Q:1When we have a lot of data, it can be difficult to decide which one is relevant and which is not. We need to have a way of telling them apart. This is where Machine Learning comes in. Classification, regression, clustering etc. are the fundamental tasks in machine learning Machine Learning has the ability to learn from data and make decisions based on that. With the help of Machine Learning, we are able to train our models in order for them to know what is relevant and what is not so that they can take care of this task for us.

1. Image Recognition:

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:

4. Product recommendations:

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:

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.

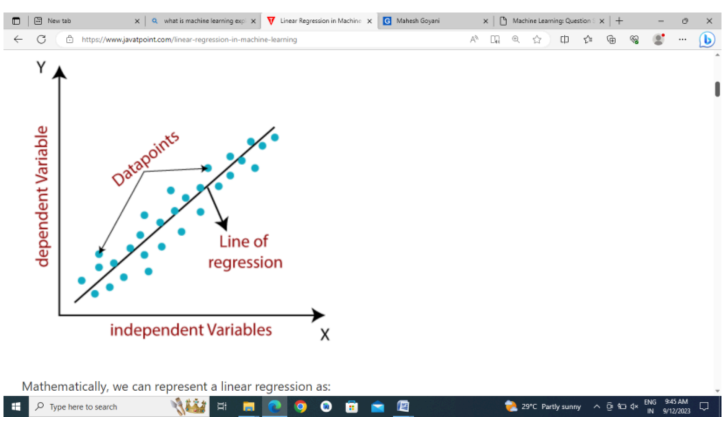
11. Automatic Language Translation:

Q:2 explain liner regression with suitable example.

Linear regression is a common technique used in statistics. Linear regression is used to find a linear relationship between two variables, one of which (the dependent variable) varies with the other (the independent variable).

The dependent variable can be a continuous or discrete variable that changes values continuously, for example height or weight. It can also be a categorical variable with two or more levels, for example blood type.

The independent variable can be any quantity that we are interested in guessing from the data we have collected on the dependent variable. For example, in an analysis of heights and weights, we would use height as the independent and weight as the dependent variables. In an analysis of blood types and eye colors, eye color would be the independent and blood type would be the dependent variables.



Types of Linear Regression

- **SimpleLinearRegression:**
  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.

- **MultipleLinearregression:**
  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:3 what is classification explain SVM

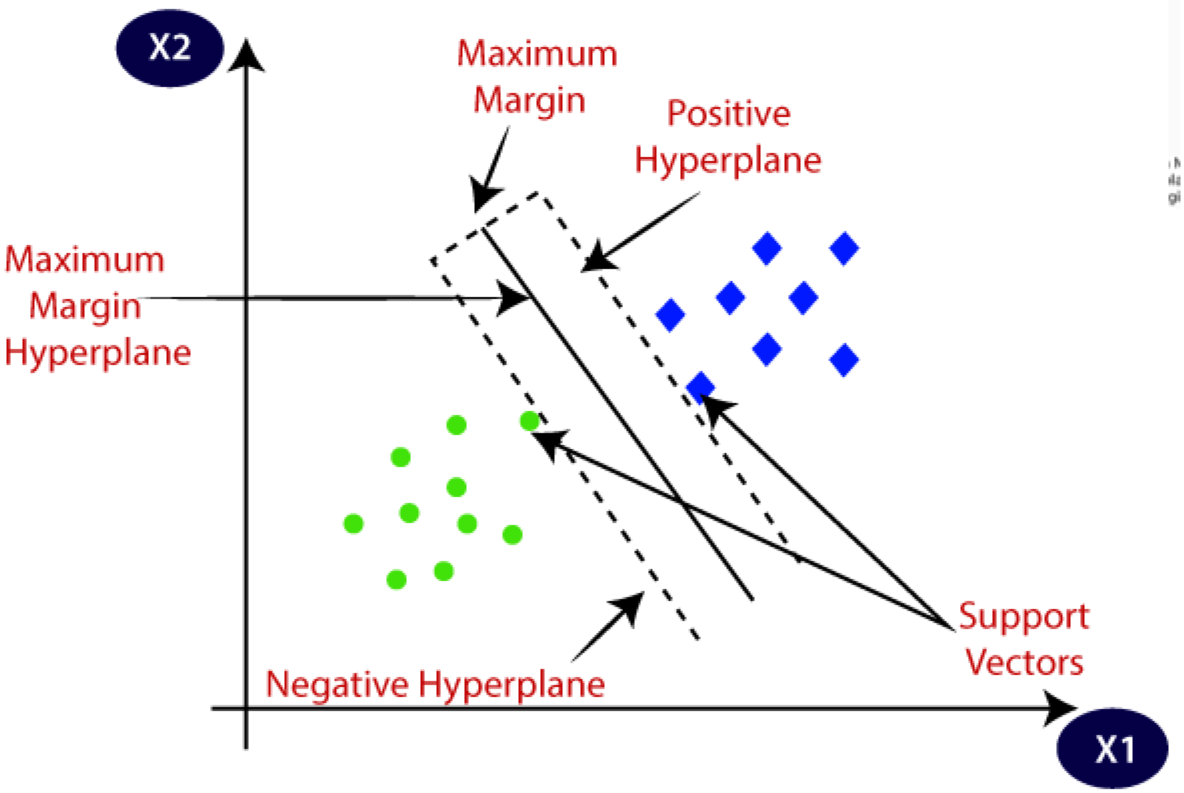
the Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog,** etc. Classes can be called as targets/labels or categories.

Support Vector Machine Algorithm

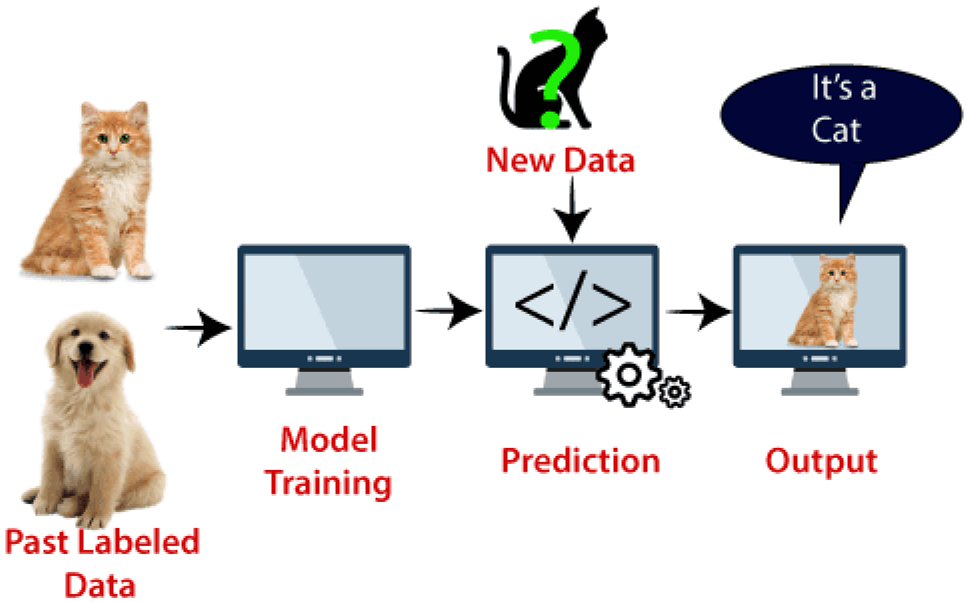
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

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

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:



SVM algorithm can be used for **Face detection, image classification, text categorization,** etc.

Types of SVM

- **Linear SVM:**
- **Non-linear SVM:**

Advantage:

- SVM classifiers perform well in high-dimensional space and have excellent accuracy. SVM classifiers require less memory because they only use a portion of the training data.
- SVM performs reasonably well when there is a large gap between classes.
- High-dimensional spaces are better suited for SVM.
- When the number of dimensions exceeds the number of samples, SVM is useful.
- SVM uses memory effectively.

Disadvantage:

- SVM requires a long training period; as a result, it is not practical for large datasets.

- The inability of SVM classifiers to handle overlapping classes is another drawback.
- Large data sets are not a good fit for the SVM algorithm.
- When the data set contains more noise, such as overlapping target classes, SVM does not perform as well.
- The SVM will perform poorly when the number of features for each data point is greater than the number of training data samples.

Q:4 bagging and boosting

Bagging	Boosting
The most effective manner of mixing predictions that belong to the same type.	A manner of mixing predictions that belong to different sorts.
The main task of it is decrease the variance but not bias.	The main task of it is decrease the bias but not variance.
Here each of the model is different weight.	Here each of the model is same weight.
Each of the model is built here independently.	Each of the model is built here dependently.
This training records subsets are decided on using row sampling with alternative and random sampling techniques from the whole training dataset.	Each new subset consists of the factors that were misclassified through preceding models.
It is trying to solve by over fitting problem.	It is trying to solve by reducing the bias.
If the classifier is volatile (excessive variance), then apply bagging.	If the classifier is stable and easy (excessive bias) the practice boosting.
In the bagging base, the classifier is works parallelly.	In the boosting base, the classifier is works sequentially.
Example is random forest model by using bagging.	Example is AdaBoost using the boosting technique.

Q:5Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- It is mainly used in *text classification* that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

Why is it called Naïve Bayes?

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

- **Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes**: It is called Bayes because it depends on the principle of Bayes' Theorem.

Bayes' Theorem:

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

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

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Problem: If the weather is sunny, then the Player should play or not?

Solution: To solve this, first consider the below dataset:

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No

12	Overcast	Yes
13	Overcast	Yes

Frequency table for the Weather Conditions:

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	5

Likelihood table weather condition:

Weather	No	Yes	
Overcast	0	5	5/14= 0.35
Rainy	2	2	4/14=0.29
Sunny	2	3	5/14=0.35
All	4/14=0.29	10/14=0.71	

Applying Bayes'theorem:

P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)

P(Sunny|Yes)= 3/10= 0.3

P(Sunny)= 0.35

P(Yes)=0.71

So P(Yes|Sunny) = 0.3*0.71/0.35= **0.60**

P(No|Sunny)= P(Sunny|No)*P(No)/P(Sunny)

P(Sunny|NO)= 2/4=0.5

P(No)= 0.29

P(Sunny)= 0.35

So P(No|Sunny)= 0.5*0.29/0.35 = **0.41**

So as we can see from the above calculation that **P(Yes|Sunny)>P(No|Sunny)**

Hence on a Sunny day, Player can play the game.

Advantages of Naïve Bayes Classifier:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems**.

Disadvantages of Naïve Bayes Classifier:

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

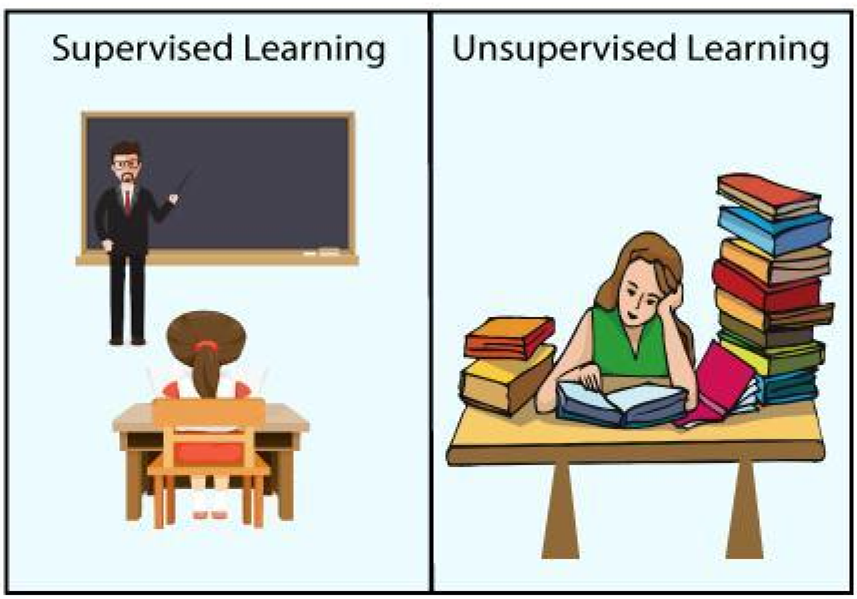
Applications of Naïve Bayes Classifier:

- It is used for **Credit Scoring**.
- It is used in **medical data classification**.
- It can be used in **real-time predictions** because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as **Spam filtering** and **Sentiment analysis**

Q:18

Difference between Supervised and Unsupervised Learning

Supervised and Unsupervised learning are the two techniques of machine learning. But both the techniques are used in different scenarios and with different datasets. Below the explanation of both learning methods along with their difference table is given.



Supervised Learning	Unsupervised Learning
Supervised learning algorithms are trained using labeled data.	Unsupervised learning algorithms are trained using unlabeled data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
Supervised learning model predicts the output.	Unsupervised learning model finds the hidden patterns in data.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Supervised learning can be categorized in **Classification** and **Regression** problems.	Unsupervised Learning can be classified in **Clustering** and **Associations** problems.

Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where we have only input data and no corresponding output data.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.
Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	The machine learns by using labelled data	The machine is trained on labelled data without any guidance	An agent interacts with its environment by producing actions & discovers errors or rewards
Types of Problems	Regression or Classification	Association or Classification	Reward Based
Types of Data	Labelled Data	Unlabelled Data	No pre-defined data
Training	External Supervision	No Supervision	No Supervision
Approach	Map Labelled input to known output	Understand pattern and discover output	Follow trail and error method
Popular Algorithms	Linear regression, Logistic regression, SVM, KNN, etc	K-means, C-means, etc	Q-Learning, SARSA, etc

Q:8)K-Means Clustering Algorithm

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?

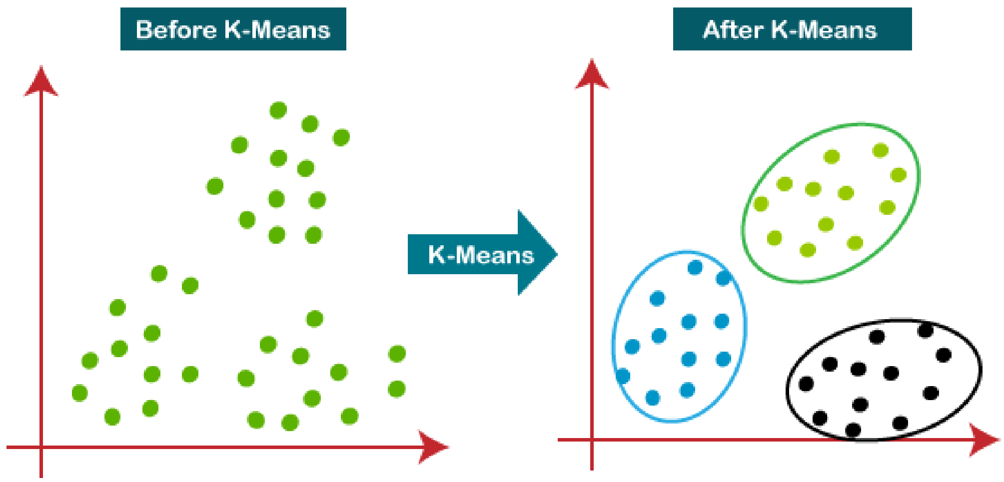
K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

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.

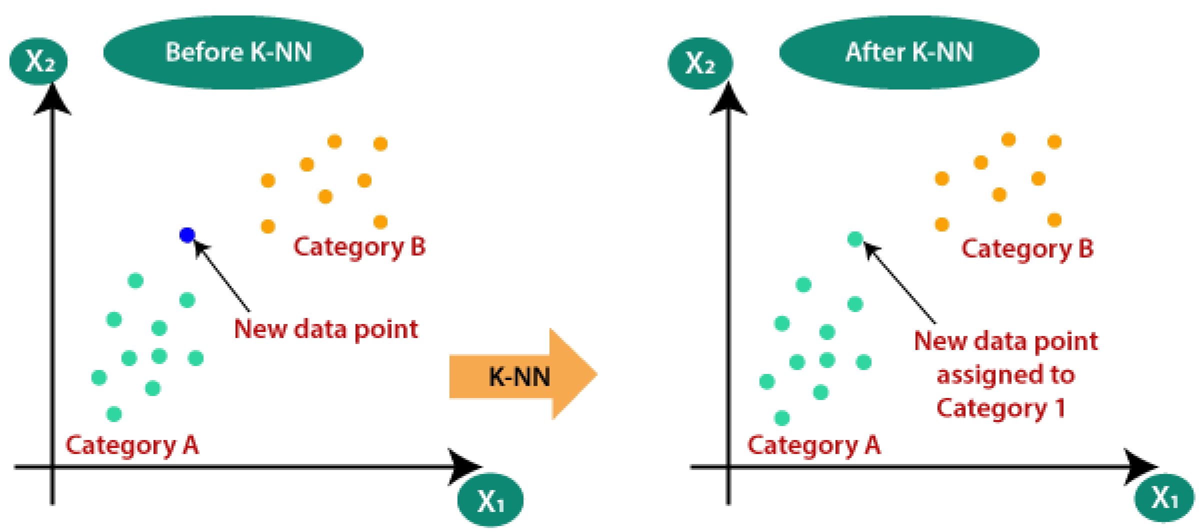
9)K-Nearest Neighbor(KNN) Algorithm for Machine Learning

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

- **Example:** Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

Why do we need a K-NN Algorithm?

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:



How does K-NN work?

The K-NN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.

Q:10. Reinforcement Learning

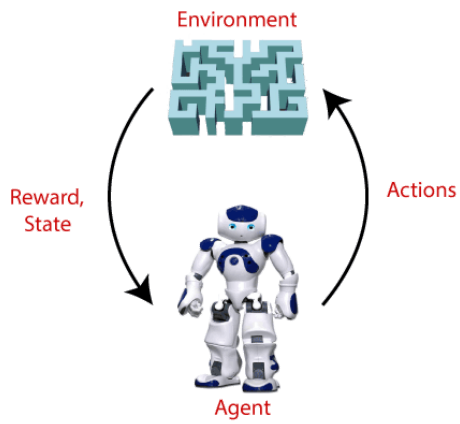
Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent gets rewarded for each good action and get punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards.

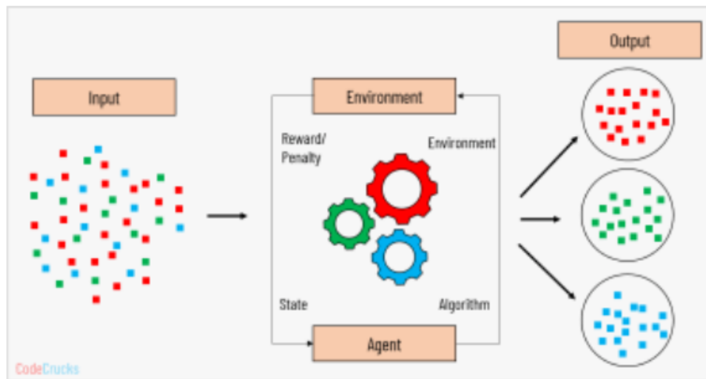
In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only.

The reinforcement learning process is similar to a human being; for example, a child learns various things by experiences in his day-to-day life. An example of reinforcement learning is to play a game, where the Game is the environment, moves of an agent at each step define states, and the goal of the agent is to get a high score. Agent receives feedback in terms of punishment and rewards.

Due to its way of working, reinforcement learning is employed in different fields such as **Game theory, Operation Research, Information theory, multi-agent systems.**

A reinforcement learning problem can be formalized using **Markov Decision Process(MDP).** In MDP, the agent constantly interacts with the environment and performs actions; at each action, the environment responds and generates a new state.





Categories of Reinforcement Learning

Reinforcement learning is categorized mainly into two types of methods/algorithms:

- **Positive Reinforcement Learning:** Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.
- **Negative Reinforcement Learning:** Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

Real-world Use cases of Reinforcement Learning

- **VideoGames:**
  RL algorithms are much popular in gaming applications. It is used to gain super-human performance. Some popular games that use RL algorithms are **AlphaGO** and **AlphaGO Zero**.
- **ResourceManagement:**
  The "Resource Management with Deep Reinforcement Learning" paper showed that how to use RL in computer to automatically learn and schedule resources to wait for different jobs in order to minimize average job slowdown.
- **Robotics:**
  RL is widely being used in Robotics applications. Robots are used in the industrial and manufacturing area, and these robots are made more powerful with reinforcement learning. There are different industries that have their vision of building intelligent robots using AI and Machine learning technology.

- **TextMining**
  Text-mining, one of the great applications of NLP, is now being implemented with the help of Reinforcement Learning by Salesforce company.

Advantages and Disadvantages of Reinforcement Learning

Advantages

- It helps in solving complex real-world problems which are difficult to be solved by general techniques.
- The learning model of RL is similar to the learning of human beings; hence most accurate results can be found.
- Helps in achieving long term results.

Disadvantage

- RL algorithms are not preferred for simple problems.
- RL algorithms require huge data and computations.
- Too much reinforcement learning can lead to an overload of states which can weaken the results.

Approaches to implement Reinforcement Learning

There are mainly three ways to implement reinforcement-learning in ML, which are:

1. **Value-based:**
   The value-based approach is about to find the optimal value function, which is the maximum value at a state under any policy. Therefore, the agent expects the long-term return at any state(s) under policy π.
2. **Policy-based:**
   Policy-based approach is to find the optimal policy for the maximum future rewards without using the value function. In this approach, the agent tries to apply such a policy that the action performed in each step helps to maximize the future reward.
   The policy-based approach has mainly two types of policy:
   - **Deterministic:** The same action is produced by the policy (π) at any state.
   - **Stochastic:** In this policy, probability determines the produced action.
3. **Model-based:** In the model-based approach, a virtual model is created for the environment, and the agent explores that environment to learn it. There is no

particular solution or algorithm for this approach because the model representation is different for each environment.

Elements of Reinforcement Learning

There are four main elements of Reinforcement Learning, which are given below:

1. Policy
2. Reward Signal
3. Value Function
4. Model of the environment

1) Policy: A policy can be defined as a way how an agent behaves at a given time. It maps the perceived states of the environment to the actions taken on those states. A policy is the core element of the RL as it alone can define the behavior of the agent. In some cases, it may be a simple function or a lookup table, whereas, for other cases, it may involve general computation as a search process. It could be deterministic or a stochastic policy:

2) Reward Signal: The goal of reinforcement learning is defined by the reward signal. At each state, the environment sends an immediate signal to the learning agent, and this signal is known as a **reward signal**. These rewards are given according to the good and bad actions taken by the agent. The agent's main objective is to maximize the total number of rewards for good actions. The reward signal can change the policy, such as if an action selected by the agent leads to low reward, then the policy may change to select other actions in the future.

3) Value Function: The value function gives information about how good the situation and action are and how much reward an agent can expect. A reward indicates the **immediate signal for each good and bad action**, whereas a value function specifies **the good state and action for the future**. The value function depends on the reward as, without reward, there could be no value. The goal of estimating values is to achieve more rewards.

4) Model: The last element of reinforcement learning is the model, which mimics the behavior of the environment. With the help of the model, one can make inferences about how the environment will behave. Such as, if a state and an action are given, then a model can predict the next state and reward.

The model is used for planning, which means it provides a way to take a course of action by considering all future situations before actually experiencing those situations. The approaches for solving the RL problems **with the help of the model** are termed as the **model-based approach**. Comparatively, an approach **without using a model** is called a **model-free approach**.

Linear Regression	**Logistic Regression**
Linear regression is used to predict the continuous dependent variable using a given set of independent variables.	Logistic Regression is used to predict the categorical dependent variable using a given set of independent variables.
Linear Regression is used for solving Regression problem.	Logistic regression is used for solving Classification problems.
In Linear regression, we predict the value of continuous variables.	In logistic Regression, we predict the values of categorical variables.
In linear regression, we find the best fit line, by which we can easily predict the output.	In Logistic Regression, we find the S-curve by which we can classify the samples.
Least square estimation method is used for estimation of accuracy.	Maximum likelihood estimation method is used for estimation of accuracy.
The output for Linear Regression must be a continuous value, such as price, age, etc.	The output of Logistic Regression must be a Categorical value such as 0 or 1, Yes or No, etc.
In Linear regression, it is required that relationship between dependent variable and independent variable must be linear.	In Logistic regression, it is not required to have the linear relationship between the dependent and independent variable.
In linear regression, there may be collinearity between the independent variables.	In logistic regression, there should not be collinearity between the independent variable.

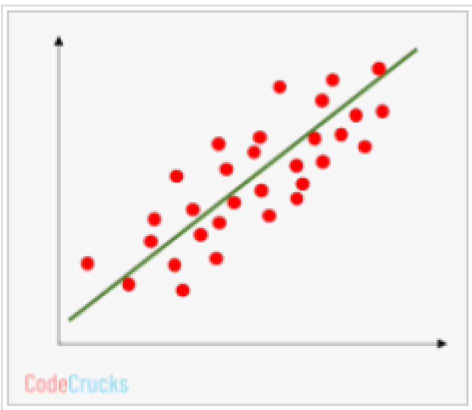
Parameter	Linear Regression	Logistic Regression
Fundamental task	Regression	Classification
Basic	The data is modelled using a straight line	The probability of some obtained event is represented as a linear function of a combination of predictor variables
Domain of predicted variable	Continuous	Discrete
Linear relationship between DV and IV	Required	Not required
The independent variable	Could be correlated with each other. (Especially in multiple linear regression)	Should not be correlated with each other (no multicollinearity exist).
Collinearity	There may be collinearity between the independent variables.	In logistic regression, there should not be collinearity between the independent variable.
Accuracy mesaure	Least square method	Maximum Likelihood estimation
Problem example:	House price prediction Student performance prediction	Tumor prediction (present or absent) Spam email classification (Spam or not spam)

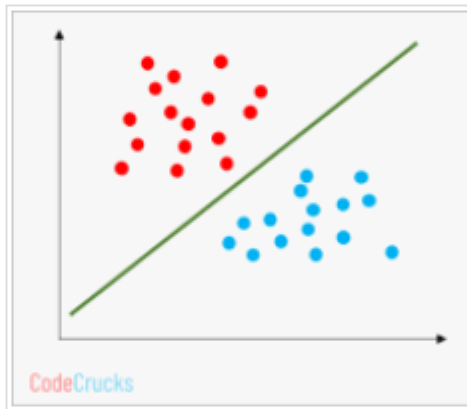
What is the difference between regression and classification?

Classification and regression are sometimes confusing to the new readers. Classification is used to provide discrete outcomes, as well as to categories data into specified categories. Classifying emails into spam and non-spam groups, for example.

Regression, on the other hand, works with continuous data. Predicting stock prices at a specific point in time, for example.

Geometric representation of both the concepts is shown below:





Types of ML Classification Algorithms:

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Kernel SVM
- Naïve Bayes
- Decision Tree Classification
- Random Forest Classification
- **Types of Regression Algorithm:**

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

Regression Algorithm	Classification Algorithm
In Regression, the output variable must be of continuous nature or real value.	In Classification, the output variable must be a c value.
The task of the regression algorithm is to map the input value (x) with the continuous output variable(y).	The task of the classification algorithm is to map th value(x) with the discrete output variable(y).
Regression Algorithms are used with continuous data.	Classification Algorithms are used with discrete dat
In Regression, we try to find the best fit line, which can predict the output more accurately.	In Classification, we try to find the decision bo which can divide the dataset into different classes.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to classification problems such as Identification of emails, Speech Recognition, Identification of cance etc.
The regression Algorithm can be further divided into Linear and Non-linear Regression.	The Classification algorithms can be divided into Classifier and Multi-class Classifier.

Q:14)What is Principal Component Analysis (PCA)? How it works?

We deal with multidimensional data in the real world. As the dimensions of data grow larger, data visualization and computation become more difficult. In such a case, we may need to reduce the dimensions in order to easily analyze and visualize the data. We accomplish this by removing irrelevant dimensions and keeping only the most relevant dimensions. This is where Principal Component Analysis comes into play (PCA).
The goals of Principal Component Analysis are to find a new set of

uncorrelated dimensions (orthogonal) and rank them based on variance.

Steps for PCA:

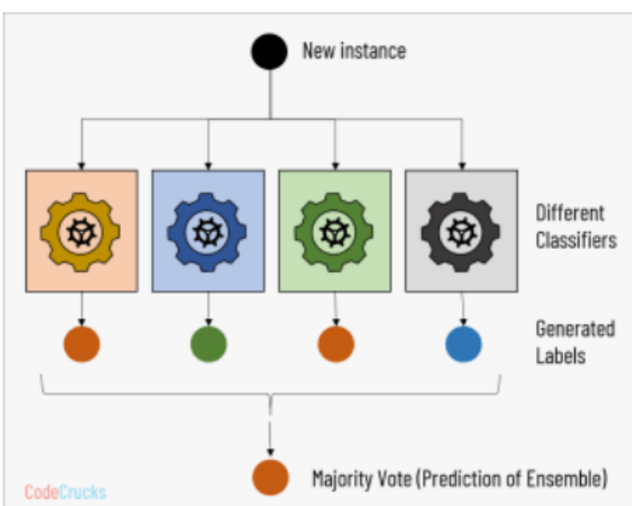
1. Find mean centered data from original dataset
2. Compute the covariance matrix for each data object
3. Compute the Eigen vectors and Eigen values of covariance matrix in descending order.
4. Project the data on first few eigen vectors to get reduced representation

Applications of Principal Component Analysis

- PCA is mainly used as the dimensionality reduction technique in various AI applications such **as computer vision, image compression, etc.**
- It can also be used for finding hidden patterns if data has high dimensions. Some fields where PCA is used are Finance, data mining, Psychology, etc.

Q:16)What is ensemble learning, and how does it work?

Ensemble learning is a strategy for creating more powerful machine learning models by combining numerous models.



There are numerous causes for a model's uniqueness. The following are a few reasons:

- Various populations
- Various hypotheses
- Various modelling methodologies

We will encounter an error when working with the model's training and testing data. Bias, variation, and irreducible error are all possible causes of this inaccuracy.

The model should now always exhibit a bias-variance trade-off, which we term a bias-variance trade-off.

This trade-off can be accomplished by ensemble learning.

There are a variety of ensemble approaches available, however there are two general strategies for aggregating several models:

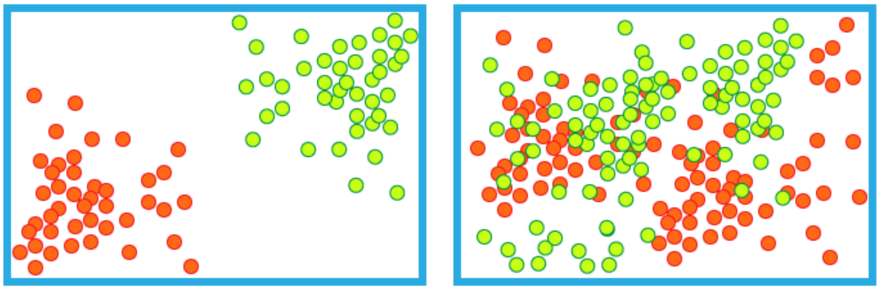
- Bagging, a native method: take a training set and use it to produce new training sets.
- Boosting, a more elegant method: boosting is used to optimize the best weighting scheme for a training set, comparable to bagging.

Q22)Introduction to Entropy in Machine Learning

Entropy is defined as the randomness or measuring the disorder of the information being processed in Machine Learning. Further, in other words, we can say that **entropy is the machine learning metric that measures the unpredictability or impurity in the system**.

What is Entropy in Machine Learning

Entropy is the measurement of disorder or impurities in the information processed in machine learning. It determines how a decision tree chooses to split data.

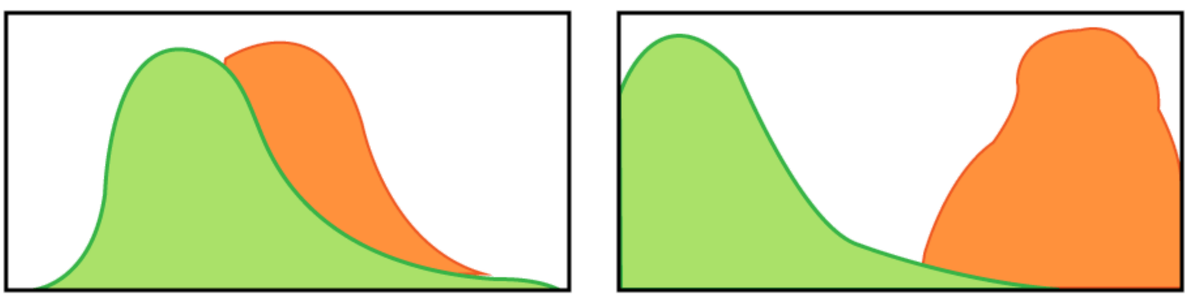


Low Entropy High Entropy

We can understand the term entropy with any simple example: flipping a coin. When we flip a coin, then there can be two outcomes. However, it is difficult to conclude what would be the exact outcome while flipping a coin because there is no direct relation between flipping a coin and its outcomes. There is a 50% probability of both outcomes; then, in such scenarios, entropy would be high. This is the essence of entropy in machine learning.

Use of Entropy in Decision Tree

In decision trees, heterogeneity in the leaf node can be reduced by using the cost function. At the root level, the entropy of the target column can be determined by the Shannon formula, in which Mr. Shannon has described the weighted entropy as the entropy calculated for the target column at every branch. However, in simple words, you can understand the weighted entropy as the individual weight of each attribute. Further, weights are considered as the probability of each class individually. The more the decrease in entropy, the more information is gained.



Low information gain High entropy High information gain Low entropy

What is the information gain in Entropy?

Information gain is defined as the pattern observed in the dataset and reduction in the entropy.

Mathematically, information gain can be expressed with the below formula:

Information Gain = (Entropy of parent node)-(Entropy of child node)

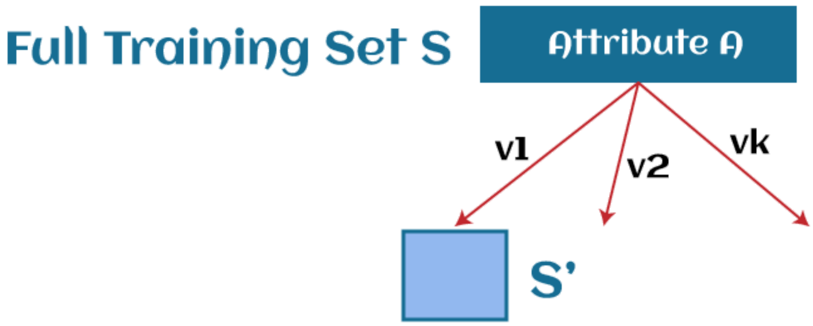
Let's understand it with an example having three scenarios as follows:

	Entropy	Information Gain
Scenario 1	0.7812345	0.2187655
Scenario 2	0	1
Scenario 3	1	0

How to build decision trees using information gain:

After understanding the concept of information gain and entropy individually now, we can easily build a decision tree. See steps to build a decision tree using information gain:

1. An attribute with the highest information gain from a set should be selected as the parent (root) node. From the image below, it is attributed A.



2. Build child nodes for every value of attribute A.
3. Repeat iteratively until you finish constructing the whole tree.

Q:23)What is Ensemble learning in Machine Learning?

Ensemble learning is one of the most powerful machine learning techniques that use the combined output of two or more models/weak learners and solve a particular computational intelligence problem. E.g., a Random Forest algorithm is an ensemble of various decision trees combined.

An ensembled model is a machine learning model that combines the predictions from two or more models."

There are 3 most common ensemble learning methods in machine learning. These are as follows:

- Bagging
- Boosting
- Stacking

1. Bagging

Bagging is a method of ensemble modeling, which is primarily used to solve supervised machine learning problems. It is generally completed in two steps as follows:

- **Bootstrapping:** It is a random sampling method that is used to derive samples from the data using the replacement procedure. In this method, first, random data samples are fed to the primary model, and then a base learning algorithm is run on the samples to complete the learning process.
- **Aggregation:** This is a step that involves the process of combining the output of all base models and, based on their output, predicting an aggregate result with greater accuracy and reduced variance.

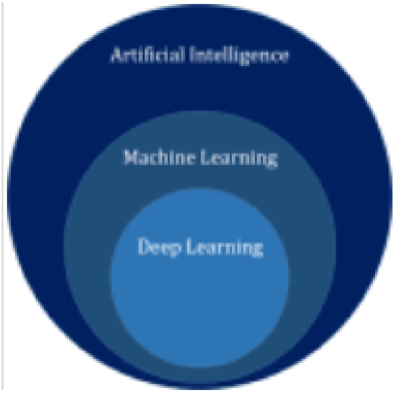
What are the Implementation Steps of Bagging?

- **Step 1:** Multiple subsets are made from the original information set with identical tuples, deciding on observations with replacement.
- **Step 2:** A base model is created on all subsets.
- **Step 3:** Every version is found in parallel with each training set and unbiased.
- **Step 4:** The very last predictions are determined by combining the forecasts from all models.

Machine Learning	Deep Learning
Allows machines to make judgments on their own based on previous data.	Allows machines to make decisions using artificial neural networks.
It requires only a minimal quantity of data for training	It necessitates a substantial amount of training data.
Works well on low-end systems, so massive machines are not required.	Because it necessitates a significant amount of computational power, it necessitates the usage of high-end equipment.
The majority of features must be identified ahead of time and manually coded.	The machine learns the characteristics from the data that is sent to it.
The problem is separated into two pieces that are solved separately and then combined.	The issue is resolved from beginning to end.

Deep learning or popularly known as deep neural network are special class of machine learning algorithms. Philosophy of deep learning algorithms is derived from the way human brain works. Deep learning algorithms are quite useful in solving computer vision problems. Deep learning algorithms got popularity after the ImageNet challenge organized in 2012. AlexNet proposed to solve the ImageNet – 2012 was a big break through in dep learning.

Relation between AI, ML and DL algorithms is depicted in following figure.



Deep Learning allows machines to make various business-related decisions using artificial neural networks, which is one of the reasons why it needs a vast amount of data for training. Since there is a lot of computing power required, it requires high-end systems as well. The systems acquire various properties and features with the help of the given data, and the problem is solved using an end-to-end method.

Deep learning algorithms are in fact a subset of machine learning algorithms. Machine learning algorithm requires hand crafted features as an input where as deep learning model does not require manual feature extraction.

Deep learning applications

Self-Driving Cars, Voice Controlled Assistance, Automatic Image Caption Generation, Automatic Machine Translation

Advantages

- It lessens the need for feature engineering.
- It eradicates all those costs that are needless.
- It easily identifies difficult defects.
- It results in the best-in-class performance on problems.

Disadvantages

- It requires an ample amount of data.
- It is quite expensive to train.

- It does not have strong theoretical groundwork.