Machine Learning (Unit-III Chapter-2)

1). Introduction to Machine Learning(ML)

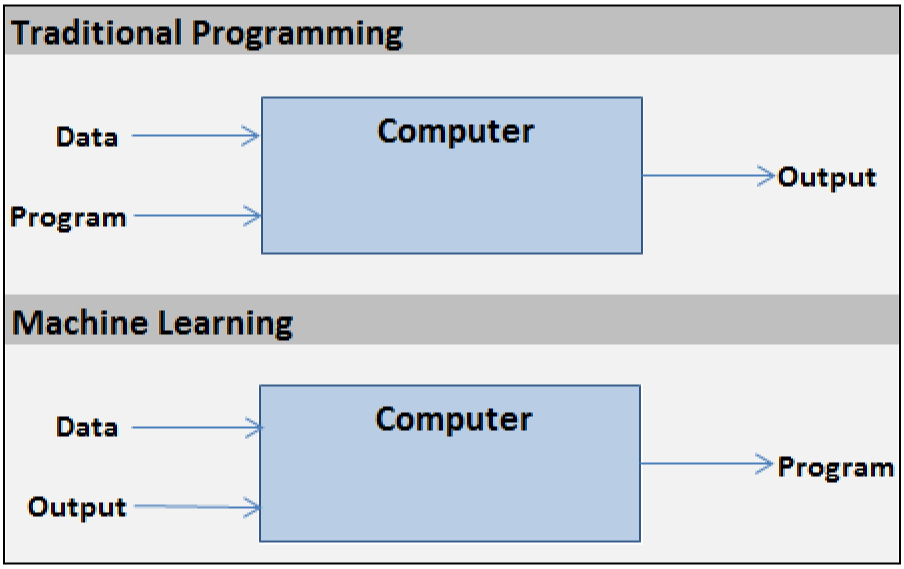
Machine Learning is an Application of Artificial Intelligence (AI) it gives devices the ability to learn from their experiences and improve their self without doing any coding. For Example, when you shop from any website it's shows related search like:- People who bought also saw this.

Definition:- Machine Learning is the "**Field of study that gives computers the capability to learn without being explicitly programmed".**

Machine Learning is a subset of Artificial Intelligence. Machine Learning is the study of making machines more human-like in their behaviour and decisions by giving them the ability to learn and develop their own programs. This is done with minimum human intervention, i.e., no explicit programming. The learning process is automated and improved based on the experiences of the machines throughout the process. Good quality data is fed to the machines, and different algorithms are used to build ML models to train the machines on this data. The choice of algorithm depends on the type of data at hand, and the type of activity that needs to be automated.

Difference between Traditional programming and Machine Learning :-

In traditional programming, we would feed the input data and a well written and tested program into a machine to generate output. When it comes to machine learning, input data along with the output is fed into the machine during the learning phase, and it works out a program for itself. To understand this better, refer to the illustration below:



Applications of Machine Learning :- The following are the some of the applications of Machine Learning:

- **Facial recognition/Image recognition**
    - The most common application of machine learning is Facial Recognition, and the simplest example of this application is the iPhone X. There are a lot of use-cases of facial recognition, mostly for security purposes like identifying criminals, searching for missing individuals, aid forensic investigations, etc. Intelligent marketing, diagnose diseases, track attendance in schools, are some other uses.
- **Automatic Speech Recognition**

Abbreviated as ASR, automatic speech recognition is used to convert speech into digital text. Its applications lie in authenticating users based on their voice and performing tasks based on the human voice inputs. Speech patterns and vocabulary are fed into the system to train the model. Presently ASR systems find a wide variety of applications in the following domains:

- Medical Assistance
- Industrial Robotics
- Forensic and Law enforcement
- Defence & Aviation
- Telecommunications Industry
- Home Automation and Security Access Control
- I.T. and Consumer Electronics

- **Financial Services**
    - Machine learning has many use cases in Financial Services. Machine Learning algorithms prove to be excellent at detecting frauds by monitoring activities of each user and assess that if an attempted activity is typical of that user or not.
    - Machine Learning also helps in making better trading decisions with the help of algorithms that can analyse thousands of data sources simultaneously. Credit scoring and underwriting are some of the other applications.

- **Marketing and Sales**
    - Machine Learning is improving lead scoring algorithms by including various parameters such as website visits, emails opened, downloads, and clicks to score each lead. It also helps businesses to improve their dynamic pricing models by using regression techniques to make predictions.
    - Sentiment Analysis is another essential application to consumer response to a specific product or a marketing initiative. Machine Learning for Computer Vision helps brands identify their products in images and videos online.

- **Healthcare**
    - A vital application of Machine Learning is in the diagnosis of diseases. Radiotherapy is also becoming better with Machine Learning taking over.
    - Early-stage drug discovery is another crucial application which involves technologies such as precision medicine and next-generation sequencing. Clinical trials cost a lot of time and money to complete and deliver results. Applying Machine Learning based predictive analytics could improve on these factors and give better results.
    - Machine Learning technologies are also critical to make outbreak predictions. Scientists around the world are using these technologies to predict epidemic outbreaks.

- **Recommendation Systems**
    - Many businesses today use recommendation systems to effectively communicate with the users on their site. It can recommend relevant products, movies, web-series, songs, and much more. Most prominent use-cases of recommendation systems are e-commerce sites like Amazon, Flipkart, and many others, along with Spotify, Netflix, and other web-streaming channels.

2). Types of Machine Learning

The following are the 4 different types of machine learning:

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

1. Supervised Learning:- The supervised learning model has a set of input variables (x), and an output variable (y). An algorithm identifies the mapping function between the input and output variables. The relationship is y = f(x).

The learning is monitored or supervised in the sense that we already know the output and the algorithm are corrected each time to optimise its results. The algorithm is trained over the data set and changed until it achieves an acceptable level of performance.

Supervised learning problems can be grouped as:

1. **Regression problems** – Used to predict future values and the model is trained with the historical data. E.g., Predicting the future price of a product.
2. **Classification problems** – Various labels train the algorithm to identify items within a specific category. E.g., Disease or no disease, Apple or an orange, Beer or wine.

2.**Unsupervised Learning**:- This approach is the one where the output is unknown, and we have only the input variable at hand. The algorithm learns by itself and discovers an impressive structure in the data. The goal is to convert the underlying distribution in the data to gain more knowledge about the data.

Unsupervised learning problems can be grouped as:

1. **Clustering** - This means grouping the input variables with the same characteristics together. E.g., grouping users based on search history
2. **Association** - Here, we discover the rules that govern meaningful associations among the data set. E.g., People who watch 'X' will also watch 'Y.'

3.Semi-supervised Learning:- In semi-supervised learning, data scientists train model with a minimal amount of labelled data and a large amount of unlabelled data. Usually, the first step is to cluster similar data with the help of an unsupervised machine learning algorithm. The next step is to label the unlabelled data using the characteristics of the limited labelled data available. After labelling the complete data, one can use supervised learning algorithms to solve the problem.

4.Reinforcement Learning:- In this approach, machine learning models are trained to make a series of decisions based on the rewards and feedback they receive for their actions. The machine learns to achieve a goal in complex and uncertain situations and is rewarded each time it achieves it during the learning period.

Reinforcement learning is different from supervised learning in the sense that there is no answer available, so the reinforcement agent decides the steps to perform a task. The machine learns from its own experiences when there is no training data set present.

3). Machine Learning concepts

A Machine Learning system **learns from historical data, builds the prediction models, and 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 output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem.

Features of Machine Learning:-

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

Need for Machine Learning:-

The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly. As a human, we have some limitations as we cannot access the huge amount of data manually, so for this, we need some computer systems and here comes the machine learning to make things easy for us.

We can train machine learning algorithms by providing them the huge amount of data and let them explore the data, construct the models, and predict the required output automatically. The performance of the machine learning algorithm depends on the amount of data, and it can be determined by the cost function. With the help of machine learning, we can save both time and money.

The importance of machine learning can be easily understood by its uses cases, Currently, machine learning is used in **self-driving cars**, **cyber fraud detection**, **face recognition**, and **friend suggestion by Facebook**, etc. Various top companies such as Netflix and Amazon have build machine learning models that are using a vast amount of data to analyze the user interest and recommend product accordingly.

Following are some key points which show the importance of Machine Learning:-

- Rapid increment in the production of data
- Solving complex problems, which are difficult for a human
- Decision making in various sector including finance
- Finding hidden patterns and extracting useful information from data.

Major building blocks of a Machine Learning system are : the model, the parameters, and the learner.

- **Model** is the system which makes predictions
- The **parameters** are the factors which are considered by the model to make predictions
- The **learner** makes the adjustments in the parameters and the model to align the predictions with the actual results

There are 5 basic steps used to perform a machine learning task:-

1. **Collecting data**: Collection of the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes.
2. **Preparing the data**: Any analytical process develops on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers.
3. **Training a model**: This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test; the first part (training data) is used for developing the model. The second part (test data), is used as a reference.
4. **Evaluating the model**: To test the accuracy, the second part of the data (holdout / test data) is used. This step determines the precision in the choice of the algorithm based on the outcome. A better test to check accuracy of model is to see its performance on data which was not used at all during model build.
5. **Improving the performance**: This step might involve choosing a different model altogether or introducing more variables to augment the efficiency. That’s why significant amount of time needs to be spent **in data collection and preparation.**

4). Concepts of Supervised and Unsupervised Learning

Supervised Learning

In supervised learning, the computer is taught by example. It learns from past data and applies the learning to present data to predict future values. In this case, both input and desired output data provide help to the prediction of future values. For accurate predictions, the input data is labeled or tagged as the right answer.

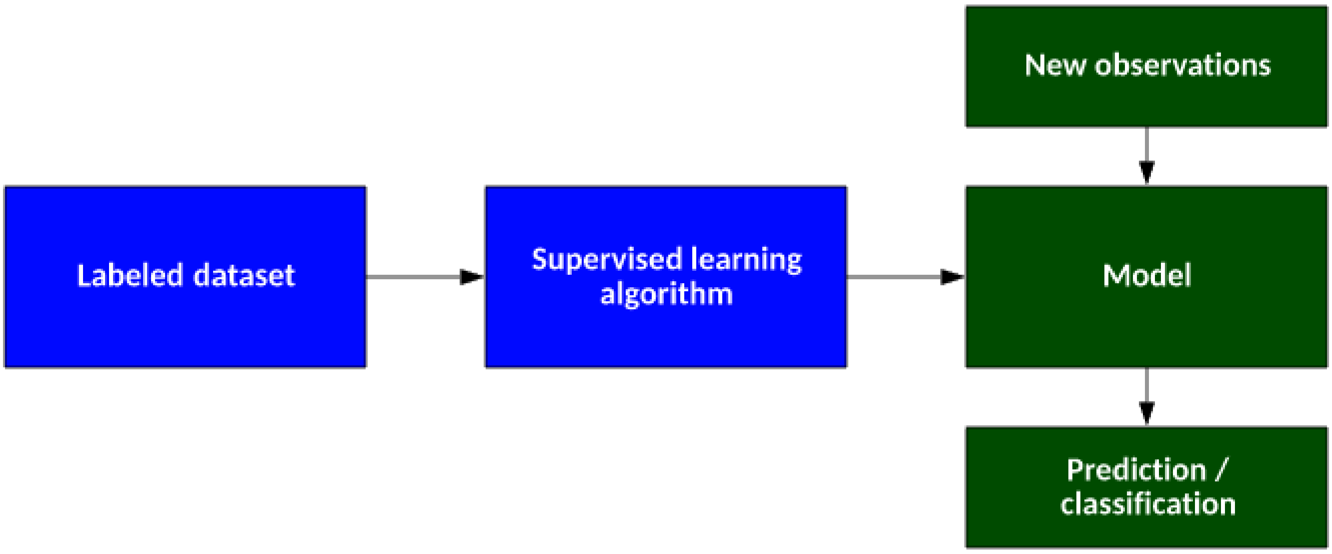
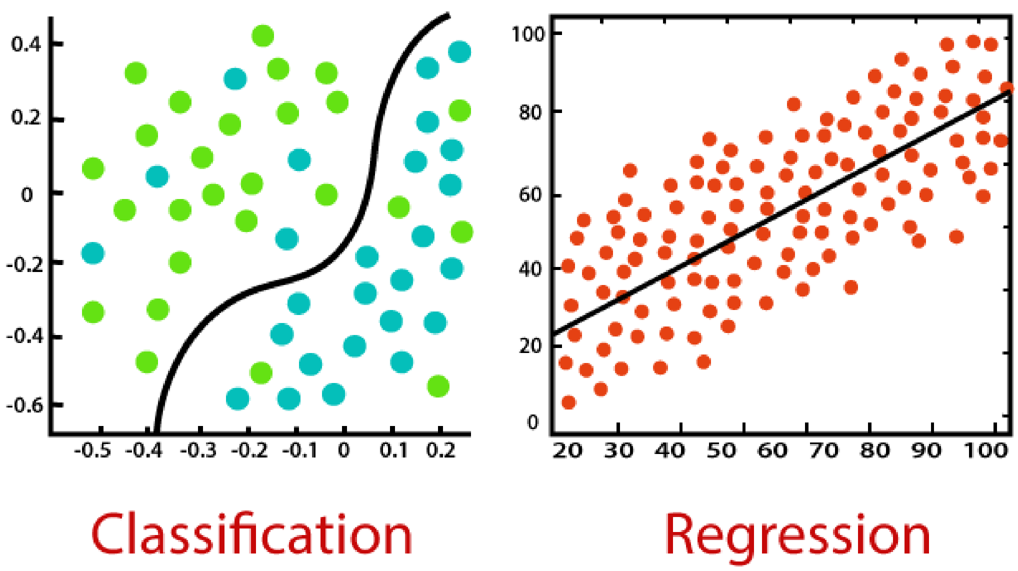


Fig : Supervised Machine Learning Categorisation

It is important to remember that all supervised learning algorithms are essentially complex algorithms, categorized as either classification or regression models.

1) Classification Models :– Classification models are used for problems where the output variable can be categorized, such as "Yes" or "No", or "Pass" or "Fail." Classification Models are used to predict the category of the data. Real-life examples include spam detection, sentiment analysis, scorecard prediction of exams, etc.

2) Regression Models :– Regression models are used for problems where the output variable is a real value such as a unique number, dollars, salary, weight or pressure, for example. It is most often used to predict numerical values based on previous data observations. Some of the more familiar regression algorithms include linear regression, logistic regression, polynomial regression, and ridge regression.



Applications of supervised learning algorithms in real life, including:

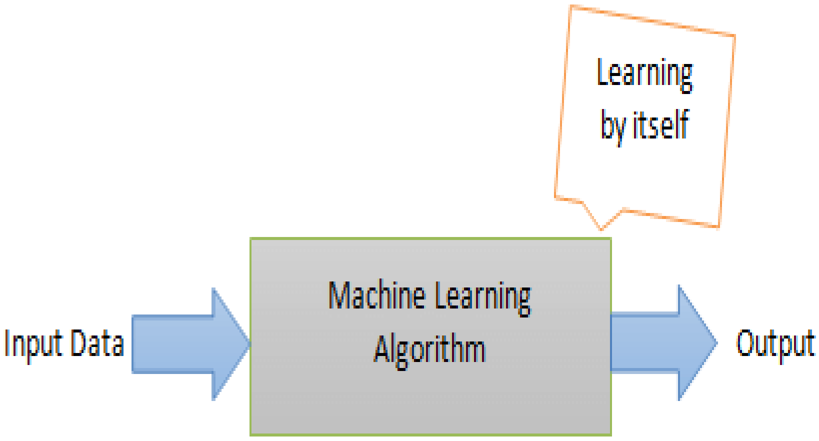
- Text categorization
- Face Detection
- Signature recognition
- Customer discovery
- Spam detection
- Weather forecasting
- Predicting housing prices based on the prevailing market price
- Stock price predictions, among others

Unsupervised Learning

Unsupervised learning, on the other hand, is the method that trains(tests) machines to use data that is neither classified nor labeled. It means no training data can be provided and the machine is made to learn by itself. The machine must be able to classify the data without any prior information about the data.

The idea is to expose the machines to large volumes of varying data and allow it to learn from that data to provide insights(an accurate and deep understanding) that were previously unknown and to identify hidden patterns. As such, there aren't necessarily defined outcomes from unsupervised learning algorithms. Rather, it determines what is different or interesting from the given dataset.

The machine needs to be programmed to learn by itself. The computer needs to understand and provide insights from both structured and unstructured data. Here's an accurate illustration of unsupervised learning:



Unsupervised Machine Learning Categorization:-

1) **Clustering** is one of the most common unsupervised learning methods. The method of clustering involves organizing unlabelled data into similar groups called clusters. Thus, a cluster is a collection of similar data items. The primary goal here is to find similarities in the data points and group similar data points into a cluster.

2) **Anomaly(outliers) detection** is the method of identifying rare items, events or observations which differ significantly from the majority of the data. We generally look for anomalies or outliers in data because they are suspicious. Anomaly detection is often utilized in bank fraud and medical error detection.

Applications of unsupervised learning algorithms include:

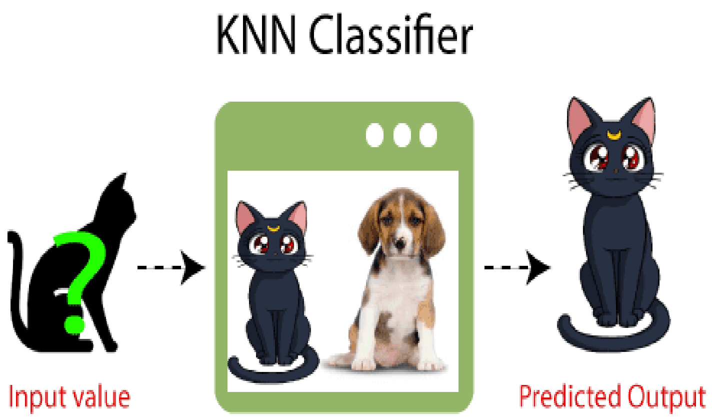
- Fraud detection
- Malware detection
- Identification of human errors during data entry
- Conducting accurate basket analysis, etc.

UNIT-IV

1). K-Nearest Neighbour(KNN) Algorithm for Machine Learning (or) Nearest Neighbour Classification.

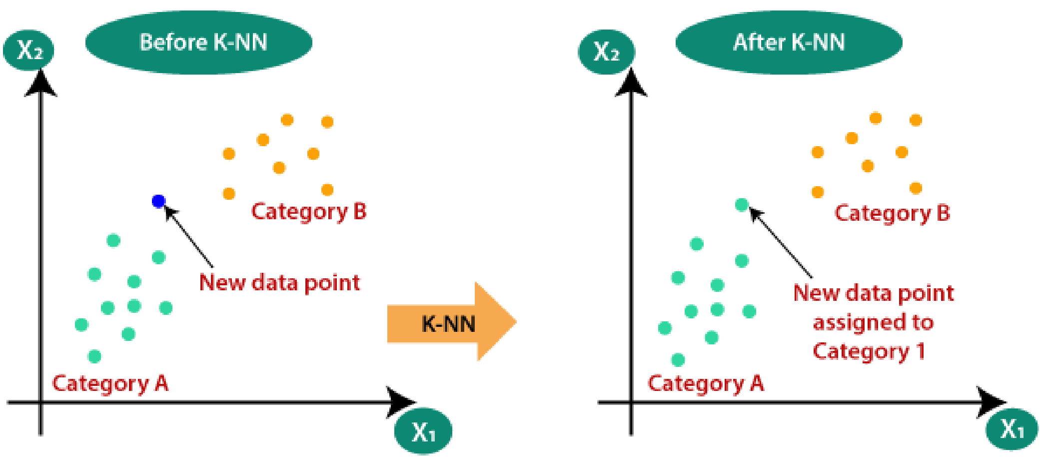
- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- KNN algorithm assumes the similarity between the new data and available data and put the new data into the category that is most similar to the available data categories.
- KNN algorithm 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 KNN algorithm.
- KNN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- KNN 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.



Need of KNN 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 KNN algorithm. With the help of KNN, we can easily identify the category or class of a particular data set. Consider the below diagram:

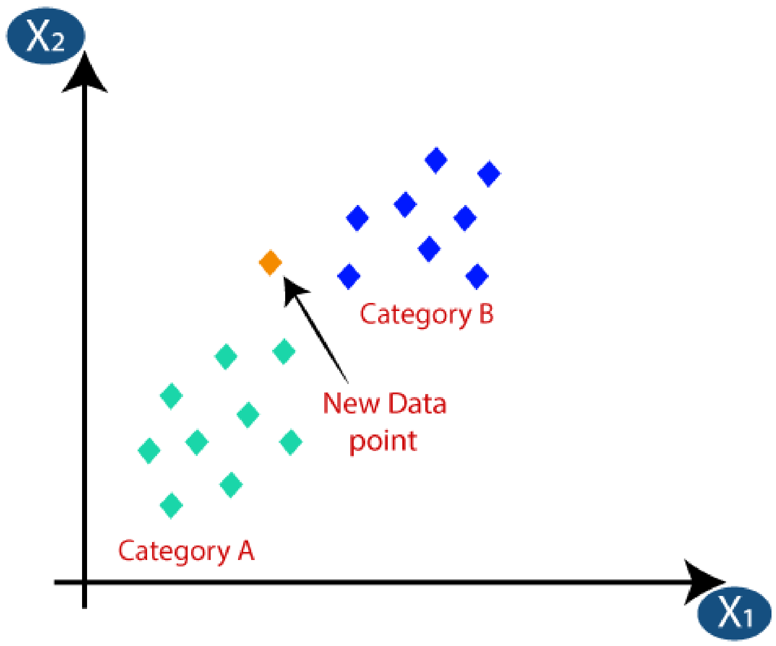


Working procedure of KNN:-

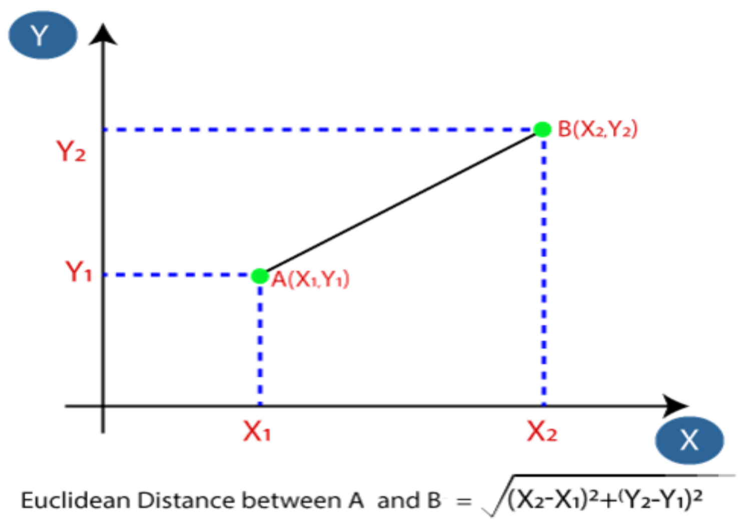
The KNN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbours
- **Step-2:** Calculate the Euclidean distance of **K number of neighbours**
- **Step-3:** Take the K nearest neighbours as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbours, 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 neighbour is maximum.
- **Step-6:** Our model is ready.

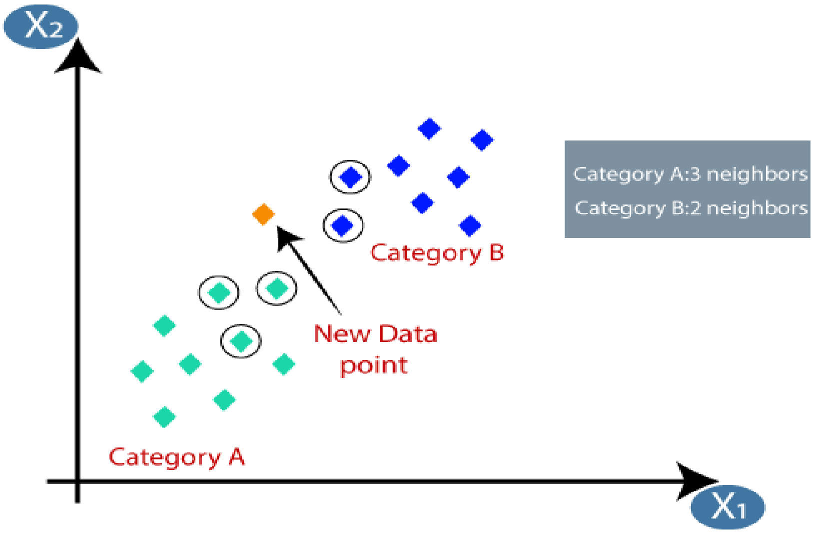
Suppose we have a new data point and we need to put it in the required category. Consider the below image:



- Firstly, we will choose the number of neighbours, so we will choose the k=5.
- Next, we will calculate the **Euclidean distance** between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:



- By calculating the Euclidean distance we got the nearest neighbours, as three nearest neighbours in category A and two nearest neighbours in category B. Consider the below image:



- As we can see the 3 nearest neighbours are from category A, hence this new data point must belong to category A.

Selecting the value of K in the KNN Algorithm:-

Below are some points to remember while selecting the value of K in the KNN algorithm:

- There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
- A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for K are good, but it may find some difficulties.

Advantages of KNN Algorithm:-

- It is simple to implement.
- It is robust to the noisy training data
- It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:-

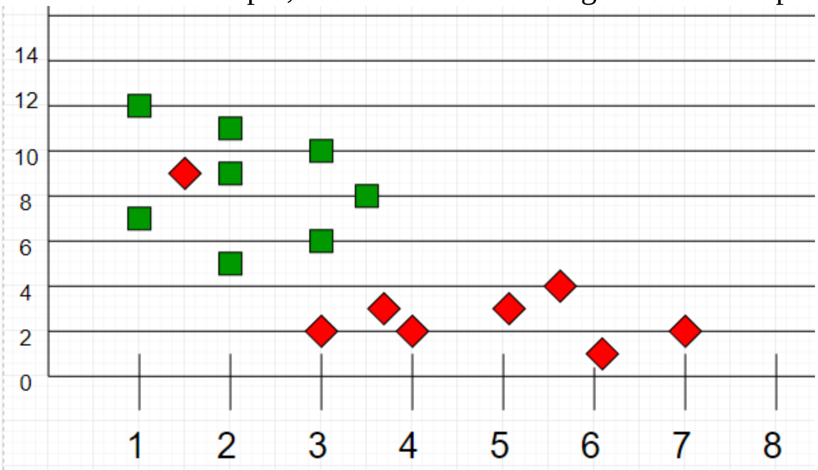
- Always needs to determine the value of K which may be complex some time.
- The computation cost is high because of calculating the distance between the data points for all the training samples.

2). A Probabilistic interpretation of Nearest Neighbours.

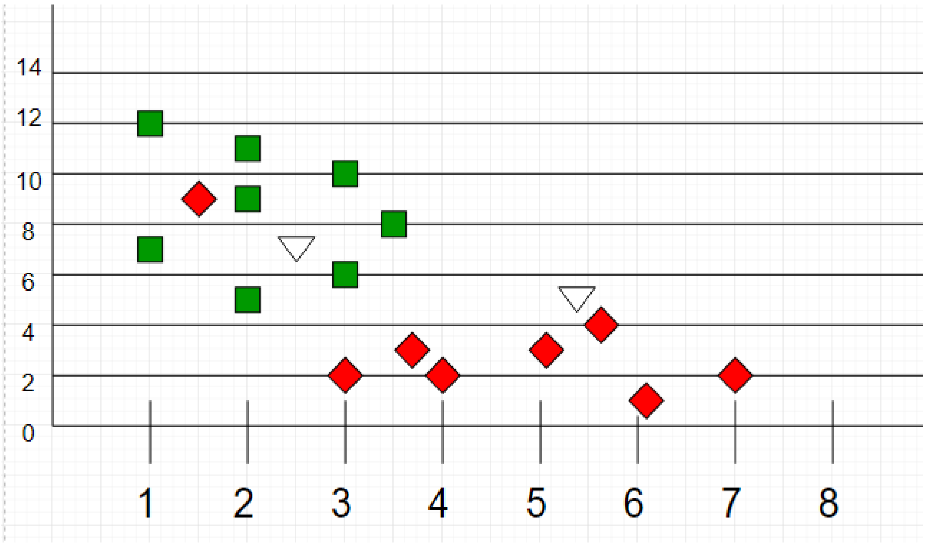
K-Nearest Neighbours is one of the most essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data. We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

As an example, consider the following table of data points containing two features:



Now, given another set of data points (also called testing data), allocate these points a group by analyzing the training set. Note that the unclassified points are marked as 'White'.



Intuition

If we plot these points on a graph, we may be able to locate some clusters or groups. Now, given an unclassified point, we can assign it to a group by observing what group its nearest neighbours belong to. This means a point close to a cluster of points classified as 'Red' has a higher probability of getting classified as 'Red'. Intuitively, we can see that the first point (2.5, 7) should be classified as 'Green' and the second point (5.5, 4.5) should be classified as 'Red'.

Algorithm

Let m be the number of training data samples. Let p be an unknown point.

1. Store the training samples in an array of data points arr[]. This means each element of this array represents a tuple (x, y).
2. for i=0 to m:
3. Calculate Euclidean distance d(arr[i], p).
4. Make set S of K smallest distances obtained. Each of these distances corresponds to an already classified data point.
5. Return the majority label among S.

3).Dimensionality Reduction Techniques

The number of input features, variables, or columns present in a given dataset is known as dimensionality, and the process to reduce these features is called **dimensionality reduction.**

A dataset contains a huge number of input features in various cases, which makes the predictive modeling task more complicated. Because it is very difficult to visualize or make predictions for the training dataset with a high number of features, for such cases, dimensionality reduction techniques are required to use.

Dimensionality reduction technique can be defined as, "It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information." These techniques are widely used in machine learning for obtaining a better fit predictive model while solving the classification and regression problems.

It is commonly used in the fields that deal with **high-dimensional data,** such as speech recognition, signal processing, bioinformatics, etc. It can also be used for data visualization, noise reduction, cluster analysis**,** etc**.**

Benefits of applying Dimensionality Reduction:-

Some benefits of applying dimensionality reduction technique to the given dataset are given below:

- By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
- Less Computation training time is required for reduced dimensions of features.
- Reduced dimensions of features of the dataset help in visualizing the data quickly.
- **I**t removes the redundant features (if present).

Disadvantages of dimensionality Reduction:-

There are also some disadvantages of applying the dimensionality reduction, which are given below:

- Some data may be lost due to dimensionality reduction.
- In the PCA dimensionality reduction technique, sometimes the principal components required to consider are unknown.

Approaches of Dimension Reduction:-

There are two ways to apply the dimension reduction technique, which are given below:

1. **Feature Selection**
2. **Feature Extraction**

1.Feature Selection:- Feature selection is the process of selecting the subset of the relevant features and leaving out the irrelevant features present in a dataset to build a model of high accuracy. In other words, it is a way of selecting the optimal features from the input dataset. Three methods are used for the feature selection:

1). Filters Methods:-

In this method, the dataset is filtered, and a subset that contains only the relevant features is taken. Some common techniques of filters method are:

- **Correlation**
- **Chi-Square Test**
- **ANOVA**
- **Information Gain, etc.**

2). Wrappers Methods:-

The wrapper method has the same goal as the filter method, but it takes a machine learning model for its evaluation. In this method, some features are fed to the ML model, and evaluate the performance. The performance decides whether to add those features or remove to increase the accuracy of the model. This method is more accurate than the filtering method but complex to work. Some common techniques of wrapper methods are:

- **Forward Selection**
- **Backward Selection**
- **Bi-directional Elimination**

3). Embedded Methods:-

Embedded methods check the different training iterations of the machine learning model and evaluate the importance of each feature. Some common techniques of Embedded methods are:

- **LASSO**
- **Elastic Net**
- **Ridge Regression, etc.**

2).Feature Extraction:-

Feature extraction is the process of transforming the space containing many dimensions into space with fewer dimensions.This approach is useful when we want to keep the whole information but use fewer resources while processing the information. Some common feature extraction techniques are:

- **Principal Component Analysis(PCA)**
- **Linear Discriminant Analysis**
- **Kernel PCA**
- **Quadratic Discriminant Analysis**

4). Principal Component Analysis (PCA)

Principal Component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the **Principal Components**. It is one of the popular tools that is used for exploratory data analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances.

Some common terms used in PCA algorithm:

- **Dimensionality:** It is the number of features or variables present in the given dataset. More easily, it is the number of columns present in the dataset.
- **Correlation:** It signifies that how strongly two variables are related to each other. Such as if one changes, the other variable also gets changed. The correlation value ranges from -1 to +1. Here, -1 occurs if variables are inversely proportional to each other, and +1 indicates that variables are directly proportional to each other.
- **Orthogonal:** It defines that variables are not correlated to each other, and hence the correlation between the pair of variables is zero.
- **Eigenvectors:** If there is a square matrix M, and a non-zero vector v is given. Then v will be eigenvector if Av is the scalar multiple of v.
- **Covariance Matrix:** A matrix containing the covariance between the pair of variables is called the Covariance Matrix.

Principal Components in PCA

As described above, the transformed new features or the output of PCA are the Principal Components. The number of these PCs are either equal to or less than the original features present in the dataset. Some properties of these principal components are given below:

- The principal component must be the linear combination of the original features.
- These components are orthogonal, i.e., the correlation between a pair of variables is zero.
- The importance of each component decreases when going to 1 to n, it means the 1 PC has the most importance, and n PC will have the least importance.

Steps for PCA algorithm

1. **Getting the dataset**
   Firstly, we need to take the input dataset and divide it into two subparts X and Y, where X is the training set, and Y is the validation set.
2. **Representing data into a structure**
   Now we will represent our dataset into a structure. Such as we will represent the two-dimensional matrix of independent variable X. Here each row corresponds to the data items, and the column corresponds to the Features. The number of columns is the dimensions of the dataset.
3. **Standardizing the data**
   In this step, we will standardize our dataset. Such as in a particular column, the features with high variance are more important compared to the features with lower variance.
   If the importance of features is independent of the variance of the feature, then we will divide each data item in a column with the standard deviation of the column. Here we will name the matrix as Z.
4. **Calculating the Covariance of Z**
   To calculate the covariance of Z, we will take the matrix Z, and will transpose it. After transpose, we will multiply it by Z. The output matrix will be the Covariance matrix of Z.
5. **Calculating the Eigen Values and Eigen Vectors**
   Now we need to calculate the eigenvalues and eigenvectors for the resultant covariance matrix Z. Eigenvectors or the covariance matrix are the directions of the axes with high information. And the coefficients of these eigenvectors are defined as the eigenvalues.
6. **Sorting the Eigen Vectors**
   In this step, we will take all the eigenvalues and will sort them in decreasing order, which means from largest to smallest. And simultaneously sort the eigenvectors accordingly in matrix P of eigenvalues. The resultant matrix will be named as P*.
7. **Calculating the new features Or Principal Components**
   Here we will calculate the new features. To do this, we will multiply the P* matrix to the Z. In the resultant matrix Z*, each observation is the linear combination of original features. Each column of the Z* matrix is independent of each other.
8. **Remove less or unimportant features from the new dataset.**
   The new feature set has occurred, so we will decide here what to keep and what to remove. It means, we will only keep the relevant or important features in the new dataset, and unimportant features will be removed out.

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.