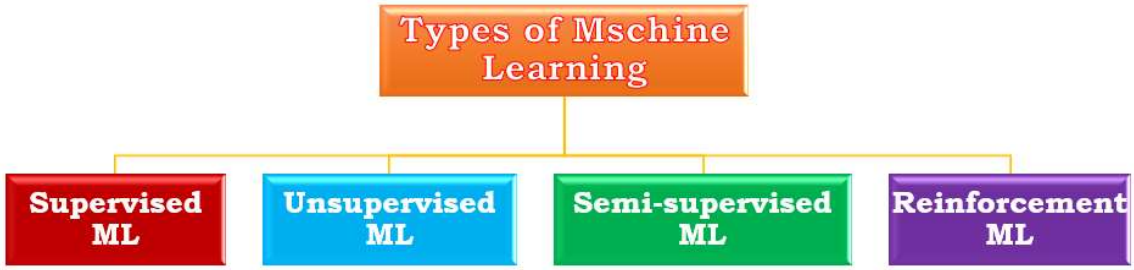
Machine Learning



Algorithm of Supervised ML

Logistic Regression

Logistic regression is a classification algorithm. It is used to predict a binary outcome based on a set of independent variables.\ It means a binary outcome is one where there are only two possible scenarios—either the event happens "yes" or it does not happen "No".

Used

- **Logistic regression is used to calculate the probability of a binary event occurring, and to deal with issues of classification.**
- **Logistic regression is used to predict the likelihood of all kinds of “yes” or “no” outcomes.**

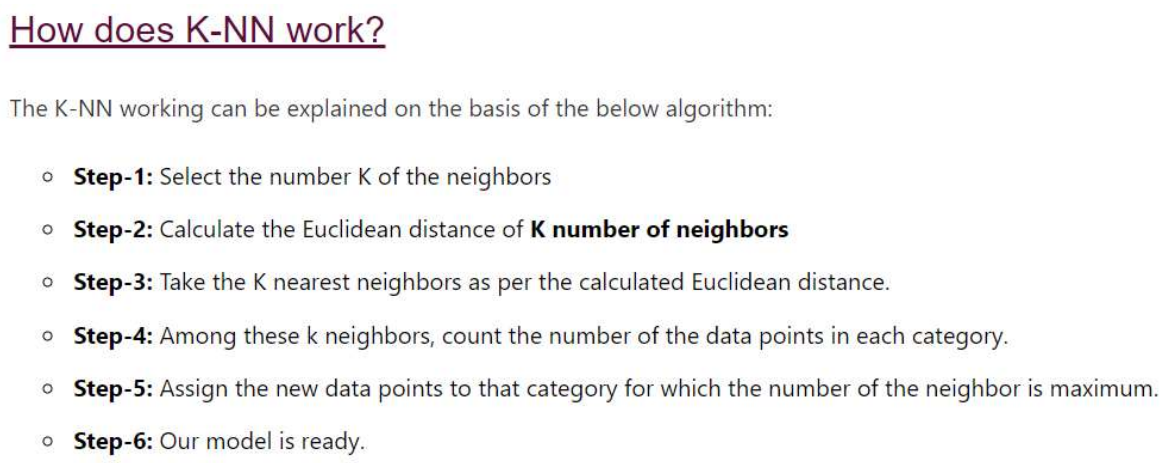
K-Nearest Neighbor(KNN)

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. It 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.



Support Vector Machine Algorithm

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.\ 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.

Kernel SVM

Introduce Kernel functions for sequence data, graphs, text, images, as well as vectors.\ SVM algorithms use a set of mathematical functions that are defined as the kernel. The function\ of kernel is to take data as input and transform it into the required form. Different SVM algorithms\ use different types of kernel functions. These functions can be different types. For example linear,\ nonlinear, polynomial, radial basis function (RBF), and sigmoid.

Naive Bayes

Naive 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.

- ***Naive:*** **It is called Naive because it assumes that the occurrence of a certain feature is independent of the occurrence of other features.**

- ***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:

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

Decision Tree Classification

Decision Tree algorithm is one of the simplest yet most powerful Supervised Machine Learning algorithms.\ Decision Tree algorithm can be used to solve both regression and classification problems in Machine Learning.\ That is why it is also known as CART or Classification and Regression Trees. As the name suggests, in Decision\ Tree, we form a tree-like model of decisions and their possible consequences.

Terminologies Related to Decision Tree Algorithms

- **Root Node:** This node gets divided into different homogeneous nodes. It represents the entire sample.
- **Splitting:** It is the process of splitting or dividing a node into two or more sub-nodes.
- **Interior Nodes:** They represent different tests on an attribute.
- **Branches:** They hold the outcomes of those tests.
- **Leaf Nodes:** When the nodes can't be split further, they are called leaf nodes.
- **Parent and Child Nodes:** The node from which sub-nodes are created is called a parent node. And, the sub-nodes are called the child nodes.

Random Forest Classification

Random Forest is a classifier that contains a number of decision trees on various subsets of the given\ dataset and takes the average to improve the predictive accuracy of that dataset. *Instead of relying on \ one decision tree, the random forest takes the prediction from each tree and based on the majority votes \ of predictions, and it predicts the final output.*

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

  How does Random Forest algorithm work?

  Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

  The Working process can be explained in the below steps and diagram:

  Step-1: Select random K data points from the training set.

  Step-2: Build the decision trees associated with the selected data points (Subsets).

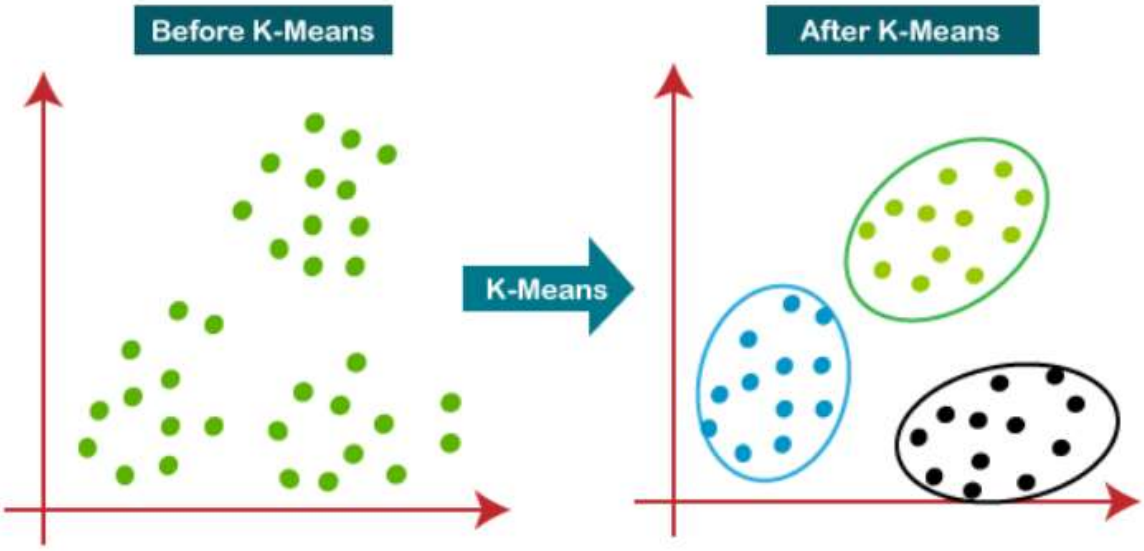
  Step-3: Choose the number N for decision trees that you want to build.

  Step-4: Repeat Step 1 & 2.

Algorithm of Unsupervised ML

K-mean Clustring

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.\ 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.



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.

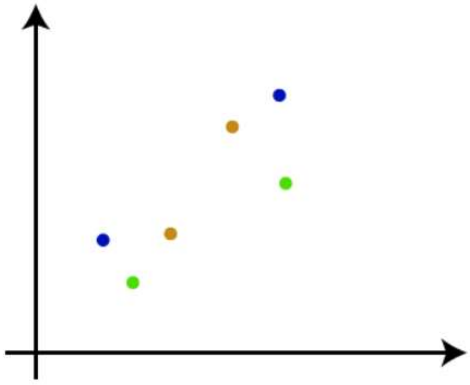
Hierarchical Clustering

Hierarchical clustering is another unsupervised machine learning algorithm, which is used to group the \ unlabeled datasets into a cluster and also known as hierarchical cluster analysis or HCA.

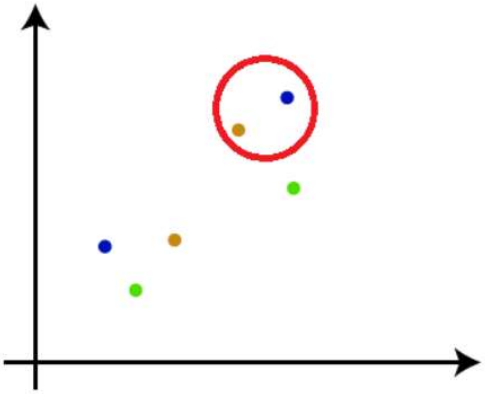
How the Agglomerative Hierarchical clustering Work?

The working of the AHC algorithm can be explained using the below steps:

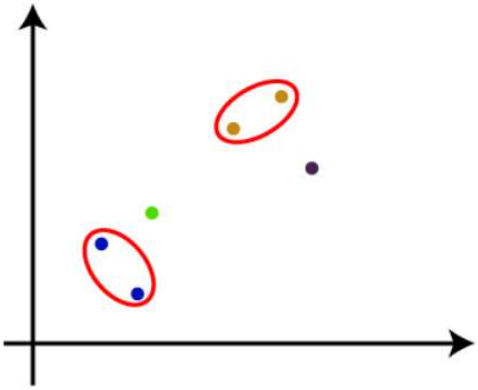
- **Step-1:** Create each data point as a single cluster. Let's say there are N data points, so the number of clusters will also be N.



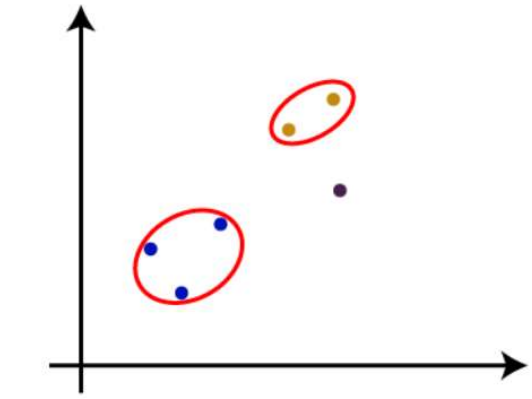
- **Step-2:** Take two closest data points or clusters and merge them to form one cluster. So, there will now be N-1 clusters.

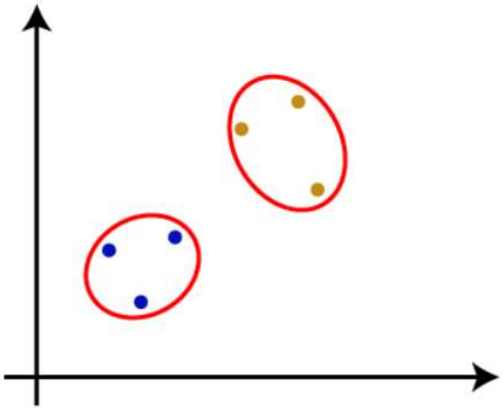


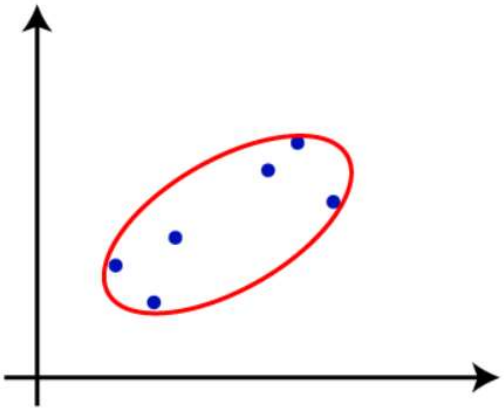
- **Step-3**: Again, take the two closest clusters and merge them together to form one cluster. There will be N-2 clusters.



- **Step-4:** Repeat Step 3 until only one cluster left. So, we will get the following clusters. Consider the below images:







- **Step-5:** Once all the clusters are combined into one big cluster, develop the dendrogram to divide the clusters as per the problem.

Model Free Reinforcement Learning

Model-free reinforcement learning does not need a model for the environment. It helps to learn a \ value function or a policy in a direct manner. The model-free reinforcement

learning tends to identify\ situations in which it is a suitable solution.

Intuition of model-free reinforcement learning

There are many different techniques for model-free reinforcement learning, all with the same basis:

- We execute many different *episodes* of the problem we want to solve, and from that we learn a *policy*.
- During learning, we try to learn the value of applying particular actions in particular states.
- During each episode, we need to execute some actions. After each action, we get a reward (which may be 0) and we can see the new state.
- From this, we *reinforce* our estimates of applying the previous action in the previous state.
- We terminate when: (1) we run out of training time; (2) we think our policy has converged to the optimal policy (for each new episode we see no improvement); or (3) our policy is 'good enough' (for each new episode we see minimal improvement).

Model Based Reinforcement Learning

"Model-based methods rely on planning as their primary component, while model-free methods primarily rely on learning."\ Model-based reinforcement learning is that it consists of a model of the environment. The model-based reinforcement learning\ helps to define the probability of ending up in the state t+1 with the current state and action. It is also known as transition\ probability with the explicit function to provide the next state efficiently.

- AI models which are called known models are provided to the agent without any need to go through a training process.
- The agent should learn the model, represented by a neural network, with sufficient and relevant experience.
- **The model-based reinforcement learning consists of an explicit model that needs fewer samples but there \ can be a probability of being inaccurate.**