IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114137

Research on Machine learning and Its Algorithms

Kamini Ahire[1], Sheetal Wadhai[2]

Department of Computer Engineering Universal College of Engineering and Research, Sasewadi, Pune

Abstract: The science of getting computers to act without being explicitly programmed is known as machine learning. Machine learning, data mining, and statistical pattern identification are all covered in this article. It examines common machine learning algorithms like the decision tree algorithm, random forest algorithm, artificial neural network algorithm, SVM algorithm, Boosting and Bagging algorithm, and BP algorithm. The machine learning development environment you use could be just as important as the machine learning methods you employ to solve your predictive modelling problem.

INTRODUCTION:

The science of getting computers to act without being explicitly programmed is known as machine learning. Self-driving cars, realistic speech recognition, successful web search, and a much-enhanced understanding of the human genome have all been made possible by machine learning in the last decade. Machine learning is now so common that you probably use it thousands of times per day without even realizing it. Many academics believe it is the most effective technique to get closer to human-level AI. Machine Learning algorithms are systems that can learn hidden patterns from data, forecast output, and improve performance based on their own experiences. It can provide guidance for subsequent machine learning development by conducting a realistic analysis of machine learning algorithms, hence boosting the applicability of machine learning machine learning algorithms and making the industry's economic development more convenient.

2.WHAT IS MACHINE LEARNING:

Artificial intelligence has a subfield called machine learning (AI). The goal of machine learning is to comprehend the structure of data and fit that data into models that people can comprehend and use.
Machine learning is a branch of computer science that is distinct from standard computational methods. Algorithms are sets of clearly programmed instructions used by computers to calculate or solve problems in traditional computing. Machine learning techniques, on the other hand, allow computers to train on data inputs and then utilise statistical analysis to produce results that are within a certain range. As a result, machine learning makes it easier for computers to create models from sample data and automate decision-making processes based on data inputs.
Machine learning has benefited everyone who uses technology today. Social media networks can utilise facial recognition technology to help users tag and share images of pals. The technology of optical character recognition (OCR) turns text pictures into moveable type. Machine learning-powered recommendation systems propose what movies or TV shows to watch next depending on user interests. Consumers may soon be able to purchase self-driving cars that use machine learning to navigate.
Machine learning is a field that is constantly evolving. As a result, there are a few things to think about when working with machine learning methodology or analysing the impact of machine learning procedures.
We'll look at supervised and unsupervised machine learning methods, as well as common algorithmic approaches in machine learning, such as the k-nearest neighbour algorithm, decision tree learning, and deep learning, in this tutorial. We'll look at which programming languages are most commonly used in machine learning, as well as some of the advantages and disadvantages of each. We'll also talk about how machine learning algorithms perpetuate biases and what may be done to avoid them when developing algorithms.

Machine Learning Algorithms

Machine Learning algorithms are systems that can learn hidden patterns from data, forecast output, and improve performance based on their own experiences. In machine learning, multiple algorithms can be employed for different tasks, such as simple linear regression for prediction problems like stock market prediction and the KNN algorithm for categorization challenges.
This article will provide an overview of some of the most popular and widely used machine learning algorithms, as well as their use cases and categories.

Types of Machine Learning Algorithms

Machine Learning Algorithm can be broadly classified into three types:

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114137

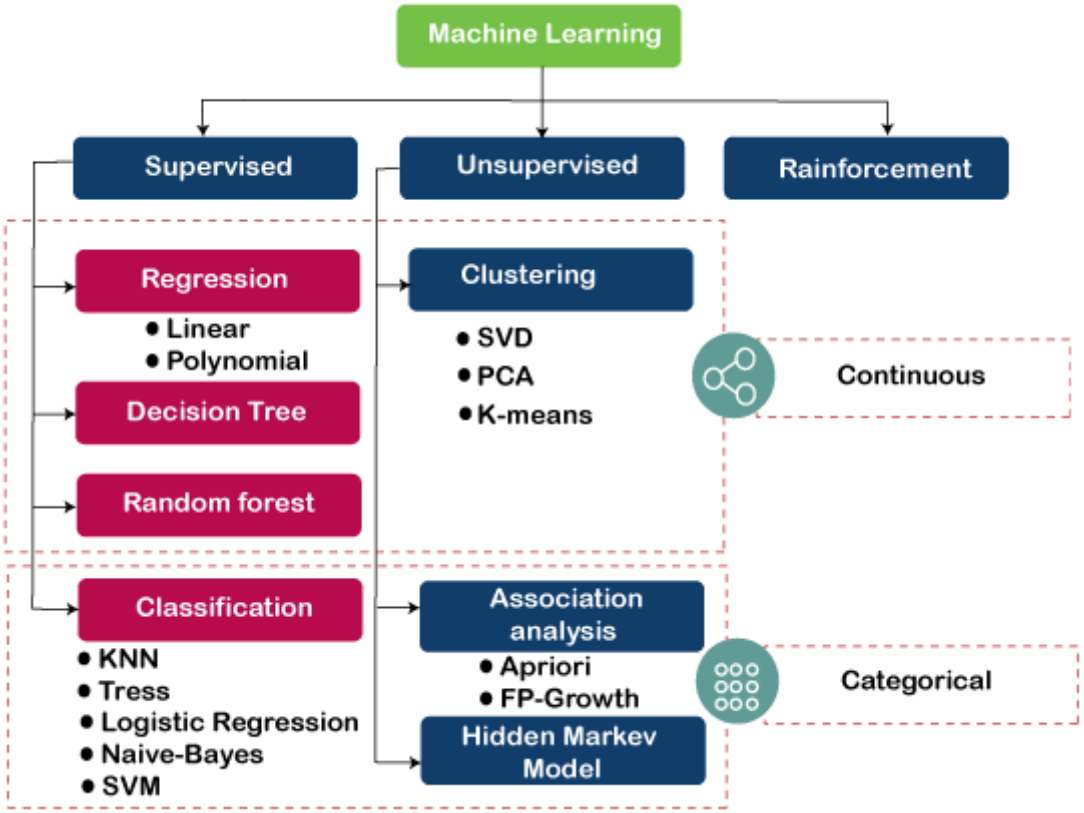
1. **Supervised Learning Algorithms**
2. **Unsupervised Learning Algorithms**
3. **Reinforcement Learning algorithm**

The below diagram illustrates the different ML algorithm, along with the categories:

38.3M

714

Java Try Catch



1) Supervised Learning Algorithm

Supervised learning is a sort of machine learning in which the machine learns with the help of a human. The labelled dataset is used to train the supervised learning models. After the model has been trained and processed, it is tested using sample test data to see if it predicts the correct outcome.

The purpose of supervised learning is to connect the input and output data. Supervised learning is based on supervision, and it is similar to when a student learns under the supervision of a teacher. Spam filtering is an example of supervised learning.

Supervised learning can be broken further into two sorts of problem:

o Classification

o Regression

Simple Linear Regression, Decision Tree, Logistic Regression, KNN algorithm, and other supervised learning algorithms are examples.

2) Unsupervised Learning Algorithm

o Unsupervised learning is a sort of machine learning in which the system learns from data without the requirement for external supervision. Unsupervised models can be trained with an unlabeled dataset that is neither classified nor categorised, and the algorithm must act on it without supervision. The model in unsupervised learning doesn't have a predefined output and instead tries to extract useful information from a large amount of data. These are used to solve the problems of association and clustering. As a result, it can be divided into two types:

o Clustering

o Association

Unsupervised learning algorithms include K-means Clustering, Apriori Algorithm, Eclat, and others.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114137

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. The agent does not receive any oversight. Reinforcement learning employs the Q-Learning algorithm.

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

1. Linear Regression

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:

$y= a_0+ a*x+ b$

Here, y= dependent variable

x= independent variable

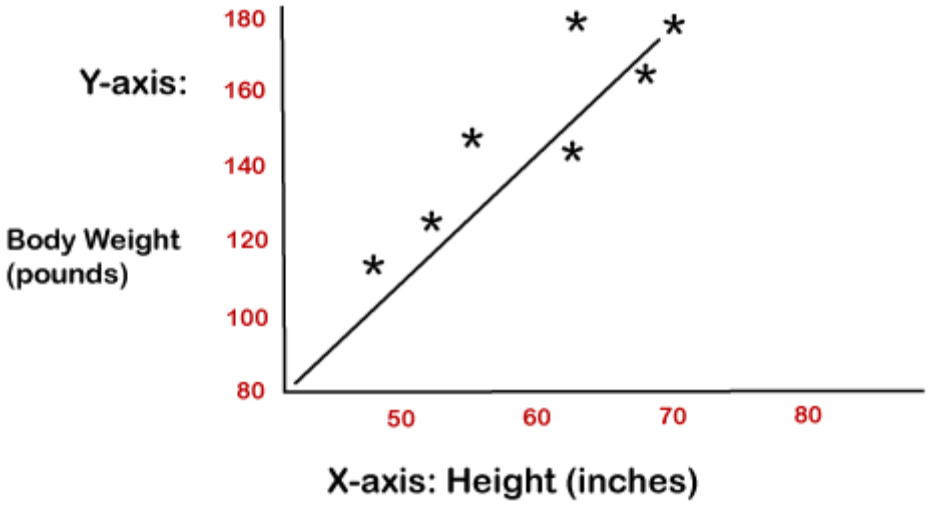
a_0 = Intercept of line.

There are two different types of linear regression:

o Simple Linear Regression: A single independent variable is utilised to predict the value of the dependent variable in simple linear regression.

o Multiple Linear Regression: Multiple independent variables are utilised to predict the value of the dependent variable in multiple linear regression.

The linear regression for weight prediction based on height is depicted in the picture below:



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114137

2. Logistic Regression

The supervised learning process of logistic regression is used to predict categorical variables or discrete values. It can be used in machine learning for classification issues, and the result of the logistic regression algorithm can be Yes or No, 0 or 1, Red or Blue, and so on.

The use of logistic regression differs from that of linear regression, for example. Logistic regression is used to solve the Classification problem and predict discrete values, whereas linear regression is used to solve the regression problem and predict continuous values. Instead of fitting the best fit line, it forms an S-shaped curve that lies between 0 and 1. The S-shaped curve is also known as a logistic function that uses the concept of the threshold. Any value above the threshold will tend to 1, and below the threshold will tend to 0.

3. Decision Tree Algorithm

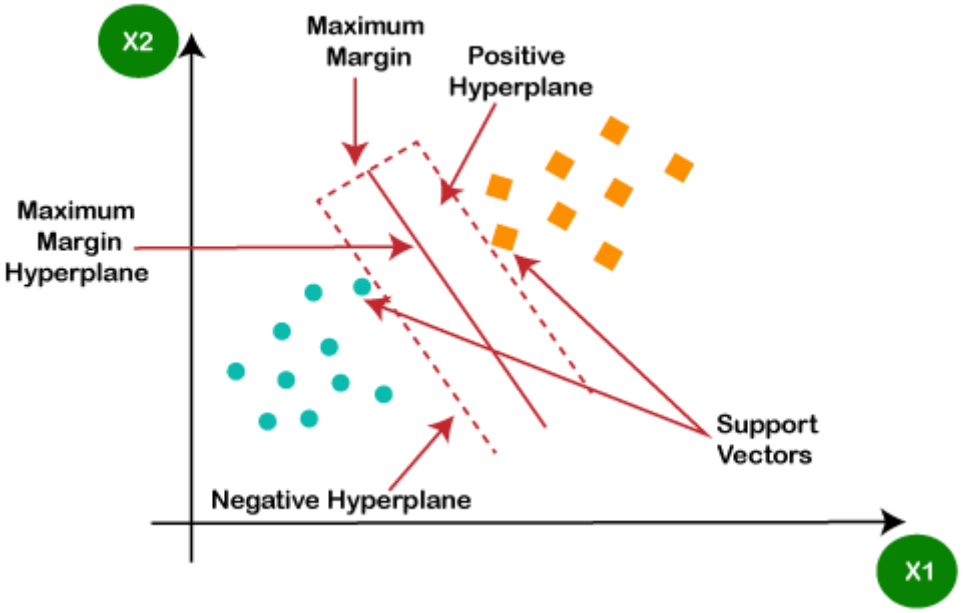
A decision tree is a supervised learning technique that can be used to tackle classification and regression problems. It is capable of working with both categorical and continuous variables. It depicts a tree-like structure with nodes and branches, beginning with the root node and expanding on subsequent branches until reaching the leaf node. The internal node represents the dataset's features, whereas branches indicate decision rules and leaf nodes represent the problem's conclusion.

Real-world uses of decision tree algorithms include identifying malignant and non-cancerous cells, making automobile buying recommendations, and so on.

4. Support Vector Machine Algorithm

A support vector machine, or SVM, is a supervised learning technique that can be used to solve issues like classification and regression. It is, however, mostly employed to solve categorization difficulties. The purpose of SVM is to generate a decision boundary or hyperplane that can divide datasets into multiple classes.

Support vectors are the data points that assist define the hyperplane, hence the algorithm is called support vector machine. Face detection, picture categorization, drug discovery, and other real-world uses of SVM Consider the diagram below:



The hyperplane has classified datasets into two classes, as seen in the diagram above.

5. Naïve Bayes Algorithm:

The Nave Bayes classifier is a supervised learning algorithm that makes predictions based on the object's likelihood. The algorithm is called Nave Bayes because it is based on the Bayes theorem and follows the naive assumption that variables are independent of one another.

The Bayes theorem is based on conditional probability, which is the probability that event(A) will occur after event(B) has already occurred. The Bayes theorem's equation is as follows:

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

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114137

The Nave Bayes classifier is one of the most effective classifiers for a given task. A naive bayesian model is simple to construct and is well suited to large datasets. It is mostly used to classify text.

6. K-Nearest Neighbour (KNN)

The K-Nearest Neighbour algorithm is a supervised learning algorithm that can be applied to classification and regression issues. This algorithm works by assuming that the new data point and existing data points are similar. The new data points are placed in the most similar groups based on their commonalities. It's also known as the lazy learner algorithm because it maintains all of the available datasets and uses K-neighbours to classify each new example. Any distance function measures the distance between the data points, and the new case is allocated to the closest class with the most similarities. Depending on the demand, the distance function can be Euclidean, Minkowski, Manhattan, or Hamming distance.

7. K-Means Clustering

K-means clustering is one of the most basic unsupervised learning algorithms for solving clustering issues. The datasets are divided into K separate clusters based on similarities and dissimilarities, which means that datasets with the most commonalities stay in one cluster while the other clusters have very few or no commonalities. K-means refers to the number of clusters, and means to the process of averaging the dataset to identify the centroid.

Each cluster is paired with a centroid in this centroid-based technique. Within a cluster, this technique seeks to reduce the distance between data points and their centroids. This approach begins with a set of randomly picked centroids that form clusters, and then uses an iterative procedure to optimise the placements of these centroids.

It may be used to detect and filter spam, as well as identify bogus news.

8. Random Forest Algorithm

Random forest is a supervised learning technique that can be utilised in machine learning for both classification and regression tasks. It's an ensemble learning strategy that combines numerous classifiers to generate predictions and improve the model's performance.

It has numerous decision trees for subsets of the given dataset and calculates the average to increase the model's forecast accuracy. There should be 64-128 trees in a random forest. The algorithm's accuracy improves as the number of trees increases.

To classify a new dataset or object, each tree provides a classification result, and the algorithm predicts the final output based on the majority votes.

Random forest is a quick method that can deal with missing and inaccurate data effectively.

9. Apriori Algorithm

The apriori algorithm is an unsupervised learning method for solving association problems. It is meant to work on databases that contain transactions and generates association rules using frequent itemsets. It establishes how firmly or weakly two objects are associated with the use of these association rules. To calculate the itemset efficiently, this approach uses a breadth-first search and a Hash Tree.

The programme searches the big dataset iteratively for the most frequent itemset.

In the year 1994, R. Agrawal and Srikant presented the apriori algorithm. It is mostly used for market basket analysis and assists in determining which products can be purchased together. It's also useful in the medical field Drug reactions in patients are studied in this subject.

10. Principal Component Analysis

The unsupervised learning approach Principle Component Analysis (PCA) is used to reduce dimensionality. It aids in lowering the dimensionality of a dataset with several features that are correlated with one another. It is a statistical procedure that uses orthogonal transformation to turn correlated feature observations into a set of linearly uncorrelated features. It's one of the most widely used tools for exploratory data analysis and predictive modelling.

PCA decreases dimensionality by considering the variance of each attribute because a high variance indicates a good separation between the classes. Image processing, movie recommendation systems, and optimising power allocation in multiple communication channels are examples of real-world PCA uses..

Some Real-World Applications of Machine Learning

- Automatic Language Translation in Google Translate
- Faster route selection in Google Map
- Driverless/Self-driving car
- Smartphone with face recognition
- Speech Recognition
- Ads Recommendation System

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022
DOI: 10.17148/IJARCCE.2022.114137

- Netflix Recommendation System
- Auto friend tagging suggestion in Facebook
- Stock market trading
- Fraud Detection
- Weather Prediction
- Medical Diagnosis
- Chatbot
- Machine Learning in Agriculture

Benefits of machine learning

- Work Automation
- Powerful predictive Ability
- Increased in sales in the e-commerce market
- ML benefits in the medical domain for enhancing medical diagnosis, drug development
- Machine Learning is used in robotic medical surgery
- ML in finance increases productivity enhances revenue and gives secure transactions.

CONCLUSION:

Machine Learning may be applied to practically every aspect of human existence to make our job more efficient, reliable, and simple. Machine learning, like everything else, has advantages and disadvantages. For example, when machine learning becomes more prevalent, many people may lose their current jobs. But, most importantly, it is good to humanity in the long run.

REFERENCES:

1. Li Kanghua, Jiang Shan. Machine Learning and Cultural Production Reform——Based on the Perspective of the Development of AI Technology [J]. Journal of Xiangtan University (Philosophy and Social Sciences), 2020, 44 (01): 74-79.
2. Jiang Na, Yang Haiyan, Gu Qingchuan, Huang Jiya. Machine learning and its algorithm and development analysis [J]. Information and Computer Science (Theoretical Edition), 2019 (01): 83-84 + 87.
3. Li Zhiwei. Development of machine learning and several learning methods [J]. Industry and Science Forum, 2018, 15 (10): 198-199.
4. Zhang Run, Wang Yongbin. Research on machine learning and its algorithm and development [J]. Journal of Communication University of China (Natural Science Edition), 2018, 23 (02): 10-18 + 24.
5. Zhang Changshui. Research on the development of machine learning and data mining [C]. 2010-2011 Development Report of Control Science and Engineering Discipline.: Chinese Society of Automation, 2018: 82-89 + 223.

 This work is licensed under a Creative Commons Attribution 4.0 International License