

Q 1) What is Machine Learning ? How does machine learning work ?

Ans :: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.



How Does Machine Learning Work?

Similar to how the human brain gains knowledge and understanding, machine learning relies on input, such as training data or knowledge graphs, to understand entities, domains and the connections between them. With entities defined, deep learning can begin.

The machine learning process begins with observations or data, such as examples, direct experience or instruction. It looks for patterns in data so it can later make inferences based on the examples provided. The primary aim of ML is to allow computers to learn autonomously without human intervention or assistance and adjust actions accordingly.

Q 2) Difference between supervised learning and unsupervised learning .

Ans:-Supervised learning is a type of machine learning that uses labeled data to train machine learning models. In labeled data, the output is already known. The model just needs to map the inputs to the respective outputs.

An example of supervised learning is to train a system that identifies the image of an animal. Attached below, you can see that we have our trained model that identifies the picture of a cat.

Working:

Supervised learning algorithms take labeled inputs and map them to the known outputs, which means you already know the target variable.

Now, let's focus on the training process for the supervised learning method.

Supervised Learning methods need external supervision to train machine learning models. Hence, the name supervised. They need guidance and additional information to return the desired result.

Applications:

Supervised learning algorithms are generally used for solving classification and regression problems.

Few of the top supervised learning applications are weather prediction, sales forecasting, stock price analysis.

1) Unsupervised Learning :-

Unsupervised learning is a type of machine learning that uses unlabeled data to train machines. Unlabeled data doesn't have a fixed output variable. The model learns from the data, discovers the patterns and features in the data, and returns the output.

Depicted below is an example of an unsupervised learning technique that uses the images of vehicles to classify if it's a bus or a truck. The model learns by identifying the parts of a vehicle, such as a length and width of the vehicle, the front, and rear end covers, roof hoods, the types of wheels used, etc. Based on these features, the model classifies if the vehicle is a bus or a truck.

Working:

Unsupervised learning finds patterns and understands the trends in the data to discover the output. So, the model tries to label the data based on the features of the input data.

The training process used in unsupervised learning techniques does not need any supervision to build models. They learn on their own and predict the output.

Applications:

Unsupervised learning is used for solving clustering and association problems.

One of the applications of unsupervised learning is customer segmentation. Based on customer behavior, likes, dislikes, and interests, you can segment and cluster similar customers into a group. Another example where unsupervised learning algorithms are used is used churn rate analysis.

Q No 3. Explain any three application of machine learning in brief.

Ans :- 1. Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**:

Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection and recognition algorithm**.

It is based on the Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

Competitive questions on Structures in HindiKeep Watching

2. Speech Recognition

While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.

Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning

algorithms are widely used by various applications of speech recognition. **Google assistant, Siri, Cortana, and Alexa** are using speech recognition technology to follow the voice instructions.

3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

- **Real Time location** of the vehicle from Google Map app and sensors
- **Average time has taken** on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations:

Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon, Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.

As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. Self-driving cars:

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:

- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

7. Virtual Personal Assistant:

We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.

These virtual assistants use machine learning algorithms as an important part.

These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

8. Online Fraud Detection:

Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction. So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction.

For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

9. Stock Market trading:

Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's **long short term memory neural network** is used for the prediction of stock market trends.

10. Medical Diagnosis:

In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.

It helps in finding brain tumors and other brain-related diseases easily.

11. Automatic Language Translation:

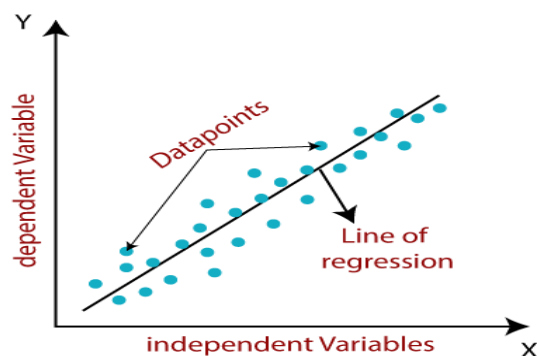
Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.

Q No 4) Explain Linear Regression model in brief with relationship between dependent and independent variable.

Ans:- Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price**, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



Mathematically, we can represent a linear regression as:

$$y = a_0 + a_1x + \varepsilon$$

Here,

Y=	Dependent	Variable	(Target	Variable)
X=	Independent	Variable	(predictor	Variable)
a0=	intercept	of the line	(Gives an additional degree of freedom)	
a1	=	Linear regression coefficient	(scale factor to each input value).	
ε	=	random error		

The values for x and y variables are training datasets for Linear Regression model representation.

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:

1) Simple Linear Regression:

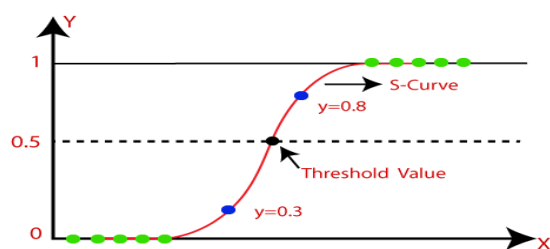
If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

2) Multiple Linear regression:

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

Q no 5) Explain logistic regression model in brief with sigmoid function.

- **Ans:- Logistic regression** is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1.**
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems.**
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.
The below image is showing the logistic function:



Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Q no 6) How does decision tree algorithm work with example.

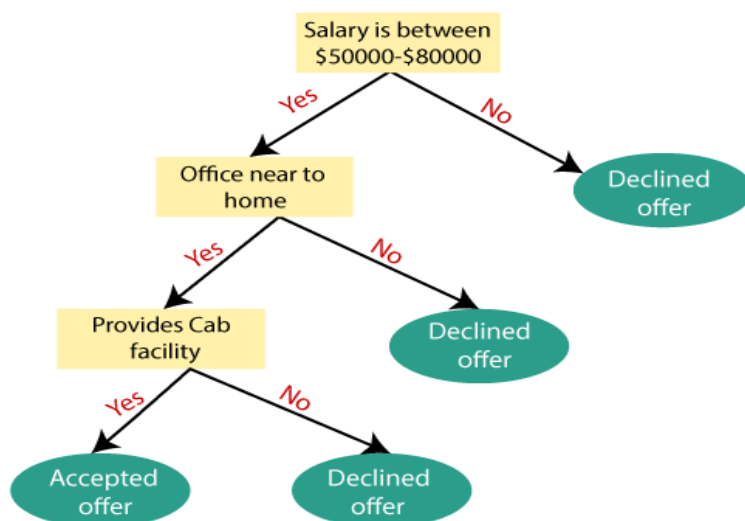
In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

Algorithm Steps

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM)**.
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Example: Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not. So, to solve this problem, the decision tree starts with the root node (Salary attribute by ASM). The root node splits further into the next decision node (distance from the office) and one leaf node based on the corresponding labels. The next decision node further gets split into one decision node (Cab facility) and one leaf node. Finally, the decision node splits into two leaf nodes (Accepted offers and Declined offer). Consider the below diagram:



Q no 7) Write short notes on following Attribute selection measure terms:-

- a) **Information gain** b) **gini index**

Ans:-1. Information Gain:

- Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
- It calculates how much information a feature provides us about a class.
- According to the value of information gain, we split the node and build the decision tree.
- A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using the below formula:

Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)]

Entropy: Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:

Entropy(s)= $-P(\text{yes})\log_2 P(\text{yes}) - P(\text{no})\log_2 P(\text{no})$

Where,

- **S= Total number of samples**
- **P(yes)= probability of yes**
- **P(no)= probability of no**

2. Gini Index:

- Gini index is a measure of impurity or purity used while creating a decision tree in the CART(Classification and Regression Tree) algorithm.
- An attribute with the low Gini index should be preferred as compared to the high Gini index.
- It only creates binary splits, and the CART algorithm uses the Gini index to create binary splits.
- Gini index can be calculated using the below formula:

Gini Index= $1 - \sum_j P_j^2$

Q no 8) Explain rule based classifier with example.

Ans:-> Rule-based classifiers are just another type of classifier which makes the class decision depending by using various “if..else” rules. These rules are easily interpretable and thus these classifiers are generally used to generate descriptive models. The condition used with “if” is called the **antecedent** and the predicted class of each rule is called the **consequent**.

IF-THEN Rules

Rule-based classifier makes use of a set of IF-THEN rules for classification. We can express a rule in the following form –

IF condition THEN conclusion

Let us consider a rule R1,

R1: IF age = youth AND student = yes
THEN buy_computer = yes

Points to remember –

- The IF part of the rule is called **rule antecedent** or **precondition**.
- The THEN part of the rule is called **rule consequent**.
- The antecedent part the condition consist of one or more attribute tests and these tests are logically ANDed.
- The consequent part consists of class prediction.

Note – We can also write rule R1 as follows –

R1: (age = youth) ^ (student = yes)(buys computer = yes)

If the condition holds true for a given tuple, then the antecedent is satisfied.

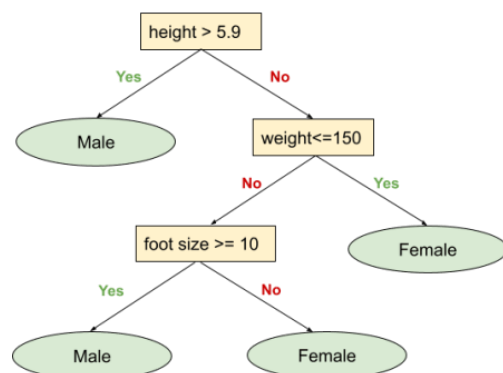
Rule Extraction

Here we will learn how to build a rule-based classifier by extracting IF-THEN rules from a decision tree.

Example

Let's consider a simple data set to find gender based on height, weight, and foot size.

Apply the **IF-THEN** rule to split the data and gender prediction.



According to the above diagram, if height is greater than 5.9 ft or if height is less than or equal to 5.9 ft and weight is greater than 150 lbs, and foot size is greater than or equal to 10 inches, then Gender is classified as '**male.**' And if height is less than 5.9 ft and weight is less than or equal to 150 lbs or if height is less than 5.9 ft and foot size is less than 10 inches then Gender is classified as '**female.**'

Assessment of Rule

Let's assess the rule 'height is greater than 5.9 ft' and say it R1

Here, the total number of records, i.e., $n = 8$

Number of records covered by rule R1 is $n_a = 3$

Number of records correctly classified by rule R1 is $n_c = 3$

coverage(R) = $n_a / |n| = 3/8 = 37.5\%$

accuracy(R) = $n_c / n_a = 3/3 = 100\%$

Here, coverage of rule R1 is **37.5%**, and accuracy is **100%**.

Q no 9) Explain Bayesian Belief network in brief .

Bayesian Belief Network is a graphical representation of different probabilistic relationships among random variables in a particular set. It is a classifier with no dependency on attributes i.e it is condition independent. Due to its feature of joint probability, the probability in Bayesian Belief Network is derived, based on a condition — $P(\text{attribute}/\text{parent})$ i.e probability of an attribute, true over parent attribute

Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

"A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."

It is also called a **Bayes network**, **belief network**, **decision network**, or **Bayesian model**.

Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.

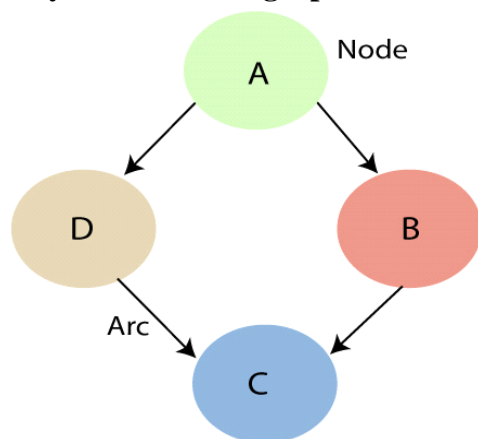
Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. It can also be used in various tasks including **prediction**, **anomaly detection**, **diagnostics**, **automated insight**, **reasoning**, **time series prediction**, and **decision making under uncertainty**.

Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

- **Directed Acyclic Graph**
- **Table of conditional probabilities.**

The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

A Bayesian network graph is made up of nodes and Arcs (directed links), where:



- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.
- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph.
- These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
- In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.

- If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
- Node C is independent of node A.

Q no 10) Explain hidden markov model(HMM) with example.

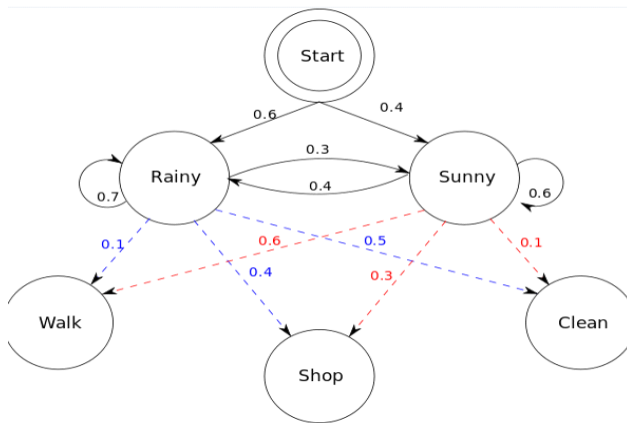
The hidden Markov model (HMM) is another type of Markov model where there are few states which are hidden. This is where HMM differs from a Markov chain. HMM is a statistical model in which the system being modeled are Markov processes with unobserved or hidden states. It is a hidden variable model which can give an observation of another hidden state with the help of the Markov assumption. The hidden state is the term given to the next possible variable which cannot be directly observed but can be inferred by observing one or more states according to Markov's assumption. **Markov assumption is the assumption that a hidden variable is dependent only on the previous hidden state.** Mathematically, the probability of being in a state at a time t depends only on the state at the time $(t-1)$. It is termed a **limited horizon assumption**. Another Markov assumption states that the **conditional distribution over the next state, given the current state, doesn't change over time**. This is also termed a **stationary process assumption**.

A Markov model is made up of two components: the state transition and hidden random variables that are conditioned on each other. However, A hidden Markov model consists of five important components:

- **Initial probability distribution:** An initial probability distribution over states, π_i is the probability that the Markov chain will start in state i . Some states j may have $\pi_j = 0$, meaning that they cannot be initial states. The initialization distribution defines each hidden variable in its initial condition at time $t=0$ (the initial hidden state).
- **One or more hidden states**
- **Transition probability distribution:** A transition probability matrix where each a_{ij} represents the probability of moving from state i to state j . The transition matrix is used to show the hidden state to hidden state transition probabilities.
- **A sequence of observations**
- **Emission probabilities:** A sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation o_t being generated from a state I . The emission probability is used to define the hidden variable in terms of its next hidden state. It represents the conditional distribution over an observable output for each hidden state at time $t=0$.

Example:-

Let's understand the above using the hidden Markov model representation shown below:



The hidden Markov model in the above diagram represents the process of predicting whether someone will be found to be walking, shopping, or cleaning on a particular day depending upon whether the day is rainy or sunny. The following represents five components of the hidden

Markov model in the above diagram:

```
states = ('Rainy', 'Sunny')

observations = ('walk', 'shop', 'clean')

start_probability = {'Rainy': 0.6, 'Sunny': 0.4}

transition_probability = {
    'Rainy' : {'Rainy': 0.7, 'Sunny': 0.3},
    'Sunny' : {'Rainy': 0.4, 'Sunny': 0.6},
}

emission_probability = {
    'Rainy' : {'walk': 0.1, 'shop': 0.4, 'clean': 0.5},
    'Sunny' : {'walk': 0.6, 'shop': 0.3, 'clean': 0.1},
}
```

Q no 11) What is dimensionality Reduction? Explain Difference between PCA & ICA.

Dimensionality reduction is the process of reducing the number of features (or dimensions) in a dataset while retaining as much information as possible. This can be done for a variety of reasons, such as to reduce the complexity of a model, to improve the performance of a learning algorithm, or to make it easier to visualize the data. There are several techniques for dimensionality reduction, including principal component analysis (PCA), singular value decomposition (SVD), and linear discriminant analysis (LDA). Each technique uses a different method to project the data onto a lower-dimensional space while preserving important information.

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done. These factors are basically variables called features. The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play. Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

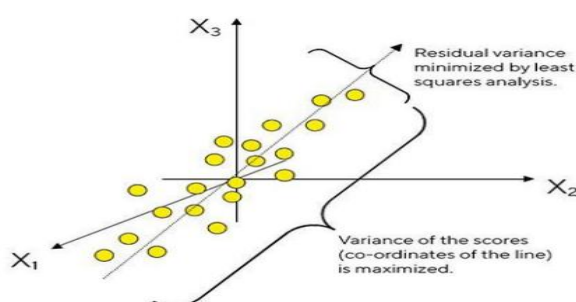
Difference between PCA and ICA are as follows:

Principal Component Analysis	Independent Component Analysis
It reduces the dimensions to avoid the problem of overfitting.	It decomposes the mixed signal into its independent sources' signals.
It deals with the Principal Components.	It deals with the Independent Components.
It focuses on maximizing the variance.	It doesn't focus on the issue of variance among the data points.
It focuses on the mutual orthogonality property of the principal components.	It doesn't focus on the mutual orthogonality of the components.
It doesn't focus on the mutual independence of the components.	It focuses on the mutual independence of the components.

Q no 12. Explain concept of Principal Component Analysis & Write the algorithm of PCA.

Principal component analysis, or PCA, is a statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of “summary indices” that can be more easily visualized and analyzed. Principal component analysis today is one of the most popular multivariate statistical techniques. It has been widely used in the areas of pattern recognition and signal processing and is a statistical method under the broad title of factor analysis. PCA forms the basis of multivariate data analysis based on projection methods. The most important use of PCA is to represent a multivariate data table as smaller set of variables (summary indices) in order to observe trends, jumps, clusters and outliers. This overview may uncover the relationships between observations and variables, and among the variables. PCA is a very flexible tool and allows analysis of datasets that may contain, for example, multicollinearity, missing values, categorical data, and imprecise measurements. The goal is to extract the important information from the data and to express this information as a set of summary indices called principal components.

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.



The Principal Components Analysis Algorithm

- Write N datapoints $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{Mi})$ as row vectors
 - Put these vectors into a matrix \mathbf{X} (which will have size $N \times M$)
 - Centre the data by subtracting off the mean of each column, putting it into matrix \mathbf{B}
 - Compute the covariance matrix $\mathbf{C} = \frac{1}{N} \mathbf{B}^T \mathbf{B}$
 - Compute the eigenvalues and eigenvectors of \mathbf{C} , so $\mathbf{V}^{-1} \mathbf{C} \mathbf{V} = \mathbf{D}$, where \mathbf{V} holds the eigenvectors of \mathbf{C} and \mathbf{D} is the $M \times M$ diagonal eigenvalue matrix
 - Sort the columns of \mathbf{D} into order of decreasing eigenvalues, and apply the same order to the columns of \mathbf{V}
 - Reject those with eigenvalue less than some η , leaving L dimensions in the data
-

Q. No. 13. What is K Means Clustering algorithm in unsupervised learning model.

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

What is K-Means Algorithm?

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if $K=2$, there will be two clusters, and for $K=3$, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

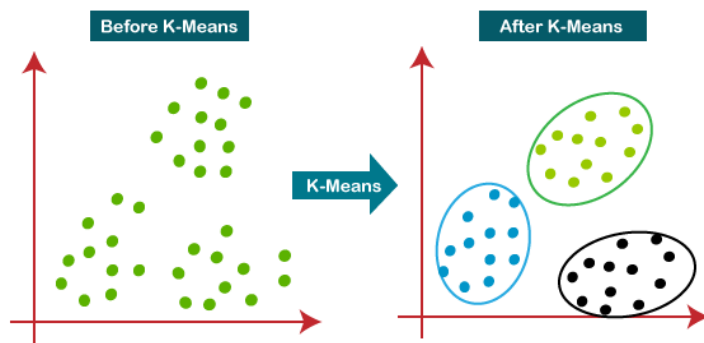
The algorithm takes the unlabeled dataset as input, divides the dataset into k -number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k -center. Those data points which are near to the particular k -center, create a cluster.

Hence each cluster has datapoints with some commonalities, and it is away from other clusters.

The below diagram explains the working of the K-means Clustering Algorithm:



How does the K-Means Algorithm Work?

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.

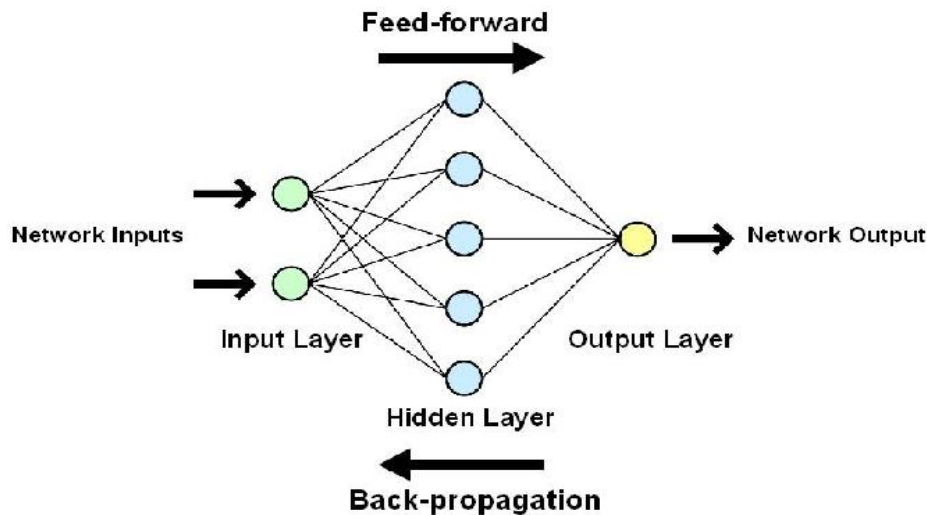
Q no 14. What is Artificial neural network ? Explain Architecture of artificial neural network.

Artificial Neural Networks (ANN) are algorithms based on brain function and are used to model complicated patterns and forecast issues. The Artificial Neural Network (ANN) is a deep learning method that arose from the concept of the human brain Biological Neural Networks. The development of ANN was the result of an attempt to replicate the workings of the human brain. The workings of ANN are extremely similar to those of biological neural networks, although they are not identical. ANN algorithm accepts only numeric and structured data.

Convolutional Neural Networks (CNN) and Recursive Neural Networks (RNN) are used to accept unstructured and non-numeric data forms such as Image, Text, and Speech. This article focuses solely on Artificial Neural Networks.

Artificial Neural Networks Architecture (neural network representation) :

1. There are three layers in the network architecture: the input layer, the hidden layer (more than one), and the output layer. Because of the numerous layers are sometimes referred to as the MLP (Multi-Layer Perceptron).



Input Layer:

As the name suggests, it accepts inputs in several different formats provided by the programmer.

Hidden Layer:

The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer:

The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

2. It is possible to think of the hidden layer as a “distillation layer,” which extracts some of the most relevant patterns from the inputs and sends them on to the next layer for further analysis. It accelerates and improves the efficiency of the network by recognizing just the most important information from the inputs and discarding the redundant information.
3. The activation function is important for two reasons: first, it allows you to turn on your computer.

Q No. 15. Explain Genetic algorithm model and discuss how it different from traditional algortihm.

The genetic algorithm models the genetic process that gives rise to evolution. In particular, it models sexual reproduction, where both parents give some genetic information to their offspring.

The Genetic Algorithm is a computational approximation to how evolution performs search, which is by producing modifications of the parent genomes in their offspring and thus producing new individuals with different fitness.

For this we need following framework-

- a method for representing problems as chromosomes
- a way to calculate the fitness of a solution
- a selection method to choose parents
- a way to generate offspring by breeding the parents

Difference between Genetic Algorithms and Traditional Algorithms

- A search space is the set of all possible solutions to the problem. In the traditional algorithm, only one set of solutions is maintained, whereas, in a genetic algorithm, several sets of solutions in search space can be used.
- Traditional algorithms need more information in order to perform a search, whereas genetic algorithms need only one objective function to calculate the fitness of an individual.
- Traditional Algorithms cannot work parallelly, whereas genetic Algorithms can work parallelly (calculating the fitness of the individualities are independent).
- One big difference in genetic Algorithms is that rather of operating directly on seeker results, inheritable algorithms operate on their representations (or rendering), frequently appertained to as chromosomes.
- One of the big differences between traditional algorithm and genetic algorithm is that it does not directly operate on candidate solutions.
- Traditional Algorithms can only generate one result in the end, whereas Genetic Algorithms can generate multiple optimal results from different generations.
- The traditional algorithm is not more likely to generate optimal results, whereas Genetic algorithms do not guarantee to generate optimal global results, but also there is a great possibility of getting the optimal result for a problem as it uses genetic operators such as Crossover and Mutation.
- Traditional algorithms are deterministic in nature, whereas Genetic algorithms are probabilistic and stochastic in nature.

Q No. 16. What are three different way to do dimensionality reduction.

->**The first is feature selection**, which typically means looking through the features that are available and seeing whether or not they are actually useful, i.e., correlated to the output variables.

->**The second method is feature derivation**, which means deriving new features from the old ones, generally by applying transforms to the dataset that simply change the axes (coordinate system) of the graph by moving and rotating them, which can be written simply as a matrix that we apply to the data.

->**The third method is simply to use clustering in order to group** together similar datapoints, and to see whether this allows fewer features to be used

Q No. 17. Explain Knapsack problem as the application of genetic algorithm.

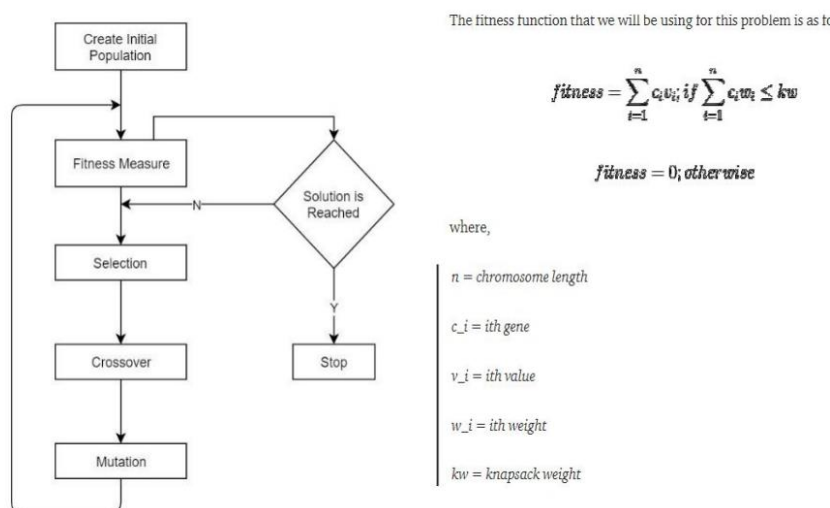
For the knapsack problem the alphabet is very simple, since we can make it binary, since for each item we just need to say whether or not we want to take it. We make the string L units long, where L is the total number of things we would like to take with us, and make each unit a binary digit. We then encode a solution using 0 for the things we will not take and 1 for the things we will. For example, If we have four things to add in a knapsack, then (0, 1, 1, 0) would mean that we take the middle two, but not the first or last. Note that this does not tell us whether or not this string is possible (that is, whether the things that we have said we will take will actually fit into the knapsack), nor whether it is a good string (whether it fills the knapsack). To work these out we need some way to decide how well each string fulfills the problem criteria. This is known as the fitness of the string.

Fitness function :-

The fitness function can be seen as an oracle that takes a string as an argument and returns a value for that string. Together with the string encoding the fitness function forms the problem-specific part of the GA. It is worth thinking about what we want from our fitness function. Clearly, the best string should have the highest fitness, and the fitness should decrease as the strings do less well on the problem. In real evolution, the fitness landscape is not static: there is competition between different species, such as predators and prey, or medical cures for certain diseases, and so the measure of fitness changes over time. We'll ignore that in the genetic algorithm.

Fitness for knapsack problem For the knapsack problem, we decided that we wanted to make the bag as full as possible. So we would need to know the volume of each item that we want to put into the knapsack, and then for a given string that says which things should be taken, and which should not, we can compute the total volume. This is then a possible fitness function. However, it does not tell us anything about whether they will fit into the bag—with this fitness function the optimal solution is to take everything. So we need to check that they will fit, and if they will not, reduce the fitness of that solution.

GENETIC ALGORITHMS • The problem addressed by GAs is to search a space.

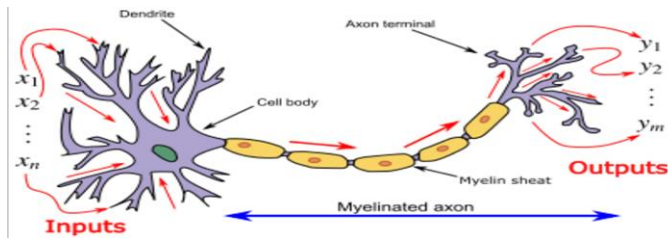


Q 18. What is artificial neural network? Explain how concept is resembled with biological neuron?

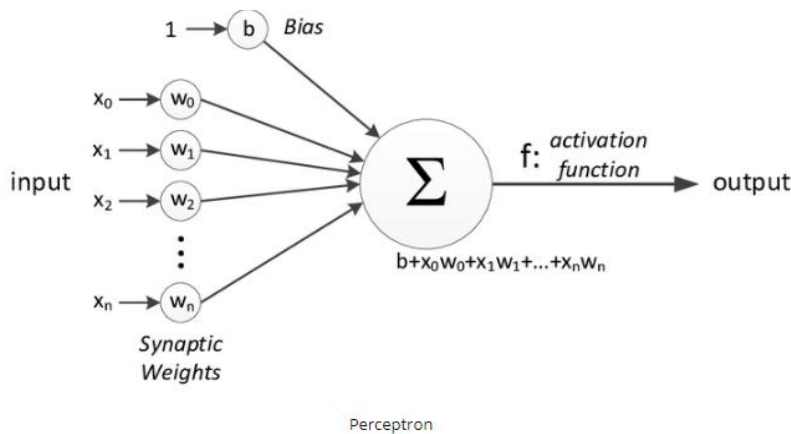
Ans:- Artificial Neural Networks or ANN is an information processing paradigm that is inspired by the way the biological nervous system such as brain process information.

It is composed of large number of highly interconnected processing elements(neurons) working in unison to solve a specific problem.

Biological Neurons (also called nerve cells) or simply neurons are the fundamental units of the brain and nervous system, the cells responsible for receiving sensory input from the external world via dendrites, process it and gives the output through Axons



- The following diagram represents the general model of ANN which is inspired by a biological neuron. It is also called Perceptron.
- A single layer neural network is called a Perceptron. It gives a single output.



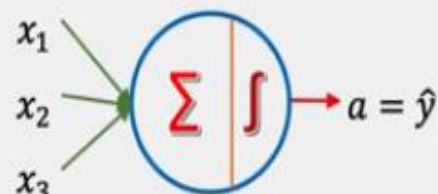
Artificial Neuron:-

The arrangements and connections of the neurons made up the network and have three layers. The first layer is called the input layer and is the only layer exposed to external signals. The input layer transmits signals to the neurons in the next layer, which is called a hidden layer. The hidden layer extracts relevant features or patterns from the received signals. Those features or patterns that are considered important are then directed to the output layer, which is the final layer of the network.

Q no. 19. Explain Concept of Activation function with respect to neural network? What are different types of Activation function ?

Activation Functions

- Mathematical equations that determine the output of a neural network.
- Attached to each neuron in the network, and determines whether it should be activated ("fired") or not, based on whether each neuron's input is relevant for the model's prediction.



- Help normalize the output of each neuron to a range between (0 – 1) or (-1 – 1).
- Must be computationally efficient as they are calculated across thousands or even millions of neurons for each data sample.

Activation Function ?

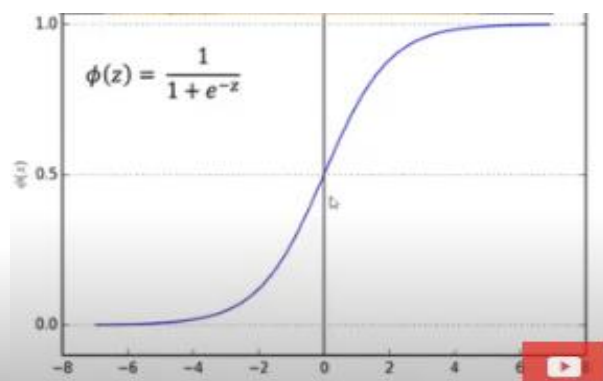
- Activation function are an extremely important feature of the artificial neuron network. They basically decide whether a neuron should be activated or not.
- It limit the output signal to a finite value.
- $\text{Logit} = (\text{Input} * \text{Weight}) + \text{Bias}$
- Bias is the information which can impact output without being dependent on the any feature.
- It does the non-linear transformation to the input to the input making it capable to learn and perform more complex relationship.

Types of Activation Function

1) Sigmoid:-

Sigmoid

- The main reason why we use sigmoid function is because it exists between **(0 to 1)**.
- Therefore, it is especially used for models where we have to **predict the probability** as output. Since probability of anything exists only between the range of **0 and 1**, sigmoid is the right choice

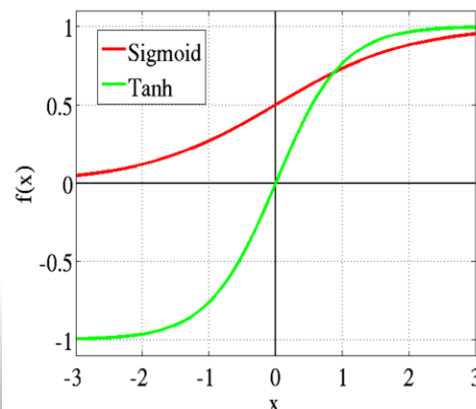


The sigmoid function is a more generalized logistic activation function which is used for multiclass classification.

2) Tanh:-

TanH

- tanh is also like logistic sigmoid but in better way. The range of the tanh function is from **(-1 to 1)**. tanh is also sigmoidal (s - shaped).
- TanH is often preferred over the sigmoid neuron because it is zero centered.
- The advantage is that the negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the tanh graph.



The tanh function is mainly used classification between two classes.

Both tanh and logistic sigmoid activation functions are used in feed-forward nets.