks:- → Regression → classification	Unsupervised learning deals with → clustering → Association	Reinforcement learning learning deals with → Exploitation → Exploration → Markov's decision process → Policy learning → Deep learning → Value learning

Date: / / Page no:

The input data in supervised learning is labelled data.	Unsupervised learning uses unlabelled data.	The data is not predefined in reinforcement learning
learning by using labelled data.	Training using unlabelled data without any guidance	Works on ~~data~~ interacting with the environment.
Maps the labelled input to the known output.	Understands the pattern & discovers the output.	Follows the trial & error method.
		24-March-2023

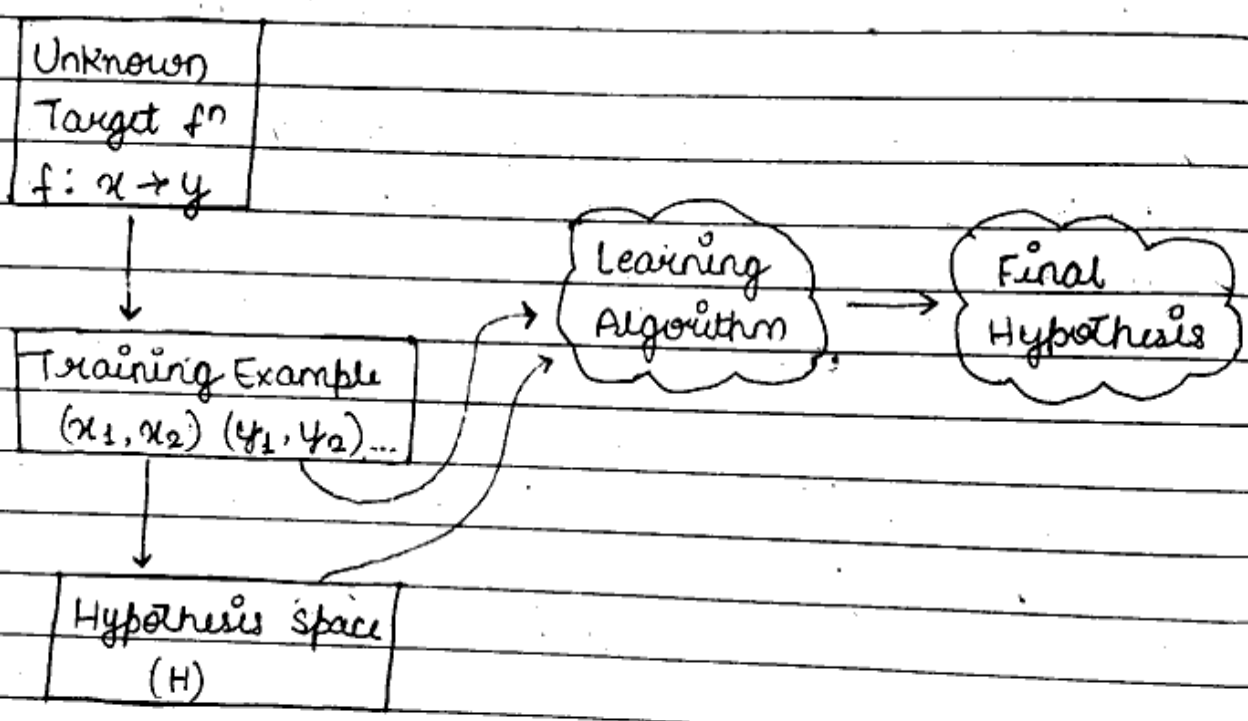
Hypothesis :

The hypothesis is defined as the sub position or proposed explanation based on insufficient evidence or assumptions. It is just a guess based on some known facts but has not yet being proven. A good hypothesis is testable, which results in true or false.

Date: / / Page no:

Hypothesis in Machine learning :-

The hypothesis is one of the commonly used concepts of statistics in machine learning. It is specifically used in supervised machine learning, wherein ML model learn a f^n that based maps the input to corresponding output with the help of an available data set.



In supervised learning techniques the main aim is to determine the possible hypothesis out of hypothesis space that best maps input to the corresponding or correct outputs.

Date: / / Page no:

Hypothesis space (H) :-

It is defined as a set of all possible legal hypothesis hence it is also known as a hypothesis set. It is used by supervised machine learning algorithms to determine the best possible hypothesis to describe the target fn or best maps input to output.

It is often constrained by choice of the forming of the problem, the choice of model & the choice of model configuration.

Hypothesis (h) :-

It is defined as the approximate fn that best describe the target in supervised machine learning algorithm it is primarily based on data as well as bias & restrictions applied to data. Hence hypothesis can be concluded as a single hypothesis that maps input to proper output & can be evaluate as well as use to make prediction.

The inductive bias (also known as learning bias)

of a learning algorithm is a set of assumptions

that a learning uses to predict output of given ~~to~~ input that it has not encountered.

In machine learning one aims to construct algorithm that are able to learn to predict a certain target output. To achieve these a learning algorithm is presented that demonstrated the intended m/p of input & output values then the learning is suppose to approximate the correct output even for examples that have not been shown during training. without any additional assumptions. This problem cannot be solved during since unseen situations might have an arbitrary output value. The kind of necessary assumptions about the nature of the target fn are sub-sumed in the phrase inductive bias.

25-March-2023

Cross-Validation in Machine learning

Cross-Validation is a technique for validating the model efficiency by training it on the subset of input data and testing on previously unseen subset of the input data.

Date / / Page no.

We can also say that it is a technique to check how a statistical data model generalize to an independent data set.

In Machine learning there is always the need to test the stability of the model. It means based only on the training data set we cannot fit our model on the training data set. For this purpose we reserve a particular sample of the data set which was not part of the training data set after that we test our model on that sample before deployment & this complete process comes under cross-validation. Hence the basic steps of cross-validation are

→ Reserve a subset of the data set as a validation set.

→ Provide the training to the model using the training data set.

→ Now, evaluate model performance using the validation set. If the model perform well with the validation set, performs the further steps, else checked for the issues.

☆ Methods used for cross-validation :-
There are some common methods that are used for cross-validation. These methods are given below :-

→ Validation set Approach
→ leave-p-out cross validation
→ leave one out cross validation
→ k-fold cross validation
→ Stratified-K-Fold cross validation

☆ Validation set Approach :-
we divide our input data set into a training set and test or validation set. In the validation set approach both the subsets are given 50% of the data set. It has one of the big disadvantages that we are just using a 50% data set to train our model so the model may miss out to capture important information of the data set

27-March-2023

☆ leave-p-out cross validation :-
In this approach, p data sets are left out of the training data. It means if there are total n data points in the original input data

set, then n-p data points will be used as the training data set & the p data points as the validation set. This complete process is repeated for all the samples, & the average error is calculated to know the effectiveness of the model. There is a disadvantage of this Technique i.e. it can be computationally difficult for the large p.

n = datasets
p = validation
n-p = Training dataset

★ Leave one out cross validation
This method is similar to the leave-p-out cross validation, but instead of p we need to take one dataset out of training. It means in this approach for each learning set only one data point is reserved & the remaining data set is used to train the model. This process repeat for each data point. Hence, for n samples, we get n diff. training set & n test set. It has the few features :-

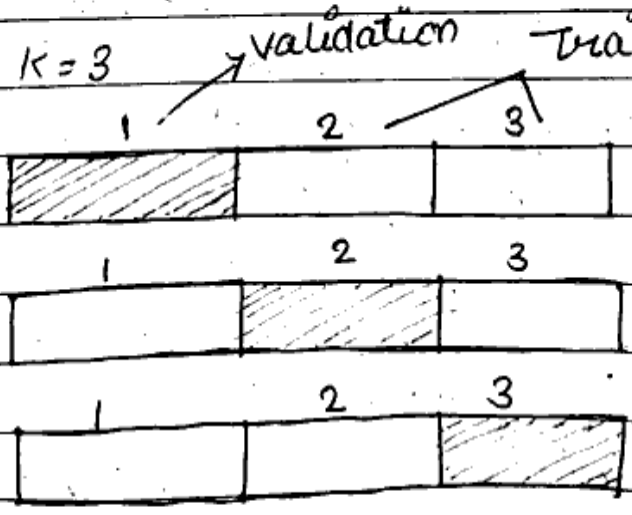
★ So this approach the bias is minimum as all the data are used.

→ The process is executed for n times. Hence, execution time is high.

→ This approach leads to high validation in testing the effectiveness of model as we iteratively check against one data point.

★ K-fold cross validation:-

This approach divides the input data sets into K groups of samples of equal size. This samples are c/d folds. For each training set, the prediction f° uses K-1 folds & the rests of the folds are used for test set



The steps for K-fold cross validation are:-

1° Split the input data set into K groups (folds).

2. For each group
   → Take one group as the reserve or test data set.
   → Used remaining groups as the training data set
   → fit the model on the training set & evaluate the performance of the model using the test set.

28-March-2023

Limitations of cross-validation :

There are some limitations of the cross-validation techniques, which are given below :-

→ For the ideal condition, it provides the optimum output. But for the inconsistent data it may produce a drastic result. so, it is one of the big disadvantages of the cross-validation. As there is no certainty of the type of data in ML.

→ In predictive modelling, the data evolves over a period, due to which it may face the differences b/w the training set & validation set. such as if we create a

model for the prediction of stock market values & the data is trained on the previous five years stock values. But the realistic future values for the next five years may drastically diff. so it is difficult to aspect correct output for such situations.

#

Applications of cross validation:

→ This Technique can be used to compare the performance of diff. predictive modelling method

→ It can great scope in the medical research field.

→ It can also be used for the meta analysis as it is already being used by the data scientist in the field of medical statistics.

29 - March - 2023

Dimensionality Reduction :-

The no. of input features, variables column present in a given dataset is known as dimensionality & the process to reduce this feature is c/d dimensionality reduction.

A dataset contains a huge number of input features in various cases which makes the predictive modelling task more complicated b/c it is very popular to visualize & make prediction for the training dataset with a high number of features for such cases dimensionality reduction techniques are required to use.

Dimensionality reduction techniques can be defined as it is a way of converting the higher dimension dataset. It provide similar information. This techniques are widely used in machine learning for obtaining a better fit predictive model while solving the classification & regression problem.

It is commonly used in the fields that deal

with high dimensional data such as speech recognition, signal processing, Bio-informatics, etc. It can be also used for data visualization, noise reduction, cluster, analysis, etc.

There are many methods to perform dimensions :-

★ Missing Values

→ while exploring data if we encounter, missing values, we do what our first step should be to identify the reason then input missing values/ drop variables using

→ It has great scope in medical research field.

→ It can also be used for the meta analysis as it is already being used by data scientist in the field of medical statistics.

appriate methods by what if we have too many missing values? should we inputa missing values as drop the variables.

Date / / Page no.

☆ Low Variance
Lets think of a scenario when we have a constant variable in our data set.

☆ Decision Tree
It can be used as a ultimate sol? tackle multiple challenges like, missing values, outliers & identifying significant variable.

☆ Random forest
Decision tree is a random forest, both similar.

☆ High Correlation
Dimensions exhibiting high correlation can lower down the performance of model.
Moreover it is not good to have multiple variables of similar info. or variation also known as Multicolinarity.

Advantages of Dimensionality Reduction :-
→ It helps in data compression & hence reduce storage, space,
→ It reduces computation time.
→ It also helps remove redundent features, if any.

Disadvantages of Dimensionality Reduction :-

→ It may lead to some amount of data lose.

Subset section :-

If you have a big dataset it is hard to say which variables are important with respect to the model. You need to select imp one's & drop the other.

Shrinkage Method :-

The idea is to shrink some of the parameters to zero. It uses an optimization formula. It is also a good method for improvement of prediction accuracy, but it is hard to say it includes interpretability b/c of very small coefficient in model. It is based on more mathematical methods.

The objective f^n is necessary to implement shrinkage method we will use some squared error f^n, Typical objective f^n.

Date / / Page no:

Principle component analysis :-

If the original data can be reconstructed from the compressed data without any loss of information, the data reduction is c/d as lossless. If instead we can reconstruct only an approx. of original data then data reduction is c/d lossy.

lossy dimensionality reduction method are,
→ Principle component analysis (PCA)
→ Wavelet transforms

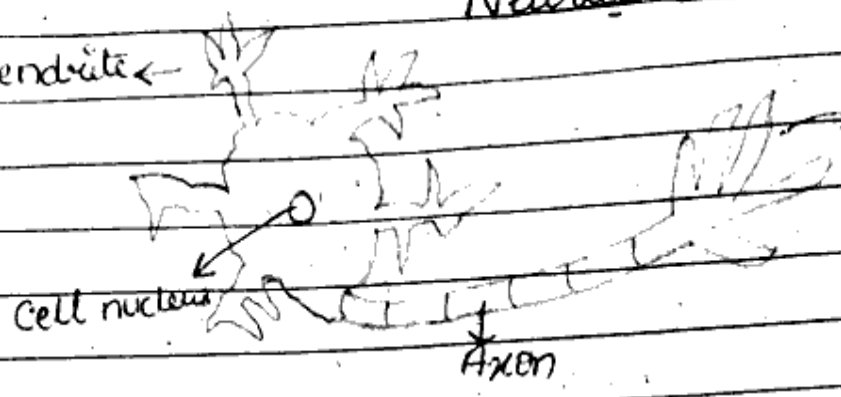
☆ PCA

Used to reduce the dimension of a data set by finding a new set of var. smaller than the original set of var. retail most of the sample information & useful for the compression & classification of data.

Date: / / Page no:

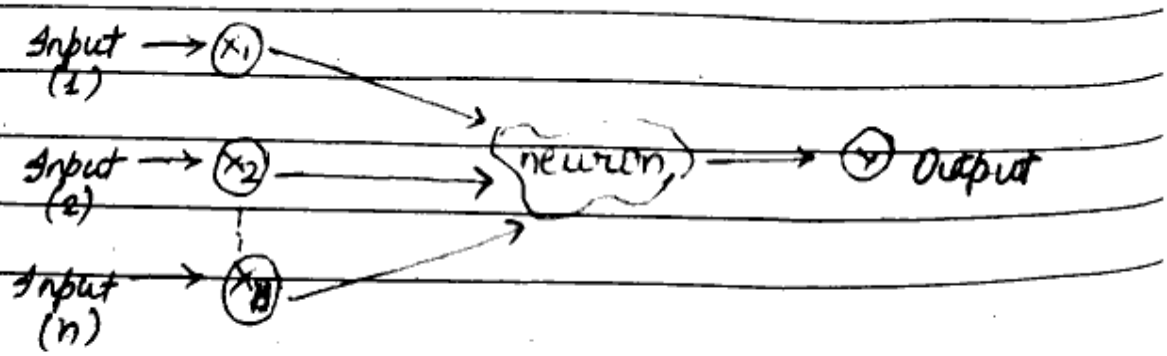
UNIT → 2

Neural Networks



The term ANN is derived from biological neural networks that develop the structure of human brain. Similar to the human brains that has neurons interconnected to one another, ANN also have neurons that are interconnected to one another in various layers of network. These neurons are known as nodes.

The above fig. illustrates the typical diagram of biological neural network.
The typical ANN looks something like the given fig.



Date / / Page no:

Dendrite from biological neural network represent input in ANN, cell nucleus represent nodes, synapse represents weight, & axon represent output,

★ Relationship b/w ANN & BNN

BNN	ANN
Dendrite	Input
Cell nucleus	Nodes
Synapse	weight
Axon	output

An ANN in the field of AI, where it attempt to mimic the network of neurons makes up a human brain so that computers will have an option to understand things & make decisions in a human like manner. The ANN is designed by programming computers To behave simply like interconnected brain cells.

There are about 100 billions neurons in the human brain, each neuron has an association point somewhere in the range of 1000 & 1 lakh (100000), in the human brain data

is stored in such a manner as to be distributed & we can extract more one piece of these data when necessary of our memory parallely.

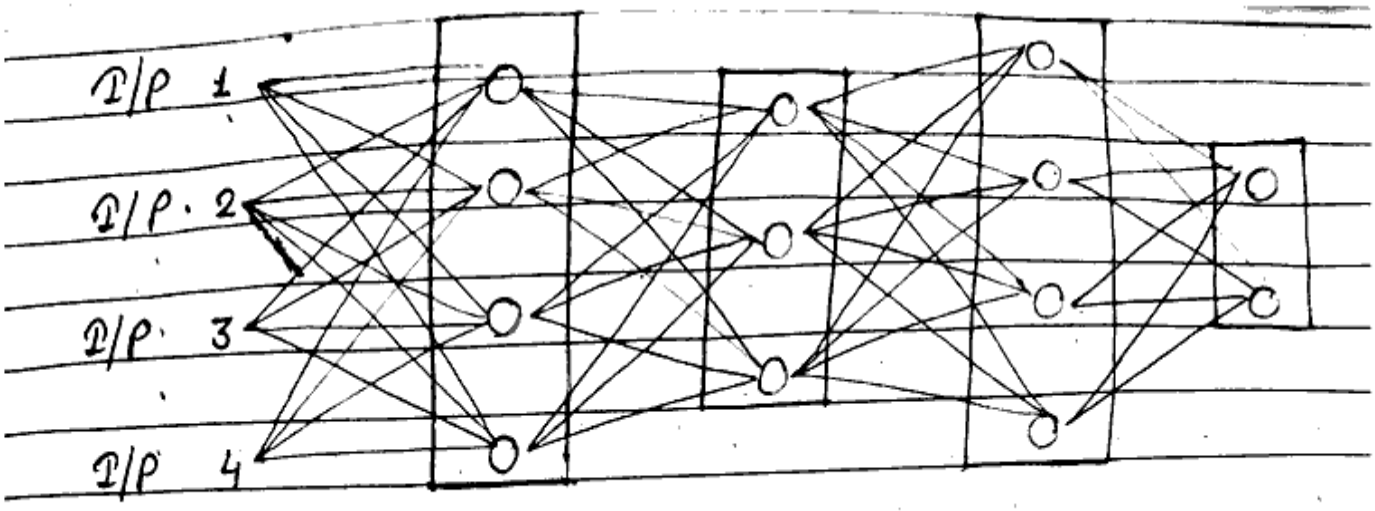
We can say that the human brain is made up of incredibly amazing parallel processor. We can understand the ANN with an example of a digital logic gate that takes an input (D/T) & gives an (O/T), "OR" gate which takes 2 input if one or both inputs are "ON" then we "ON" in O/P if both the D/P are "OFF" then we get "OFF" in O/P. Here the output depends upon D/P. our brain doesn't perform the same task the O/P to D/P. r/nship keep changing b/c of the neuron in, our brain which are learning.

The Architecture of ANN

To understand the concept of architecture of an ANN, we have to understand what a neural network consist of inorder to define a neural network that consist of a large no. of AN. which are termed unit arranged in a sequence of layer.

ANN consist of 3 layers:-

Date: / / Page no:



Input layer :-
As the name suggest, it accepts I/P in several diff. format provided by the programmer.

Hidden layer :-
The hidden layer present in b/w the I/P & O/P layer, it performs all the calculation to find hidden features & patterns.

Output layer :-
The I/P goes through a series of transformation using the hidden layer, which finally results in output i.e. convey using this layer. The artificial N.N takes input & computes the weighted sum of the I/P & include a bias.

This computation is represented in the form of transfered f^n

$$\sum_{i=1}^{n} w_i x_i + b$$

It determines weighted total is passed as an I/P to an activated f^n to produce the O/P. Activation f^n choose whether a f^n should fire (activate) or not. Only those who are fired make it to the output layer. There are using activation f^n.

Advantages of ANN:-

★ Parallel processing capabilities :-
ANN have a numerical value that can perform more than one task simultaneously.

★ Storing data on the entire network (net^k) :-
Data i.e. used in traditional programming is stored in the whole net^k not on a data base. The disappear once of a couple of pice of data in one place doesn't prevent the net^k from working.

* Capability to work with incomplete knowledge
After ANN training the info. may produce o/p even with inadiquate system or data. The loss of performance here realize upon the significant data.

* Having a memory distribution :-
For ANN is to be able to adopt, it is imp to determine the examples of / and to encourage the net^{K} acc to the desired o/p by demonstrating these examples to the net^{K}. The succession of the net^{K} is directly proportional to the choosen interface (instances), and if the event can't appear to the net^{K} in all its aspects it can produce false o/p.

* Having fault tolerance :-
Extortion of 1 or more cells of ANN doesn't prohibit it from generating o/p & this feature makes the net^{K} fault tolerance.

Date: / / Page no:

Disadvantages of ANN.

★ Assurance of proper net^k structure :-
There is no particular guideline for determining the structure of ANN. The appropriate net^k structure is accomplished through experience, trial & error.

★ Unrecognized behaviour of net^k :-
It is the most significant issue of ANN when ANN produce a testing sol^n, it doesn't provide insight concerning "why" & "how". It decreases trust in the net^k.

★ Hardware dependence :-
ANN needs processor with parallel processing power as for their structure.

★ Difficulty of showing the issue to net^k :-
ANN can work with numerical data problems must be converted into numerical value before being introduce to ANN. The presentation mechanism to be resolve here will directly impact the performance

UNIT - 3

Date: / / Page no:

Decision Tree

learning is a method for approximately target fuctions in discrete.

A decision tree is a tree where each node represents a feature, (attributes). Each link (branch) represents a decision rule & each leaf represents an outcome (categrical or continic value.

A decision tree or a classification tree is a tree in which each internal node is labelled with input features. The arcs coming from a node labelled with a feature are labelled with each of possible

A decision tree has 2 kind nodes :-

→ Each leaf node has clan label determined by majority ~~to~~ note of training,

Ex:

→ Reading that leaf

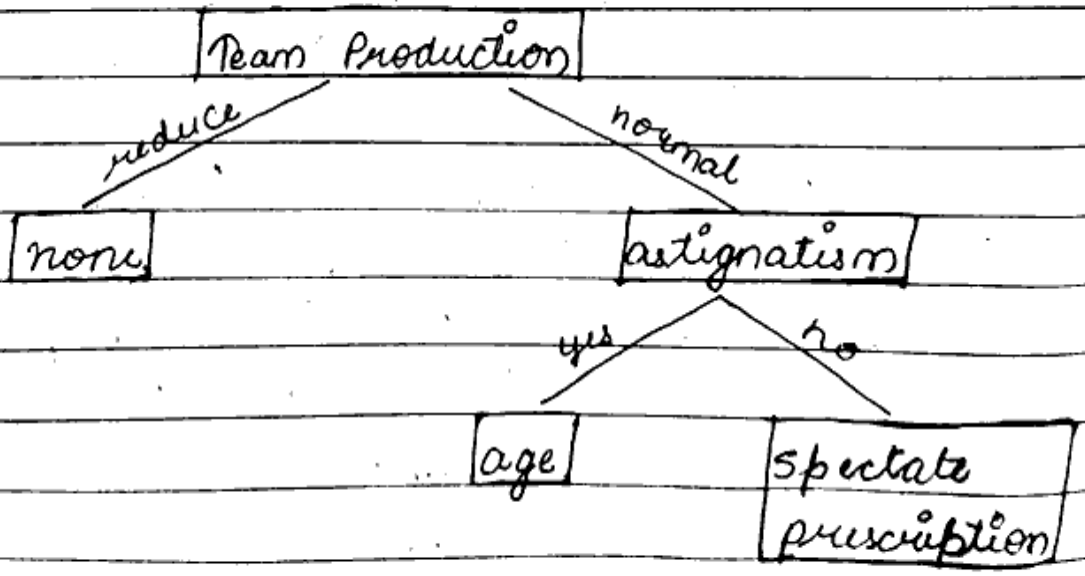
→ Each internal node is a question on features, it branches out acc. to the answer.

Date / / Page no:

→ Decision tree learning is method for approximating value discrete value target fn. The learn fn represented by a decision tree.

→ A decision tree is a tree where each non-leaf is associated with it attribute (features), each leaf node has associated with an classification. Each decision tree has associated with one of the possible value of the attribute at the node from which the arc is directed.

Example:



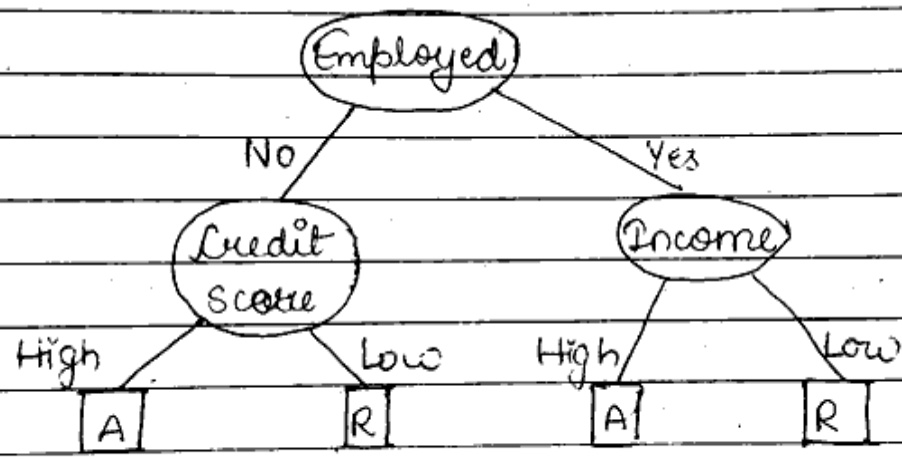
Date: / / Page no:

A decision tree is a flow chart like tree structure where each node denotes test on the attributes values each branch represents an outcome of the test & tree leaves represents classes or class distribution.

→ D.P can easily be converted to classified rules.

11-April-2023

Employed	Credit Score	Income
Y	High	High
Y	High	low
Y	low	High
N	High	low
N	low	low



There are several steps involved in the building of decision tree.

1. **Splitting** :- The process of partiting the data set into sub-set. splitting a form is particular variable. For each split two determinations are made the predictor variable c/d the splitting variable & the set of values c/d the split value.

2. Pruning:- The softening of the branches of tree. It is the process of reducing the size of the tree by turning some branch nodes into leaf nodes & removing the leaf node under the original branch.

3. Tree selection:- The process of finding the smallest tree that fits the data. Usually this is the tree that yields lowest cost validated error.

Example:-

Features
→ Age
→ Competition
→ Type
→ Profit

$$I.G = -\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$= -\frac{5}{10}\log_2\left(\frac{5}{10}\right) - \frac{5}{10}\log_2\left(\frac{5}{10}\right)$$

$$\text{Entrophy} = \frac{1}{2} + \frac{1}{2} = 1/1$$

Age	Competition	Type	Profit	
old	Yes	S/ω	Down	P=5
old	No	S/ω	Down	
old	No	H/ω	Down	
mid	Yes	S/ω	Down	
mid	Yes	H/ω	Down	
mid	No	H/ω	Up	N=5
mid	No	S/ω	Up	
new	Yes	S/ω	Up	=10
new	No	H/ω	Up	
new	No	S/ω	Up	

12 - April - 2023

Date: / / Page no:

Purpose of Decision tree:-

1° Classification

A classification tree will determine a set of logical if-then conditions to classify problems.

For ex:- Discriminating b/w three types of flowers based on certain features.

2° Regression

Regression tree is used when the target variable is numerical or continuous in nature. We fit a regression model to the target variable using each of the independent variables. Each split is made based on the sum of squared error.

Advantages of Decision Tree:-

→ Simple to understand, interpret & visualize.
→ Little effort required for data preparation.
→ Can handle both numerical & categorical data.
→ Non linear parameters don't effect its performance

Disadvantages of decision tree:-

→ Overfitting occurs when the algorithm captures

Date: / / Page no:

noise in the data.
→ The model can get unstable due to small variation in data.
→ A highly complicated decision tree tends to have a low bias - which makes it difficult for the model to work with new data.

Important Terms of Decision Tree

1. Entropy → Entropy is the measure of randomness or unpredictability in the dataset.

2. Information gain → It is the measure of decrease in entropy after the dataset is split.

3. Leaf node → Leaf node carries the classification or the decision.

4. Root node → The topmost decision node is known as the root node.

Date / / Page no:

working of Decision tree :-

→ Problem statement

To classify the diff. types of animals based on their features using decision tree.

→ How to split the data

- We have to frame the conditions that split the data in such a way that the information gain is the highest.
- Gain is the measure of decrease in entropy after splitting.

$$\sum_{i=1}^{k} P(value_i) \log_2 P(value_i)$$

Lets try to calculate the entropy for the current dataset :-

k = No. of Animals

Giraffe → 3
Tiger → 2
Monkey → 1
Elephant → 2
Total → 8

$$\Rightarrow \left(\frac{3}{8}\right) \log_2\left(\frac{3}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right) + \left(\frac{1}{8}\right) \log_2\left(\frac{1}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right)$$

Entropy → 0.571 //

Date: / / Page no:

Continue...

Age	Down	Up
old	3	0
mid	2	2
new	0	3

value of old

$$I(old) = \left[\frac{3}{3}\log\frac{3}{3} + \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right]$$

Total old in a whole dataset

$\Rightarrow 0$ $\Rightarrow 0 \times 3/10 \Rightarrow 0$

$$I(mid) = \left[\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right]$$

$\Rightarrow \frac{1}{2} + \frac{1}{2} \Rightarrow 1$ $\Rightarrow 1 \times 4/10 \Rightarrow .4$

$$I(new) = \left[\frac{0}{3}\log\left(\frac{0}{3}\right) + \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right]$$

$\Rightarrow 0$ $\Rightarrow 0 \times 3/10 \Rightarrow 0$

Entropy of Age = $0 + 4 + 0$

$\Rightarrow 0.4$

18 - April - 2023

Date: / / Page no.

Hyperplane

There can be multiple lines/ decision boundaries to segregate the classes and dimensional space but we need to find out the best decision boundary that helps to classify the data points. The best boundary is known the hyperplane.

The dimensions of the hyperplane depends on the features present in the dataset, which means if there are two features (as shown in image) then hyperplane will be a straight line. And if there are three features then hyperplane will be a two-dimensional plane.

We always create a hyperplane that has a maximum margin which means the maximum distance b/w the data points.

Support Vectors

The data points or vectors that are the closest to the hyperplane & which affect the position of the hyperplane are termed as support vectors. Since these vectors support the hyperplane hence c/d as support vectors.

Date / / Page no

Random forest Algorithm

Random forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification & regression problems in ML. It is based on the concept of Ensemble learning which is a process of combining multiple classifiers to solve a complex problem & to improve the performance of model.

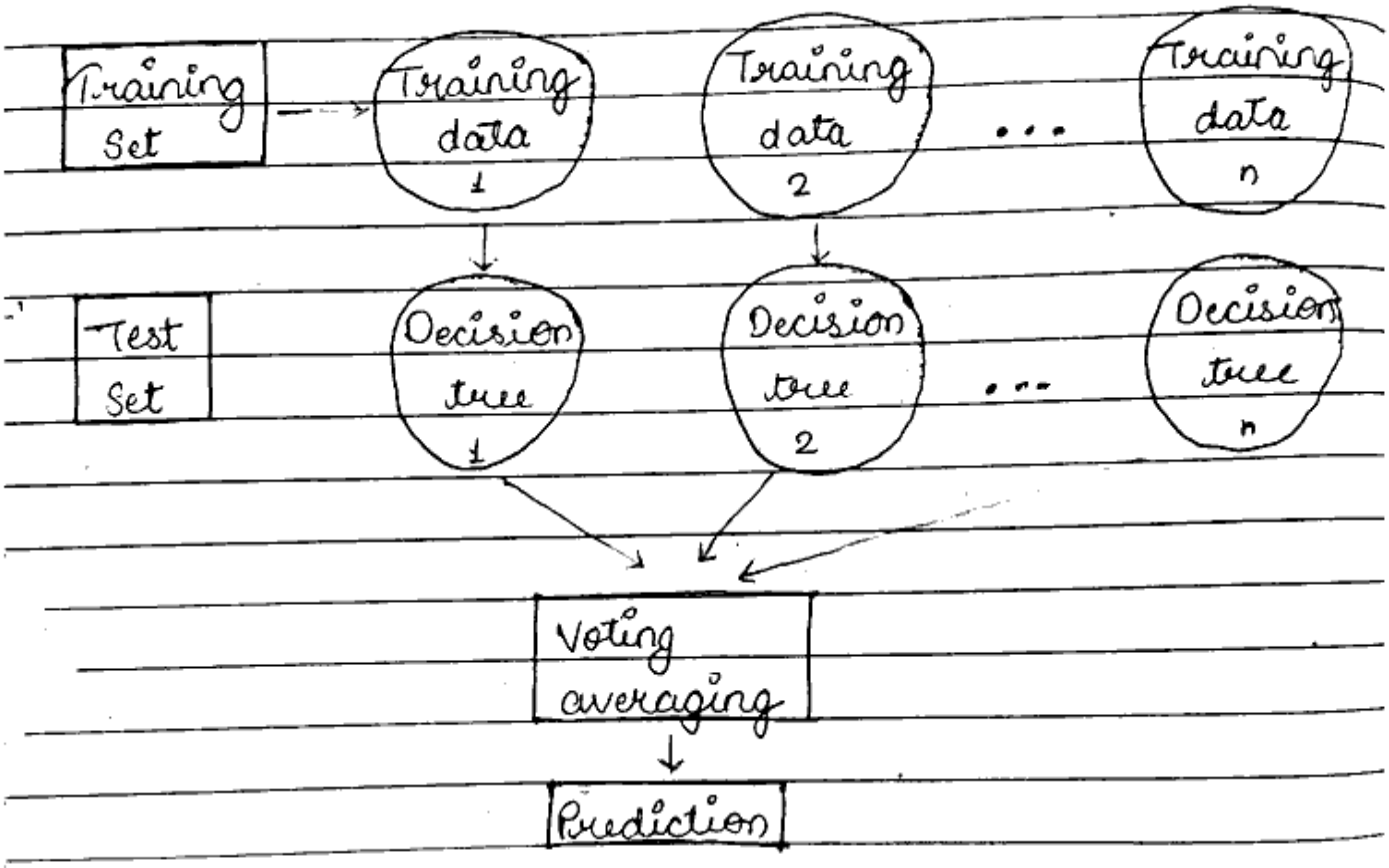
As the name suggests Random forest is a classifier that contains a no. of decision trees on various subsets of the given data set & takes the average to improve the predictive accuracy of that data set.

Instead of relying on one decision tree the random forest takes the prediction from each tree & based on the majority goals of prediction & it predicts the final output.

The greater no. of trees in the forest leads to higher accuracy & prevents the problem of overfitting.

Date: / / Page no:

The below diagram express the working of Random forest.



20-April-2023

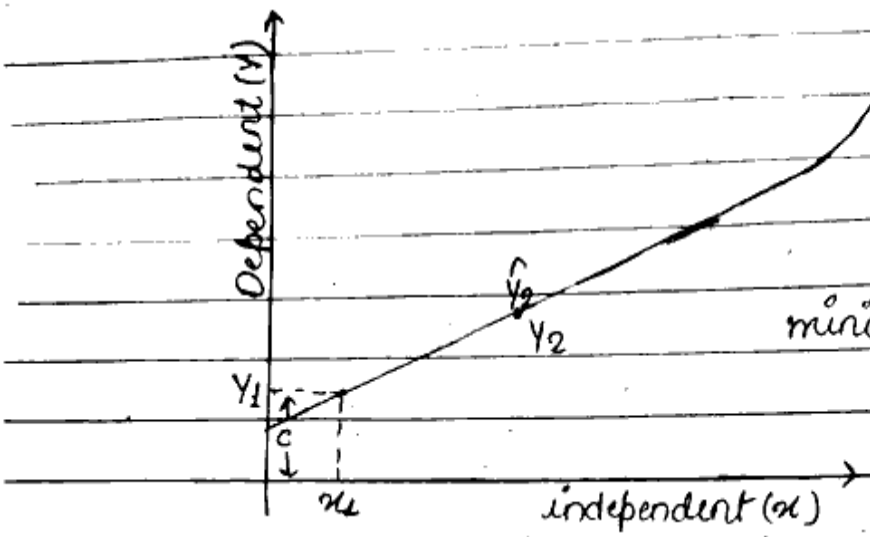
Date: / / Page no:

Ordinary least squares Method

Regression analysis is a fundamental statistical technique used in using financial to social sciences ——— . It involves modelling the r/s b/w a dependent variable & one & more independent variables. The ordinary least squares (OLS) method is one of the most commonly used technique for regression analysis.

OLS is a linear regression techniques used to find the best fitting line for a set of data points by minimizing the residuals (the diff. b/w the observed & predicted values). It does so by ~~th~~ estimating the coefficient of a linear regression model by minimizing the sum of the squared diff. b/w the observed values of the dependent variable & the predicted values from the model. It is a popular method b/c it is used & produces decent result.

Date: / / Page no:



$$Y = mx + c$$

$$\tan\theta = m$$

The OLS method can be defined as a linear regression technique that is used to estimate the unknown parameter in a model. The method relies on minimizing the sum of squared residuals b/w the actual & predicted values from the model. The residuals can be defined as the diff. b/w the actual value & the predicted value. Another word for residuals can be error.

The sum of the squared differences is also known as the residuals sum of squares. The OLS method minimizes the RSS by finding the

Date: / / Page no:

values of the coefficient that result in the smallest possible RSS. The resulting line is c/d the Regression line which represents the best fit for the data.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n + e$$

Minimize $\Sigma (y_i - \hat{y}_i)^2$

where, y_i is the actual value, $\hat{y}_i$ is the predicted value.

A linear regression model used for determining a value of a response variable $\hat{y}$ can be represented

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_n x_n + e$$

where $\hat{y}$ is the dependent variable (target variable). β_0 is the intercept. $\beta_1, \beta_2 \dots \beta_n$ are the coefficients of the independent variable. $x_1, x_2 \dots x_n$. e is the error term.

24-April-2023

Date / / Page no. :

UNIT-4

K-Mean Clustering Algorithm

	Height	Weight
1.	185	72
2.	170	56
3.	168	60
4.	179	68
5.	182	72
6.	188	77
7.	180	71
8.	180	70
9.	183	84
10.	180	88
11.	180	67
12.	177	76

Distance $= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

K_1 (185, 72)

K_2 (170, 56)

(169, 58)

(168.60)

$A = \{K_1, K_4, K_5, K_6, K_7, K_{12}\}$

$B = \{K_2, K_3\}$

Ques. write K-Mean algorithm and seperate

{5, 11, 19, 27, 23, 25, 6, 18, 2, 8, 10, 12, 31, 29, 4}

into 3 clusters.

$A = \{5, 6, 2\}$

$B = \{11, 8, 10, 12\}$

$C = \{19, 27, 23, 25, 18, 31, 29\}$

Date / / Page no:

K-Mean clustering is an unsupervised learning algo. which groups the unlabelled dataset into diff. clusters. Here, K defines the no. of predefined clusters that need to be created in the process as if K=2 there will be two clusters & for K=3 there will be three clusters & so on.

25-April-2023

It is an iterative algo. that divides the unlabelled data into K-diff. clusters. In such a way that each data set belong only one group that has similar properties. It allows us to cluster the data into diff. groups & a convenient way to discover the ~~gathe~~ categories of group in the unlabelled data set on its own without the need for any training. It is centroid based algo. where each cluster is associate with a cluster. The main aim of algo. is to minimize the sum of distance b/w the data point & their corresponding cluster.

Date / / Page no.

Hierarchical clustering in Machine Algorithm learning

Hierarchical clustering is another unsupervised ML algorithm which is used to group the unlabelled data set into a cluster & also known as hierarical cluster analysis or HCA.

So this algo we develop the hirarcy of clusters in the form of a tree & these tree saved structures is known as the Dendrogram. Sometimes the result of K-mean clustering & hierarical clustering may look similar but they both differ depending on how they work. As there is no requirement to tree determine the no. of clusters as we did in the K-means algo.

The hirachical clustering technique has two approaches.

→ Agglomerative
→ Divisive

★ Agglomerative-It is a bottom up approach. So

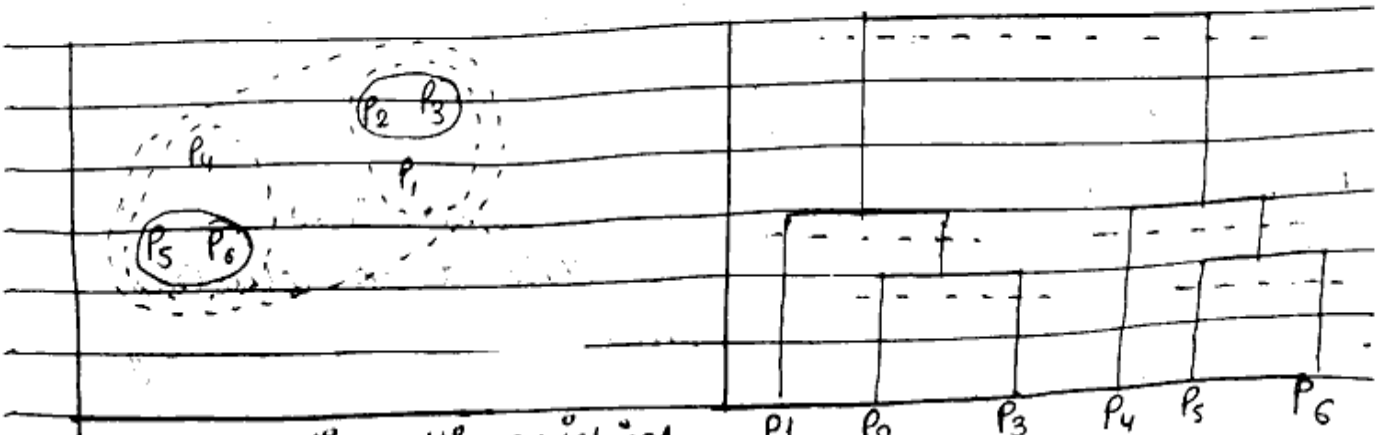
Date: / / Page no:

which the algo starts with taking all data points as single clusters & merging them until one cluster is left.

* Divisive - It is the reverse of the agglomerative algo. It is a top down approach.

Agglomerative hierarchical clustering Algo.
It is a popular example of HCA. To group the data set into clusters it follows the bottom up approach. This means this algo. considers each data set as a single cluster at the beginning & then start combining the most pair of clusters together. It does this until all the clusters are merged into a single cluster that contains all the data set.

* ~~Working of HCA:~~



Agglomerative Hierarchical Clustering (AHC)

Dataset

26- April - 2023

Date: / / Page no:

★ Working of AHC:-

The working of AHC algo can be explained with the foll steps:-

→ Create each data point as a single cluster. Lets say there are n data points. so, the no. of cluster will also be n.

→ Take two closest data points or clusters & merge them to form one cluster.

→ Again take the two cluster & merge them together to form one cluster, there will be n-2 cluster.

→ Repeat step 3 until only one cluster left. so, we will get the foll. cluster consider the below images.

→ Once all the clusters are combined into one big cluster develop the Dendrogram to divide the clusters as per the problem.

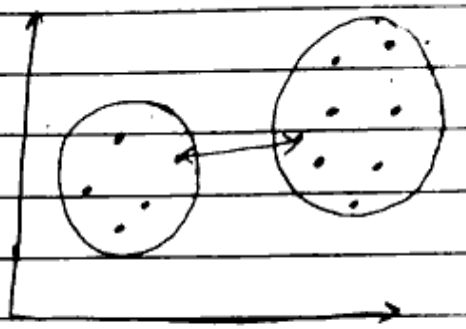
Measure for the distance b/w two clusters
The closet distance b/w the two clusters is crucial for the hierarchical clustering.

Date: / / Page no:

There are various ways to calculate the dist. b/w two clusters. And these ways decide the rule for clustering. These majors are c/d linkage Method.

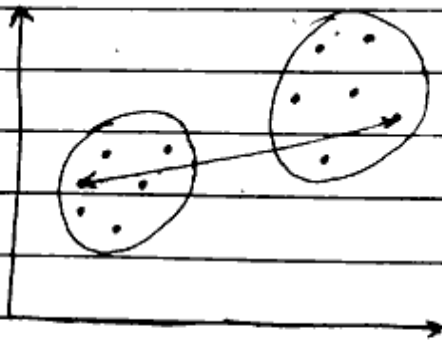
Some of the popular linkage methods are given below :-

→ Single linkage



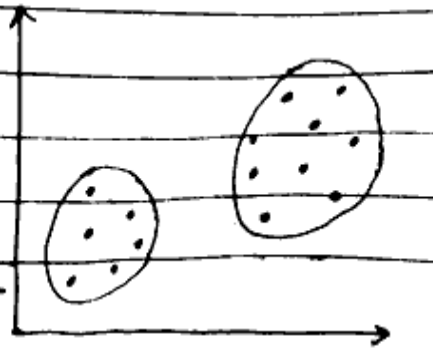
It is the shortest distance b/w the closest points of cluster consider the below image.

→ Complete linkage



It is the longest distance b/w the two data points of two diff. clusters. It is one of the popular linkage method as it forms tighter clusters than single linkage.

→ Average linkage

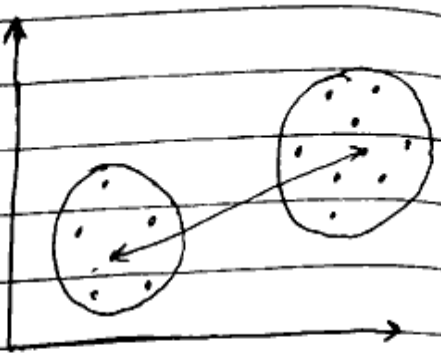


It is the linkage method in which the dist. b/w each pair of data sets is added up & then divided by the total no. of datasets. To calculate the average dist. b/w two clusters.

Date: / / Page no:

→ Centroid

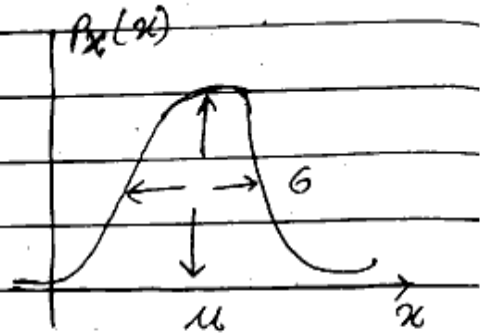
It is the linkage method in which the dist. b/w centroid of the clusters is calculated.



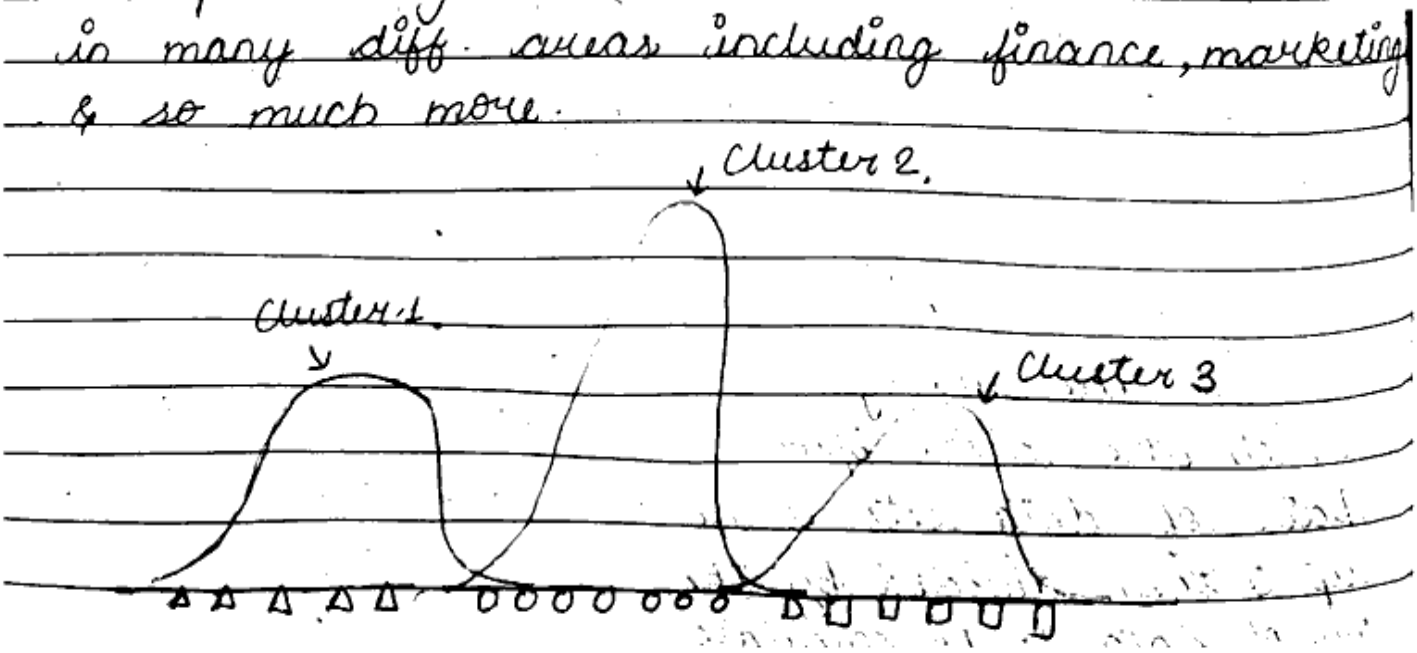
28-April-2023

Gaussian Mixture Models (GMMs)

$$P_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$



Gaussian mixture models (GMMs) are a type of ML algo. They are used to classify data in two diff categories based on the probability distribution. GMMs can be used in many diff. areas including finance, marketing & so much more.



Date: / / Page no:

GMMs are a probabilistic concept used to real world data sets. GMMs are a generalisation of causing distribution & can be used to represent any data set that can be clustered into multiple gaussian distribution.

The GMMs is a probabilistic model that assumes all the data points are generated from a gaussian distribution with unknown parameters. A GMMs can be used for clustering which is the task of grouping a set of data points into clusters. GMMs can be used to find clusters in data set where the cluster may not be clearly defined. Additionally GMMs can be used to estimate the probability that a new points belongs to each cluster. GMMs are also relatively robust outliners meaning that they can yield accurate results. Even if they are some data points that do not fit neatly into any of the clusters. These makes GMMs a flexible & powerful tool for clustering the data. It can be understood as a probabilistic model where gaussian distribution are

assumed for each group. And they have means & covariance which define their parameters.

Some real world problems which can be solved using GMMs :-
→ Finding patterns in medical data sets.
→ Modelling natural phenomenon.
→ Customer behaviour analysis.
→ Stop price prediction

Key difference b/w GMMs & the K-means algo. used for clustering
The GMMs is a type of clustering algo. that assumes that the data point is generated from a mixture of gaussian distribution with unknown parameter.
The goal of the algo. is to estimate the parameters of the gaussian distribution as well as the proportion of data points that comes from each distribution.
In contrast, K-means is a clustering algo that does not make any assumptions about the underline distribution of the data point. Instead it simply partition the data into

Date / / Page no.

K-clusters where each cluster is defined by its centroid

03-May-2023

Clustering

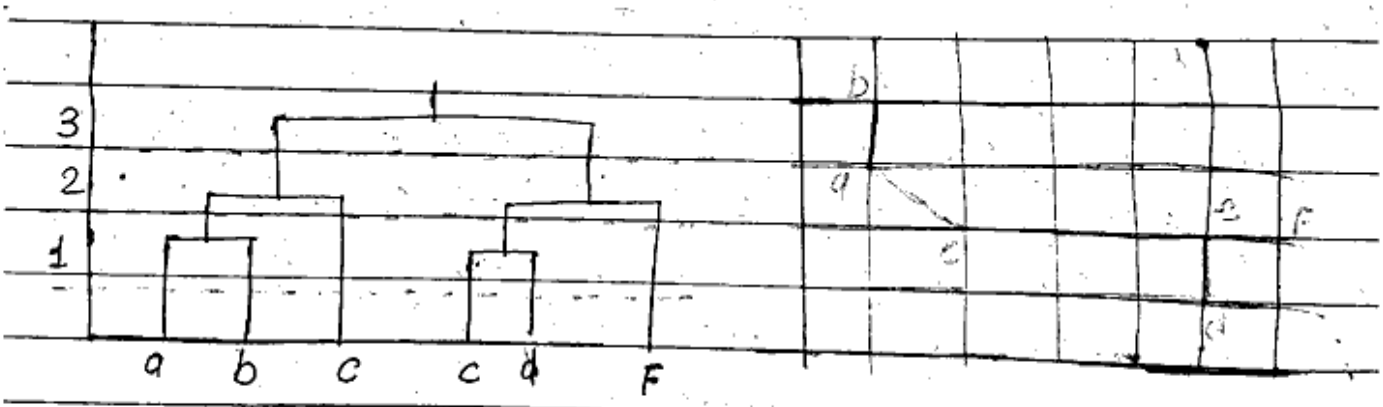
It is the task of dividing the unlabelled data or data points into diff. clusters such that similar data points fall in the same cluster than those which differ from the others. The aim of the clustering process is segregate books with similar types & assign them to cluster.

Ex:-

suppose you are the head of retail store / rental store & wish to understand the preference of your customer to scale up your business. It is possible for you to look at the details of each customer & unique business strategy for each one of them definitely not but what you can do is cluster all of your customers into 10 groups based on their purchasing habit & use the seperate strategy for customers in each of these 10 groups. And these is what we c/d clustering

Date: / / Page no:

like any learning method clustering also has its knob to adjust complexity. It is K, the no. of clusters given any K clustering will always find K centre. whether they really are meaningful groups, or whether they are imposed by the method we use there are various ways we can use to find - tune. In some applications such as colour quantization K is defined by the application.



A two dimensional data set & the dendogram showing the result of single link clustering in figure.

→ Plotting the data in 2-D using TCA may be used in uncovering the structure of data & the no of cluster in the data.

Date: / / Page no:

→ An Incremental approach may also help. setting a max. allowed dist. is equivalent to setting a max. allowed reconstruction error per instance

→ In some applications validation of the group can be done manually by checking whether clusters actually core meaningful groups of the data.

13-May-2023

Date: / / Page no:

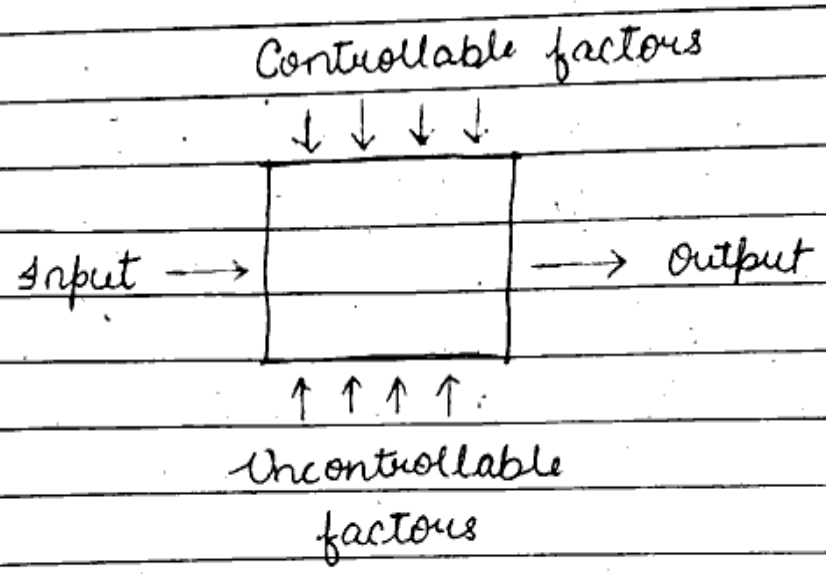
UNIT : 5

ML Experiments

In ML to we do experiments to get information about the process under scrutiny. So our case this is a learner which having bee trained on a data set generated as output for a given input. An experiment is a test or a series where we play with the factors that affect the output. These factors may be algo. used, the training set, input features / factors & so on. And we observe the changes in the response to ~~to~~ ~~not~~ be able to extract info. The aim may be to identify the most imp. factors. screen the unimportant ones or find the configuration of the factors that optimize the response.

Our aim is to plan & conduct ML experiments & analyze the data resulting from experiments to be able to eliminate the effect of chain & obtain conclusions which we can consider statistically significant. In ML we target a learner

heavy the highest generalisation accuracy & the minimal complexity & is robust i.e. minimally effects the external sources of variability. A trained learner can be shown in figure.



23-May-2023

Hypothesis Testing

Any data science project starts with exploring the data. When we perform an analysis on a sample through exploratory Data analysis & inferencial statistics we get information about the sample. Now we want to use this information to predict values for the entire population.

Date: / / Page no: .

Hypothesis testing is done to confirm our observation about the population using sample data within the desired error level. Through hypothesis testing we have enough statistical event to conclude if the hypothesis about the population is true or not.

The challenge of algorithm selection

Each model or any ML algo has several features that process the data in diff. ways. Often the data that is fed to these algo is also diff. depending on previous experiment stages. But, since ML teams & developers usually record their experiments, there's ample data available for comparison.

The challenge is to understand which parameters, data & metadata must be considered to arrive at the final choice. It's the classic paradox of having an overwhelming amount of details with no clarity.

even more challenging, we need to understand if a parameter with a high value, say a higher metric score, actually means the model is better than one with a lower score, or if it's only caused by statistical bias or misdirected metric design.

☆ Comparing ML algo's: Why ~~do~~ we do it?
Comparing ML algo's is imp. in itself, but there are some not-so-obvious benefits of comparing various experiments effectively.

→ Goals of Comparision:-

- Better-Performance
The primary objective of model comparison & selection is definitely better performance of the ML software/sol?. The objective is to narrow down on the best algo's that suit both the data & the business requirements.

- Longer lifetime
High performance can be short-lived if the chosen model is tightly coupled with

the training data & fails to interpret unseen data. so, it's also imp. to find the model that understands underlying data patterns so that the predictions are long-lasting & the need for re-training is minimal.

- Easier retraining

When models are evaluated & prepared for comparisions, minute details, & metadata get recorded which come in handy during retraining. For example, if a developer can clearly retrace the reasons behind choosing a model, the causes of model failure will immediately pop out & re-training can start with equal speed.

- Speedy Production

With the model details available at hand, it's easy to narrow down on models that can offer high processing speed & use memory resources optimally. Also during production, several parameters are required to configure the ML solutions.

☆ Parameters of ML algo's & how to compare them

- Production-based
- Development-based

→ Statistical Tests :-

On a fundamental level, ML models are statistical equations that run at great speed on multiple data points to arrive at a conclusion.

→ Null hypothesis testing :-

It is used to determine if the differences in two data samples or metric performances are statistically significant or more-or-less equal and caused only by noise or coincidence.

→ ANOVA :-

Analysis of variance, it's similar to linear discriminant analysis with the exception of the fact that it uses one or more categorical features & one continous target, providing the statistical test of whether the means of the diff. groups are similar or not.

Date: / / Page no:

→ Chi-Square :-
It's a statistical tool or test which can be used on groups of categorical features to evaluate like likelihood of association or correlation with the help of freq. distributions.

→ Student's t-Test :-
It compares the averages or means of diff. samples from normal distributions when the standard deviation is unknown to determine if the diff. are statistically significant.

→ Ten-fold cross-validation :-
The 10-fold cross-validation compares the performance of each algo. on diff. datasets that have been configured with the same random seed so as to maintain uniformity in testing.

Model features and objectives
To choose the best ML model for a given dataset, it's essential to consider the features or parameters of the model.

The parameters & model objectives help to gauge the models flexibility, assumptions, & learning style.

For Example:-
If two linear regression models are compared, one model might be aiming to reduce the mean squared error, whereas another might be aiming to reduce the mean absolute error through objective f^n's.

Compare experiments - parallel coordinates

→ Learning curves
It can help in determining if a model is on the correct learning trajectory of achieving the bias - variance tradeoff. It also provides a bias for comparing diff. ML models.

→ Bias
It is the assumption used by ML models to make the learning process easier. Variance is the measure of how much

Date: / / Page no:

the estimated target variable will change with a change in training data. The ultimate goal is to reduce both bias & variance to a minimum.

→ Less data - bias variance tradeoff
The best way to track the progress of model training is to use learning curves. These curves help to identify the optimal combinations of hyperparameters & assists massively in model selection & model evaluation.

★ The two most popular learning curves are:-

→ Training learning curve:
It effectively plots the evaluation metric score over time during a training process, thus helping to track the learning or progress of the model during training.

→ Validation learning curve :-

In this curve, the evaluation score is plotted against time on the validation set.

Sometimes it might happen that the training curve shows an improvement but the validation curve shows stunted performance. This is indicative of the fact that the model is overfitting & needs to be reverted to the previous iterations. In other words, the validation learning curve identifies how well the model is generalizing.

Date: / / Page no:

Assignment (AL - 405)

Examination, June 2022

Q.1 a) Explain the concept of hypothesis space & inductive bias in brief.

★ Hypothesis space :- (H)

It is defined as the set of all possible legal hypothesis hence it is also known as hypothesis set. It is used by supervised ML algo's to determine the best possible hypothesis to describe the best possible target fn or best maps input to output.

It is often constrained by choice of the forming of the problem, the choice of model & the choice of model configuration.

★ Inductive Bias :-

Also known as learning bias. Of a ML algo. is a set of assumptions that a learning uses to predict output of given ~~Output~~ input that it has not encountered.

So ML one aims to construct algo. that are able to learn to predict a certain target output. To achieve this a learning algo is presented that demonstrated the intended o/p of input & output values that the learning is suppose to approximate the correct output given for examples that have not been shown during training without any additional assumptions. This problem cannot be solved during since unseen situations might have an arbitrary output value. The kind of necessary assumptions about the nature of the target f° are sub-sumed in the phase inductive bias.

b) list & explain perspectives & issues in ML.

→ Inadequate Training data :-
The major issue that comes while using ML algo is the lack of quality as well as quantity of data. Although data plays a vital role in the processing

of ML algo's, many data scientists claim that inadequate data, noisy data, & unclean data are extremely exhausting the ML algo's.

→ Poor quality of data :-
Data plays a significant role in ML, & it must be of good quality as well. Noisy data, incomplete data, inaccurate data, & unclean data lead to less accuracy in classification & low-quality results.

→ Non-representative training data :-
To make sure our training model is generalized well or not, we have to ensure that sample training data must be representative of new cases that we need to generalize. The training data must cover all cases that are already occurred as well as occurring.

→ Overfitting :-
It is one of the most common issues faced by ML engineers & data scientists.

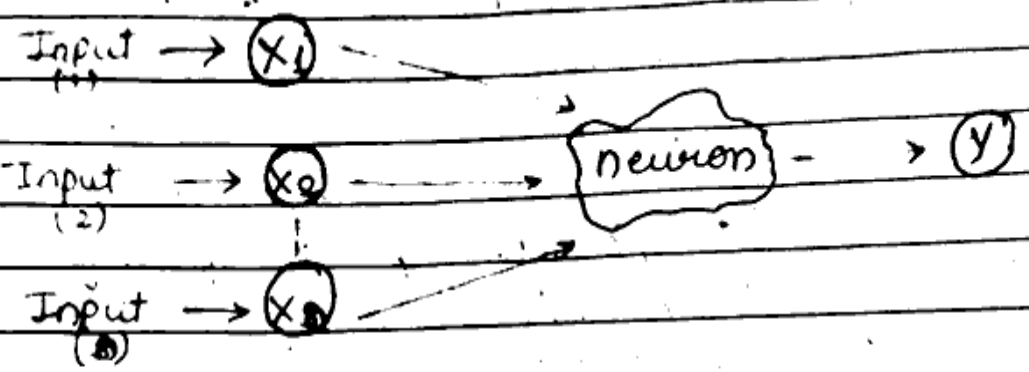
Whenever a ML model is trained with a huge amount of data, it starts capturing noise & inaccurate data into the training data set. It negatively affects the performance of the model.

→ Underfitting :-
It is just the opposite of overfitting. Whenever a ML model is trained with fewer amounts of data, & as a result, it provides incomplete & inaccurate data & destroys the accuracy of the ML model.

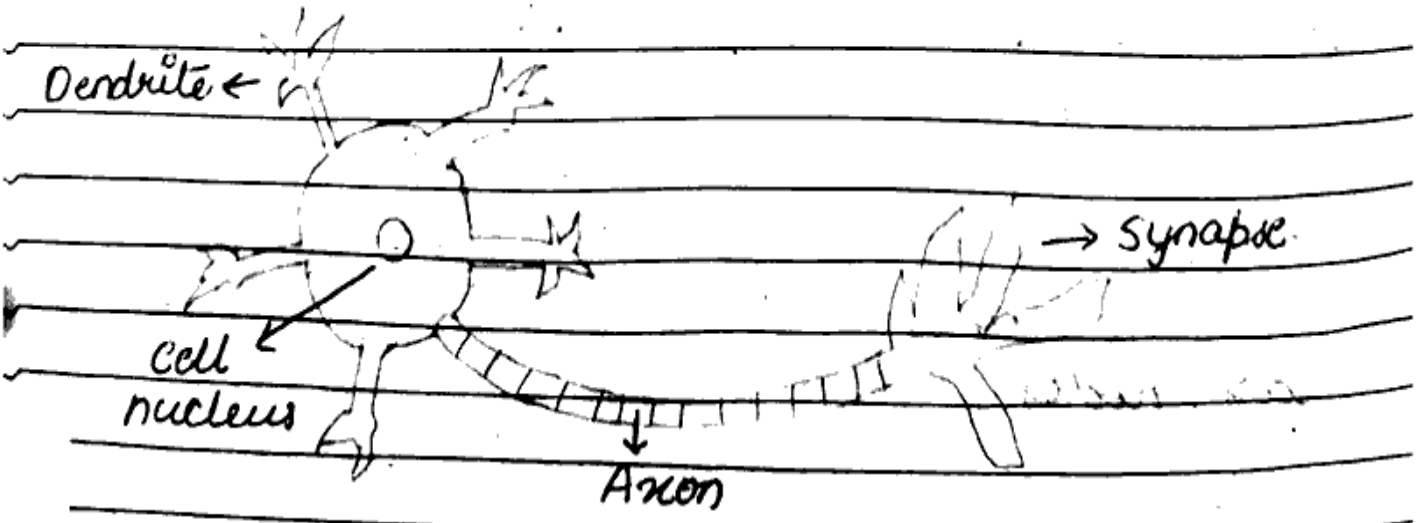
Q.2. a) Define artificial neural network. Explain the biological learning system.
The term ANN is derived from biological neural networks that develop the structure of human brain. Similar to the human brains that has neurons interconnected to one another, ANN also have neurons that are interconnected to one another in various layers of network. These neurons are known as nodes.

Date / / Page no.

The typical ANN looks something like the given figure.



Dendrite from biological neural network represent input in ANN, cell nucleus represent nodes, synapse represents weight & the axon represent output.
The fig illustrates the typical diagram of biological neural network.



b) Describe the characteristics of back propagation algo.

Backpropagation is one of the important concepts of neural network. Our task is to classify our data best.

Backpropagation algo. calculates the gradient of the error fn. Backpropagation can be written as a fn of the neural network. Backpropagation algo's are a set of methods used to efficiently train artificial neural networks following a gradient descent approach which exploits the chain rule.

The main features of backpropagation are the iterative, recursive & efficient method through which it calculates the updated weight to improve the network until it is not able to perform the task for which it is being trained. Derivatives of the activation fn to be known at network design time is required to backpropagation.

Date: / / Page no:

Q.2 a) Define decision tree learning list & explain appropriate problems for decision tree learning.

→ Decision tree is a tree where each node represents a feature (attribute). Each link (branch) represents a decision rule & each leaf represents an outcome (categorial or continue value).

→ A decision tree or a classification tree is which each internal node is labelled with input features. The arcs coming from a node labelled with a feature are labelled with each of possible.

• A decision tree has two kind nodes:-

→ Each leaf node has class label determined by majority note of training, Ex:-

→ Reading that leaf

→ Each internal node is a questions on features, it branches out acc to the answer.

→ Decision tree learning is method for approximating value discrete value target f^. The learn f^ represented by a decision tree.

→ It is a flow chart like tree structure where each node denotes test on the attributes values each branch represents an outcome of the test & tree leaves represents class or dan distribution.

b) list & explain the issues in decision tree learning.

→ Overfitting occurs when the algorithm captures noise in the data.

→ The model can get unstable due to small variation in data.

→ A highly complicated decision tree tends to have a low bias which makes it difficult for the model to work with new data.

Date: / / Page no:

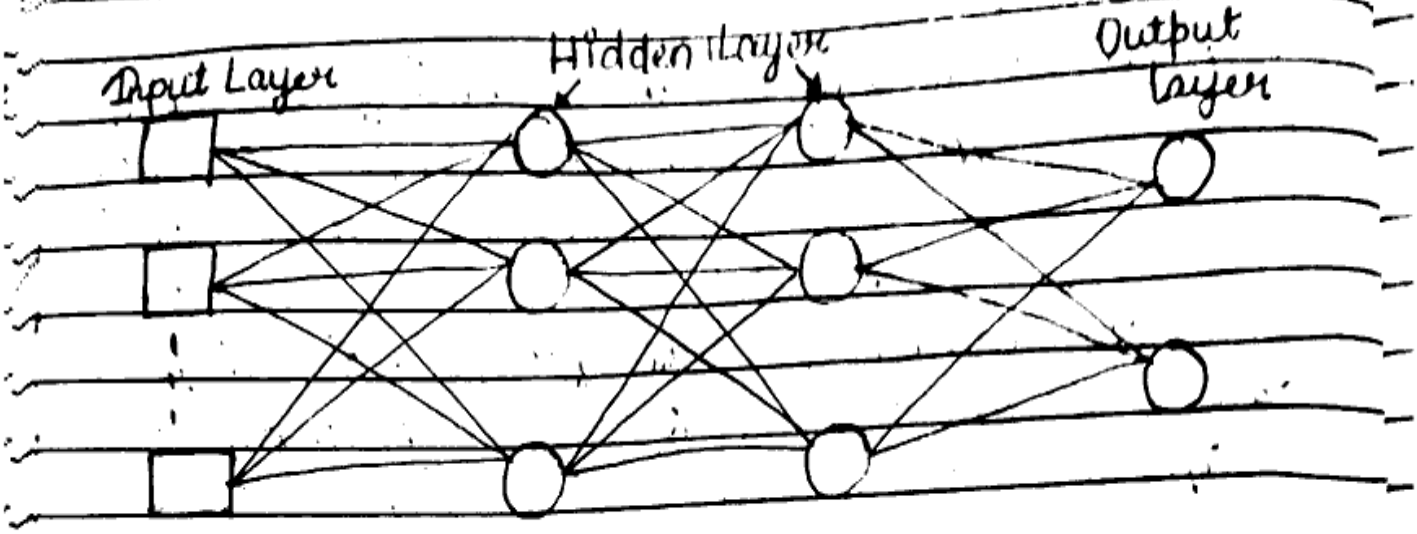
Q.4 a) Distinguish b/w supervised ML and Reinforcement learning.

Supervised	Reinforcement
→ Supervised learning deals with two main tasks Regression & classification.	→ Reinforcement learning deals with exploitation or exploration, Markov's decision processes, Policy learning, Deep learning & value learning
→ It works with the labelled data & here the output data patterns are known to the system.	Markov's decision process the agent interacts with the environment in discrete steps.
→ It is highly supervised.	It is less supervised which depends on the agent in determining the output.
→ The input data is it is labelled data.	→ The data is not predefined in it.

* It predicts based on a class type	* The learning agent works as a reward & action system.
* It maps labelled data to known output.	It follows a trial & error method.
* The goal is to generate formula based on input & output values.	* An agent learns through delayed fedback by interacting with the environment.

b) Explain the multi-layer perceptron model with a neat diagram.

A MLP is a fed forward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph b/w the input nodes connected as a directed graph b/w the input & output layers.



Input layer ⇒ It accepts input in several diff. format provided by the programmer.

Hidden layer ⇒ Present in b/w the I/P & O/P layer, it performs all the calculation to find hidden features & patterns.

Output layer ⇒ The I/P goes through a series of transformation using the hidden layer, which finally results in output i.e. convey using this layer. The artificial neural network takes input & computes the weighted sum of the I/P & include a bias.

Date / / Page no.

5. a) Explain parallel processing perception learning in neural networks.

When training neural networks, the primary ways to achieve this are model parallelism, which involves distributing the neural network across diff. processors, & data parallelism, which involves distributing training examples across diff. processors & computing updates to the neural network in parallel.

Parallel processing may also be referred to as parallel computing. Parallel processing allows individuals - as well as network & data center managers - to use ordinary desktop & laptop computers to solve complex problems that once required the assistance of a powerful supercomputer.

b) What is the goal of support vector machine (SVM)? How to compute the margin.

The goal of SVM is to divide the datasets into classes to find a max.

marginal hyperplane (MMH).

Support vectors → Datapoints that are closest to the hyperplane is c/d support vectors seperating line will be defined with the help of these data points.

Margin → It is the distance b/w the hyperplane & the observations closest to the hyperplane (support vectors). In SVM large margin is considered a good margin. There are two types of margins hard margin & soft margin.

Support vector machines is a supervised ML algo. which works both on classification & regression problems. It tries to classify data by finding a hyperplane that maximizes the margin b/w the classes in the training data. Hence, SVM is an example of a large margin classifier.