SNS COLLEGE OF TECHNOLOGY

Coimbatore – 35

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade

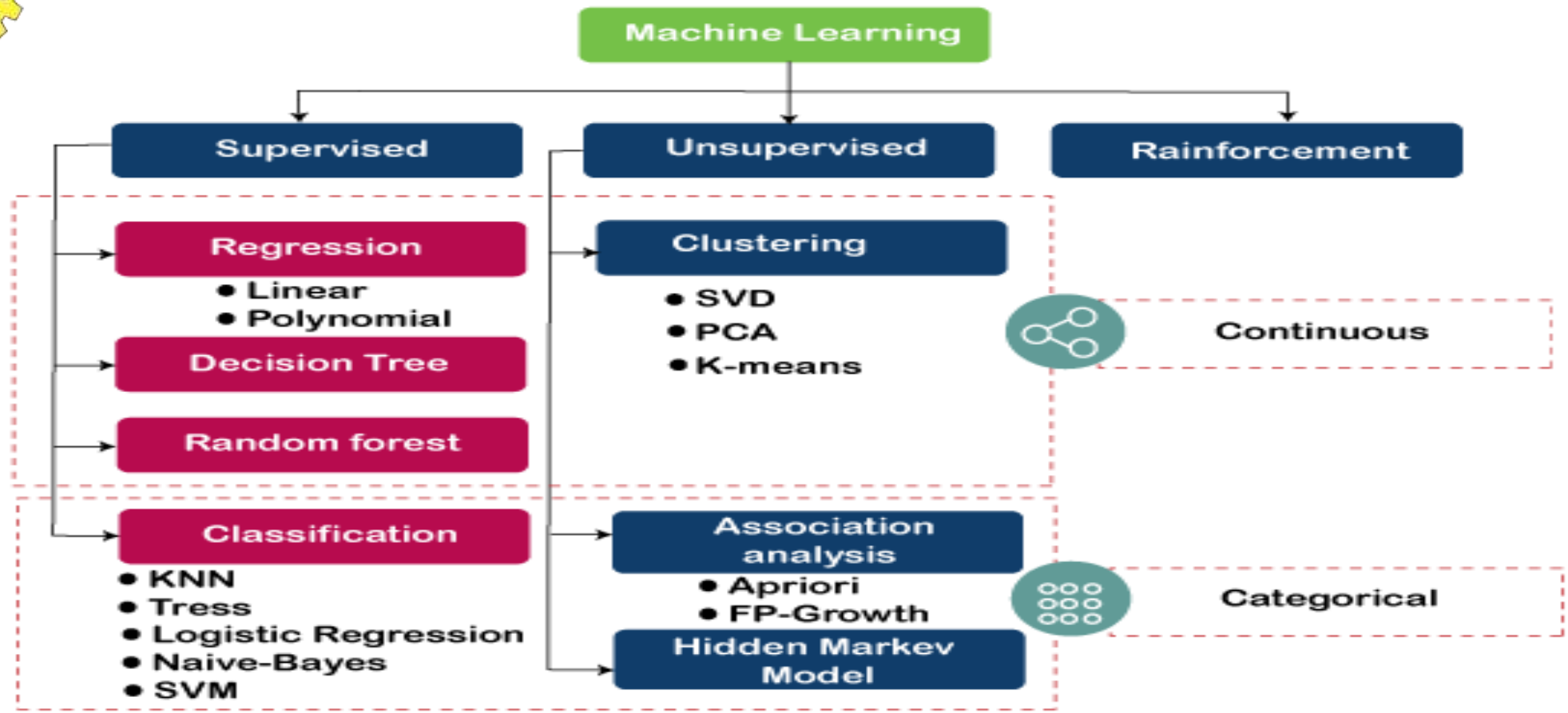
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

MACHINE LEARNING ALGORITHM

Machine Learning algorithm/ML/S.Rajarajeswari/AIML/SNSCT

Machine Learning Algorithms

Machine Learning algorithms are the programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences on their own. Different algorithms can be used in machine learning for different tasks, such as simple linear regression that can be used **for prediction problem**s like **stock market prediction,** and **the KNN algorithm can be used for classification problems**.


Machine Learning
Supervised
Unsupervised
Rainforcement
Regression
Linear
Polynomial
Decision Tree
Random forest
Classification
KNN
Tress
Logistic Regression
Naive-Bayes
SVM
Clustering
SVD
PCA
K-means
Association analysis
Apriori
FP-Growth
Hidden Markev Model
Continuous
Categorical

1) Supervised Learning Algorithm

- Supervised learning is a type of Machine learning in which the machine needs external supervision to learn. The supervised learning models are trained using the labeled dataset. Once the training and processing are done, the model is tested by providing a sample test data to check whether it predicts the correct output.
- The goal of supervised learning is to map input data with the output data. Supervised learning is based on supervision, and it is the same as when a student learns things in the teacher's supervision. The example of supervised learning is **spam filtering**.
- Classification
- Regression
- 2) **Unsupervised Learning Algorithm**
- It is a type of machine learning in which the machine does not need any external supervision to learn from the data, hence called unsupervised learning. The unsupervised models can be trained using the unlabelled dataset that is not classified, nor categorized, and the algorithm needs to act on that data without any supervision. In unsupervised learning, the model doesn't have a predefined output, and it tries to find useful insights from the huge amount of data. These are used to solve the Association and Clustering problems**. Hence further, it can be classified into two types:**
- Clustering
- Association

- 3) Reinforcement Learning
- In Reinforcement learning, an agent interacts with its environment by producing actions, and learn with the help of feedback. The feedback is given to the agent in the form of rewards, such as for each good action, he gets a positive reward, and for each bad action, he gets a negative reward. There is no supervision provided to the agent. **Q-Learning algorithm** is used in reinforcement learning. Read more…

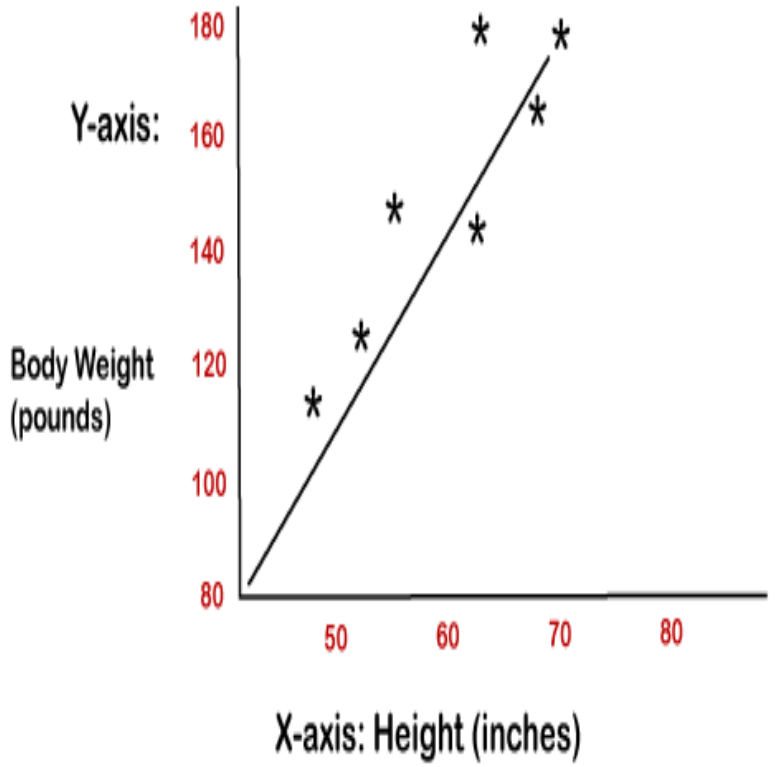
List of Popular Machine Learning Algorithm

1.Linear Regression Algorithm
2.Logistic Regression Algorithm
3.Decision Tree
4.SVM
5.Naïve Bayes
6.KNN
7.K-Means Clustering
8.Random Forest
9.Apriori
10.PCA

Linear

- Linear regression is one of the most popular and simple machine learning algorithms that is used for predictive analysis. Here, **predictive analysis** defines prediction of something, and linear regression makes predictions for *continuous numbers* such as **salary, age, etc.**
- It shows the linear relationship between the dependent and independent variables, and shows how the dependent variable(y) changes according to the independent variable (x).
- It tries to best fit a line between the dependent and independent variables, and this best fit line is knowns as the regression line.
- The equation for the regression line is:
- $y= a_0+ a*x+ b$

- a_0 = Intercept of line.
- Linear regression is further divided into two types:
- **Simple Linear Regression:** In simple linear regression, a single independent variable is used to predict the value of the dependent variable.
- **Multiple Linear Regression:** In multiple linear regression, more than one independent variables are used to predict the value of the dependent variable.



Logistic Regression

- Logistic regression is the supervised learning algorithm, which is used to **predict the categorical variables or discrete values**. It can be used for the *classification problems in machine learning*, and the output of the logistic regression algorithm can be either Yes or NO, 0 or 1, Red or Blue, etc.
- Logistic regression is similar to the linear regression except how they are used, such as Linear regression is used to solve the regression problem and predict continuous values, whereas Logistic regression is used to solve the Classification problem and used to predict the discrete values

Decision Tree Algorithm

- A decision tree is a supervised learning algorithm that is mainly used to solve the classification problems but can also be used for solving the regression problems. It can work with both categorical variables and continuous variables. It shows a tree-like structure that includes nodes and branches, and starts with the root node that expand on further branches till the leaf node. The **internal node** is used to represent the **features of the dataset, branches show the decision rules,** and **leaf nodes represent the outcome of the problem.**

- **Root Node:** Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

•**Leaf Node:** Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

•**Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions

•**Branch/Sub Tree:** A tree formed by splitting the tree

•.

•**Pruning:** Pruning is the process of removing the unwanted branches from the tree.

•**Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes

- Support Vector Machine Algorithm
- A support vector machine or SVM is a supervised learning algorithm that can also be used for classification and regression problems. However, it is primarily used for classification problems. The goal of SVM is to create a hyperplane or decision boundary that can segregate datasets into different classes.
- The data points that help to define the hyperplane are known as **support vectors**, and hence it is named as support vector machine algorithm.
- Some real-life applications of SVM are **face detection, image classification, Drug discovery**

Naïve Bayes Algorithm:

Naïve Bayes classifier is a supervised learning algorithm, which is used to make predictions based on the probability of the object. The algorithm named as Naïve Bayes as it is based on **Bayes theorem**, and follows the *naïve* assumption that says' variables are independent of each other.

The Bayes theorem is based on the conditional probability; it means the likelihood that event(A) will happen, when it is given that event(B) has already happened. The equation for Bayes theorem is given as:

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

Naïve Bayes classifier is one of the best classifiers that provide a good result for a given problem. It is easy to build a naïve bayesian model, and well suited for the huge amount of dataset. It is mostly used for text classification.

- K-Nearest Neighbour (KNN)
- K-Nearest Neighbour is a supervised learning algorithm that can be used for both classification and regression problems. This algorithm works by assuming the similarities between the new data point and available data points. Based on these similarities, the new data points are put in the most similar categories. It is also known as the lazy learner algorithm as it stores all the available datasets and classifies each new case with the help of K-neighbours. The new case is assigned to the nearest class with most similarities, and any distance function measures the distance between the data points. The distance function can be **Euclidean, Minkowski, Manhattan, or Hamming distance** based on the requirement.

K-Means Clustering

- K-means clustering is one of the simplest unsupervised learning algorithms, which is used to solve the clustering problems. The datasets are grouped into K different clusters based on similarities and dissimilarities, it means, datasets with most of the commonalties remain in one cluster which has very less or no commonalities between other clusters. In K-means, K-refers to the number of clusters, and **means** refer to the averaging the dataset in order to find the centroid.
- It is a centroid-based algorithm, and each cluster is associated with a centroid. This algorithm aims to reduce the distance between the data points and their centroids within a cluster. his algorithm starts with a group of randomly selected centroids that form the clusters at starting and then perform the iterative process to optimize these centroids' positions.
- It can be used for spam detection and filtering, identification of fake news, etc

- Random Forest Algorithm
- Random forest is the supervised learning algorithm that can be used for both classification and regression problems in machine learning. It is an ensemble learning technique that provides the predictions by combining the multiple classifiers and improve the performance of the model.
- *It contains multiple decision trees for subsets of the given dataset, and find the average to improve the predictive accuracy of the model. A random-forest should contain 64-128 trees. The greater number of trees leads to higher accuracy of the algorithm.*
- To classify a new dataset or object, each tree gives the classification result and based on the majority votes, the algorithm predicts the final output

priori Algorithm

- Apriori algorithm is the unsupervised learning algorithm that is used to solve the association problems. It uses frequent itemsets to generate association rules, and it is designed to work on the databases that contain transactions. With the help of these association rule, it determines how strongly or how weakly two objects are connected to each other. This algorithm uses a breadth-first search and Hash Tree to calculate the itemset efficiently.

- The algorithm process iteratively for finding the frequent itemsets from the large dataset

- The apriori algorithm was given by the **R. Agrawal and Srikant** in the year 1994. It is mainly used for market basket analysis and helps to understand the products that can be bought together. It can also be used in the healthcare field to find drug reactions in patients.

rinciple Component Analysis

Principle Component Analysis (PCA) is an unsupervised learning technique, which is used for dimensionality reduction. It helps in reducing the dimensionality of the dataset that contains many features correlated with each other. 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. It is one of the popular tools that is used for exploratory data analysis and predictive modeling.

Probability in machine learning

Probability

- Machine Learning is an interdisciplinary field that uses statistics, probability, algorithms to learn from data and provide insights which can be used to build intelligent applications. In this article, we will discuss some of the key concepts widely used in machine learning.

Probability and statistics are related areas of mathematics which concern themselves with analyzing the relative frequency.

Probability deals with predicting the likelihood of future events, while *statistics* involves the analysis of the frequency of past events

- **Probability**
- Most people have an intuitive understanding of degrees of probability, which is why we use words like “probably” and “unlikely” in our daily conversation, but we will talk about how to make quantitative claims about those degrees

In probability theory, an **event** is a set of outcomes of an experiment to which a probability is assigned. If **E** represents an event, then **P(E)** represents the probability that **E**will occur. A situation where **E** might happen (*success*) or might not happen (*failure*) is called a ***trial***.T

- Let us consider a basic example of tossing a coin. If the coin is fair, then it is just as likely to come up heads as it is to come up tails. In other words, if we were to repeatedly toss the coin many times, we would expect about about half of the tosses to be heads and and half to be tails. In this case, we say that the probability of getting a head is 1/2 or 0.5
- **empirical probability** of an event is given by number of times the event occurs divided by the total number of incidents observed. If for n trials and we observe s successes, the probability of success is s/n. In the above example. any sequence of coin tosses may have more or less than exactly 50% heads.

- **Theoretical probability** on the other hand is given by the number of ways the particular event can occur divided by the total number of possible outcomes. So a head can occur once and possible outcomes are two (head, tail). The true (theoretical) probability of a head is 1/2.

- **Joint Probability** Probability of events A and B denoted by **P(A and B) or P(A ∩ B)** is the probability that events A and B both occur. **P(A ∩ B) = P(A). P(B)** . This only applies if **A** and **B** are independent, which means that if **A** occurred, that doesn't change the probability of **B**, and vice versa.

Conditional probability

- Let us consider A and B are not independent, because if A occurred, the probability of B is higher. When A and B are not independent, it is often useful to compute the conditional probability, P (A|B), which is the probability of A given that B occurred: ∴ **P(A|B) = P(A ∩ B)/ P(B).**
- Similarly, **P(B|A) = P(A ∩ B)/ P(A)** . We can write the joint probability of as A and B as **P(A ∩ B)= p(A).P(B|A)**, which means : *"The chance of both things happening is the chance that the first one happens, and then the second one given the first happened."*

Introduction to Bayes Theorem in Machine Learning

- Bayes theorem is given by an English statistician, philosopher, and Presbyterian minister named **Mr. Thomas Bayes** in 17th century. Bayes provides their thoughts in decision theory which is extensively used in important mathematics concepts as Probability. Bayes theorem is also widely used in Machine Learning where we need to predict classes precisely and accurately. An important concept of Bayes theorem named **Bayesian method** is used to calculate conditional probability in Machine Learning application that includes classification tasks. Further, a simplified version of Bayes theorem (Naïve Bayes classification) is also used to reduce computation time and average cost of the projects.
- Bayes theorem is also known with some other name such as **Bayes rule or Bayes Law.** ***Bayes theorem helps to determine the probability of an event with random knowledge***. It is used to calculate the probability of occurring one event while other one already occurred. It is a best method to relate the condition probability and marginal probability.

- Bayes theorem is one of the most popular machine learning concepts that helps to calculate the probability of occurring one event with uncertain knowledge while other one has already occurred.
- Bayes' theorem can be derived using product rule and conditional probability of event X with known event Y:
  •

  According to the product rule we can express as the probability of event X with known event Y as follows;

1.P(X ? Y)= P(X|Y) P(Y) {equation 1}

• urther, the probability of event Y with known event X:

1.P(X ? Y)= P(Y|X) P(X) {equation 2}

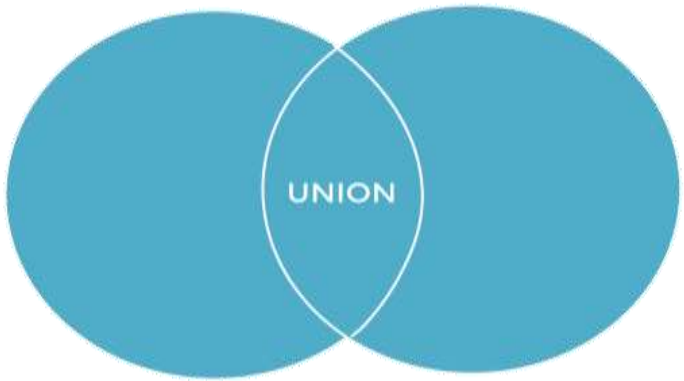
Mathematically, Bayes theorem can be expressed by combining both equations on right hand side

The below equation iscalled as Bayes Rule or Bayes Theorem.

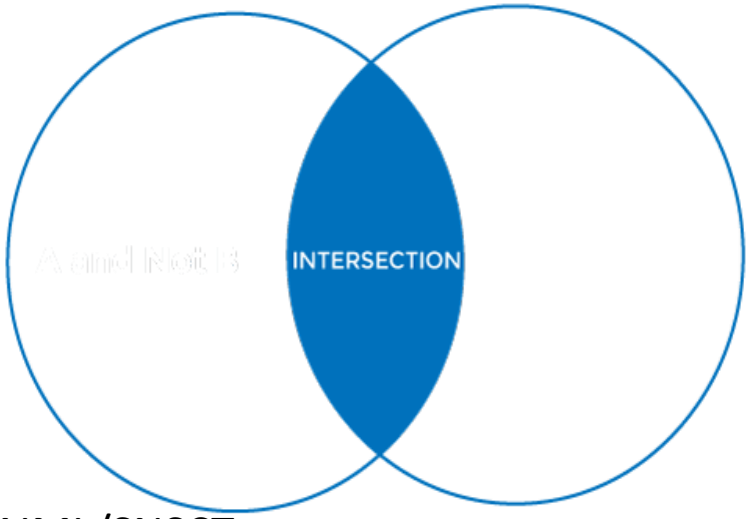
$$P(X|Y) = \frac{P(Y|X).P(X)}{P(Y)}$$

- Assume in our experiment of rolling a dice, there are two event A and B such that;
- A = Event when an even number is obtained = {2, 4, 6}
- B = Event when a number is greater than 4 = {5, 6}
- **Probability of the event A "P(A)"**= Number of favourable outcomes / Total number of possible outcomes
  P(E) = 3/6 =1/2 =0.5
- Similarly, **Probability of the event B "P(B)"**= Number of favourable outcomes / Total number of possible outcomes
  =2/6
  =1/3
  =0.333

- **Union of event A and B:**
  A∪B = {2, 4, 5, 6}



- **Intersection of event A and B:**
  A∩B= {6}



- **Disjoint Event:** If the intersection of the event A and B is an empty set or null then such events are known as **disjoint event** or **mutually exclusive events** also.
- **Random Variable:**
- It is a real value function which helps mapping between sample space and a real line of an experiment. A random variable is taken on some random values and each value having some probability. However, it is neither random nor a variable but it behaves as a function which can either be discrete, continuous or combination of both.

- **Exhaustive Event:**
- As per the name suggests, a set of events where at least one event occurs at a time, called exhaustive event of an experiment.
- Thus, two events A and B are said to be exhaustive if either A or B definitely occur at a time and both are mutually exclusive for e.g., while tossing a coin, either it will be a Head or may be a Tail.

- **Independent Event:**
- Two events are said to be independent when occurrence of one event does not affect the occurrence of another event. In simple words we can say that the probability of outcome of both events does not depends one another.
- Mathematically, two events A and B are said to be independent if:
- $P(A \cap B) = P(AB) = P(A)*P(B)$

Marginal Probability:

Marginal probability is defined as the probability of an event A occurring independent of any other event B. Further, it is considered as the probability of evidence under any consideration.

- P(A) = P(A|B)*P(B) + P(A|~B)*P(~B)

- Here ~B represents the event that B does no
  occur