An Overview of Machine Learning Algorithms, Tasks, and Applications in Brief.

By

Ravi Bhushan[1]
Ph.D. Scholar
School of Computer Science and Engineering.
Galgotias University, Gr.Noida

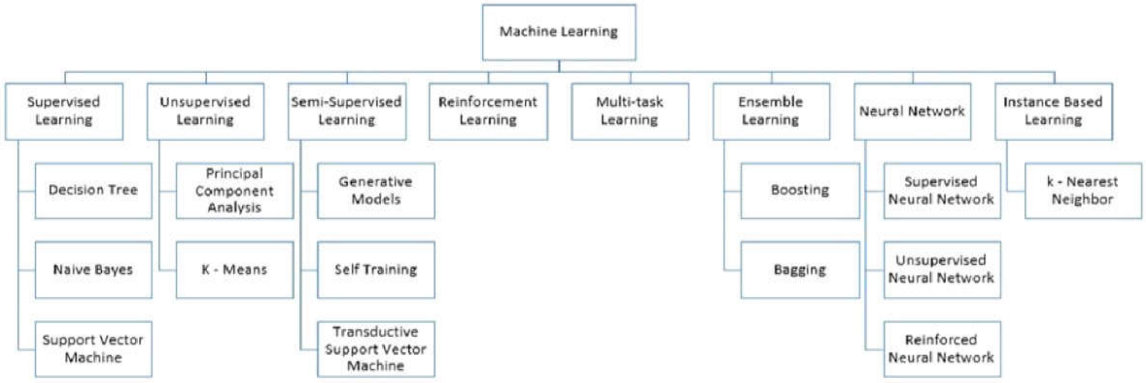
Dr. vineeta khemchandani[2]
Professor
School of Computer applications and Technology Galgotias University, Gr.Noida

Abstract: The scientific study of statistical models and techniques that computer systems employ to do certain tasks without explicit programming is called machine learning (ML). gaining an understanding of algorithms in a number of applications that we regularly use. A learning algorithm that has found out how to rank websites is one of the reasons web search engines like Google operate so well. The main advantage of machine learning is that once algorithms have mastered data handling, they can do jobs autonomously. A succinct summary and outlook on the many applications of machine learning techniques are given in this paper. A brief overview and outlook on the several uses of machine learning algorithms have been provided in this study.

Keywords: Reinforcement learning, Supervised learning, Unsupervised learning, Machine learning, and Algorithms

Introduction

Since the dawn of time, humans have employed a vast array of tools to better accomplish a variety of tasks. The human brain's ingenuity led to the creation of numerous machines. Human existence was made easier by these devices, which enabled people to fulfil a range of needs, including industry, transport, and computing. Additionally, machine learning is one of them. Machine learning is the science of teaching computers to learn without explicit programming, according to Arthur Samuel. For his checkers-playing program, Arthur Samuel gained notoriety. Machine learning (ML) is the process of teaching machines to handle data more effectively. We may find ourselves unable to decipher the data after examining it. When that occurs, we use machine learning. Because there are so many datasets available, there is a growing need for machine learning. To extract pertinent data, machine learning is used in many sectors. Using data to learn is the goal of machine learning. A lot of research has been done on how to teach robots to learn without explicit programming. Numerous programmers and mathematicians use a variety of techniques to solve this problem, which involves large data sets.



Machine learning uses a variety of techniques to address data issues. Data scientists want to emphasise that there isn't a single, universally applicable method that works well for every situation. The type of method used is determined by the type of problem you want to answer, the amount of variables, the model that would work best for it, and other factors. A brief overview of some of the most popular machine learning (ML) algorithms is provided here.

Supervised Learning

The machine learning task of learning a function that converts an input to an output using sample input-output pairs is known as supervised learning. It uses a collection of training examples and labelled training data to infer a function. The machine learning algorithms that require outside help are known as

supervised algorithms. There are train and test datasets within the input dataset. The output variable in the train dataset has to be categorised or forecasted. All algorithms to make predictions or classify the test dataset, find patterns in the training dataset. The supervised machine learning algorithm workflow is depicted in the figure below. This article has covered the most well-known supervised machine learning methods.

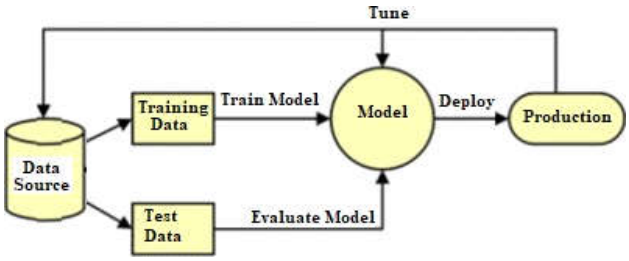


Figure: Supervised learning Workflow

Decision Tree

A decision tree is a graph that shows decisions and the outcomes of those choices as a tree. The graph's edges stand in for the conditions or decision rules, while the nodes represent an event or option. There are nodes and branches in every tree. Every node represents a group's properties that need to be categorised, and every branch represents a value that the node is capable of accepting.

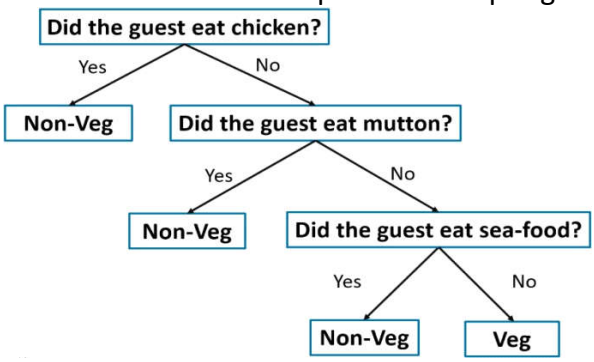


Figure: Decision Tree

Navie Bayes

The Bayes Theorem serves as the foundation for this classification method, which assumes predictor independence. To put it simply, a Naive Bayes classifier posits that the existence of a certain characteristic inside a class is independent of the existence of any other feature. The primary market for Naïve Bayes is the text categorisation sector. Its primary applications are in clustering and classification, which are based on the conditional probability of occurrence.

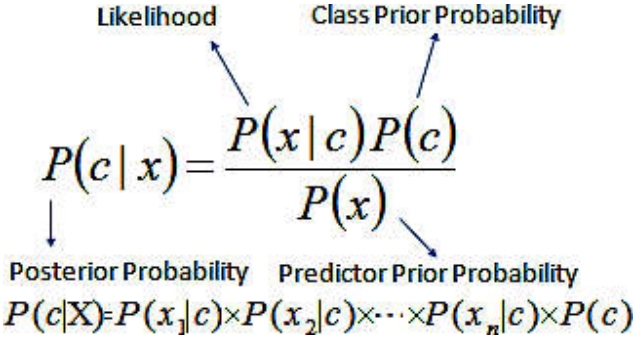


Figure: Navie Bayes

Support Vector Machine

The most popular cutting-edge machine learning method is Support Vector Machine (SVM). Support-vector machines are supervised learning models with related learning algorithms that evaluate data for regression and classification in machine learning. The kernel approach, which involves implicitly translating inputs into high-dimensional feature spaces, allows SVMs to effectively conduct non-linear classification in addition to linear classification. The primary idea is to construct boundaries between the classes. In order to minimise

the classification error, the margins are drawn so that the distance between the margin and the classes is as large as possible.

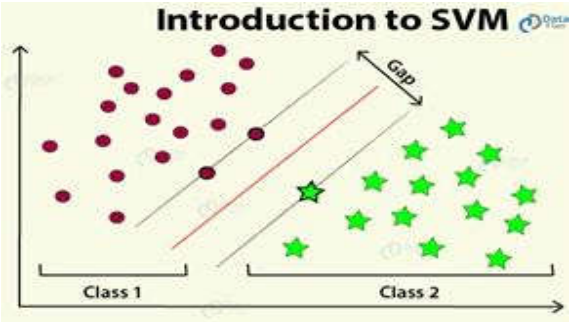


Figure: Support Vector Machine

Unsupervised Learning:

Because there are no right answers or teachers, unlike supervised learning mentioned above, these are referred to as unsupervised learning. To find and display the intriguing structure in the data, algorithms are left to their own devices. From the data, the unsupervised learning algorithms pick up only a few features. When fresh data is presented, it recognises the data's class using the previously learnt features. Feature reduction and clustering are its primary applications.

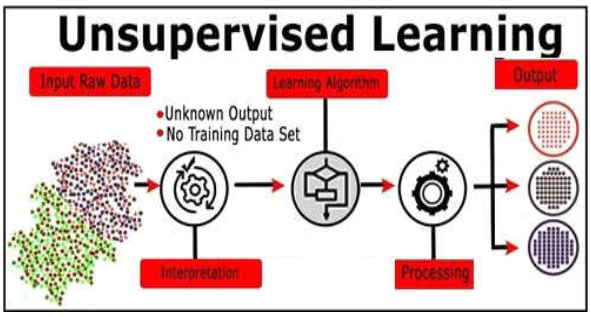


Figure: Unsupervised Learning

Principal Component Analysis

principle component analysis is a statistical technique that creates principle components—a collection of values of linearly uncorrelated variables—from a set of observations of potentially correlated variables using an orthogonal transformation. In order to speed up and simplify the computations, the data's dimension is decreased. It uses linear combinations to explain the variance-covariance structure of a set of variables. It is frequently applied as a method of dimensionality reduction.

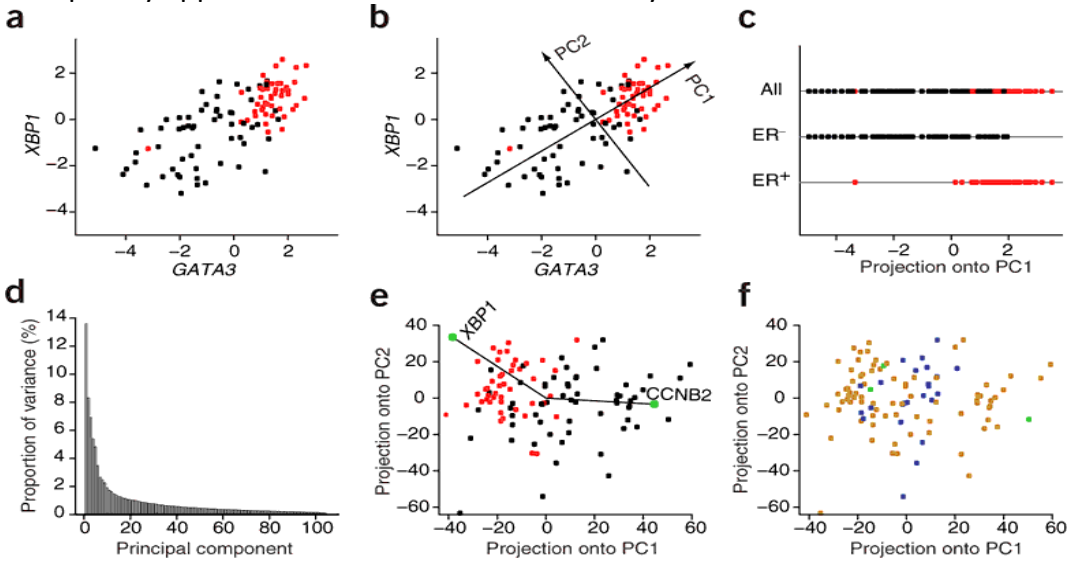


Figure: Principal Component Analysis

K-Means Clustering

The well-known clustering problem can be solved by one of the most straightforward unsupervised learning techniques, K-means. The process uses a straightforward method to group a given data set into a predetermined number of clusters. Define k centres, one for each cluster, as the primary idea. It is important to strategically situate these centres since different locations yield varied outcomes. Therefore, it is preferable to position them as far apart as feasible.

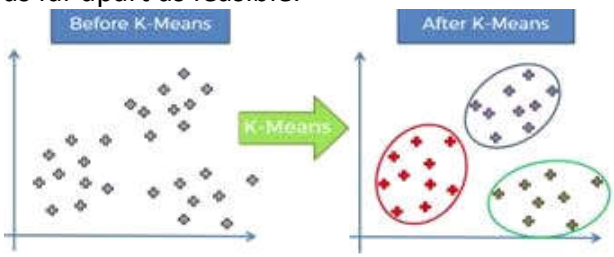


Figure: K-Means Clustering

Semi Supervise Learning:

Unsupervised and supervised machine learning techniques are combined to create semi-supervised machine learning. In fields like data mining and machine learning when unlabelled data is already available and obtaining labelled data is a laborious procedure, it may be beneficial. A machine learning algorithm is trained on a "labelled" dataset, where each record contains the result information, in increasingly popular supervised machine learning techniques. The following section discusses a few semi-supervised learning algorithms.

Transductive SVM

Semisupervised learning has made extensive use of transductive support vector machines (TSVM) to handle partially labelled input. Because its basis in generalisation is not well understood, there has been mystery about it. It is employed to label the unlabelled data so that the margin between the labelled and unlabelled data is as small as possible. Using TSVM to find an exact answer is an NP-hard task.

Generative Models

A model that can produce data is called a generative model. Both the class (i.e., the entire data) and the features are modelled. Since I can create data points using this probability distribution if we model P(x,y), all algorithms that model P(x,y) are generative. For each component, one labelled example is sufficient to verify the distribution of the mixture.

Self-Training

Self-training involves using a subset of labelled data to train a classifier. Unlabelled data is then supplied into the classifier. In the training set, the unlabelled points and the anticipated labels are totalled. After that, same process is carried out once again. The term "self-training" comes from the fact that the classifier is learning itself.

Reinforcement Learning

The study of how software agents should behave in a given environment to maximise a concept of cumulative reward is known as reinforcement learning. Among the three fundamental machine learning paradigms are reinforcement learning, supervised learning, and unsupervised learning.

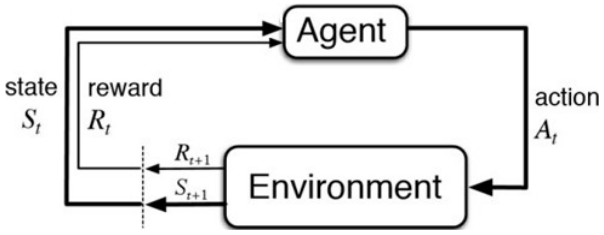


Figure: Reinforcement Learning

Multitask Learning

A kind of machine learning called "multi-task learning" seeks to solve several tasks simultaneously by utilising the similarities between them. This can serve as a regularizer and increase the effectiveness of learning. In a formal sense, Multi-Task Learning (MTL) will improve the learning of a particular model by utilising the knowledge contained in all n tasks. This is because traditional deep learning approaches aim to solve only one task using a single model. This is the case if there are n tasks, or a subset of them, that are related to each other but not exactly the same.

Ensemble Learning

Ensemble learning is the act of intentionally generating and combining several models, like classifiers or experts, to address a specific computational intelligence problem. Ensemble learning is mostly used to enhance a model's performance or lessen the possibility of a bad model being selected by accident. Choosing the best features, data fusion, incremental learning, non-stationary learning, error-correcting, and giving a confidence level to the model's choice are some other uses for ensemble learning.

Boosting:

"Boosting" describes a class of algorithms that transform weak learners into strong ones. Boosting is an ensemble learning strategy that reduces variance and bias. The foundation of boosting is the question, "Can a set of weak learners create a single strong learner?" put out by Kearns and Valiant. A classifier that has an arbitrary high correlation with the true classification is considered a strong learner, while a weak learner is a classifier.

Bagging

In situations where improving a machine learning algorithm's accuracy and stability is necessary, bagging or bootstrap aggregating is used. Both regression and classification can use it. Additionally, bagging helps manage overfitting and reduces variance.

Neural Networks

In order to identify underlying links in a set of data, a neural network is a collection of algorithms that simulates how the human brain functions. Neural networks, in this context, are networks of neurones, whether they be artificial or natural. Because they can adjust to changing input, neural networks produce optimal results without requiring a change of the output criteria. As trading systems are developed, the idea of neural networks—which has its origins in artificial intelligence—is rapidly gaining traction.

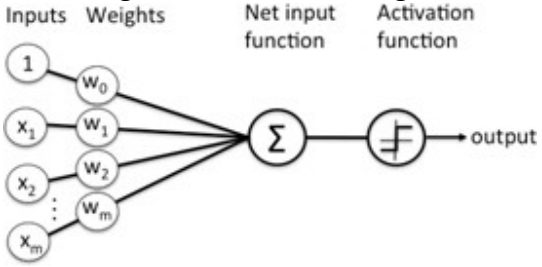


Figure: Neural Networks

In the same way, an artificial neural network acts. It operates on three levels. The input layer receives input. Within the concealed layer, the input is processed. The calculated output is finally sent via the output layer.

Supervised Neural Network

In a supervised neural network, the input's output is predetermined. The neural network's actual output and its anticipated output are contrasted. The parameters are adjusted and then re-fed into the neural network based on the error. In feed forward neural networks, supervised neural networks are employed.

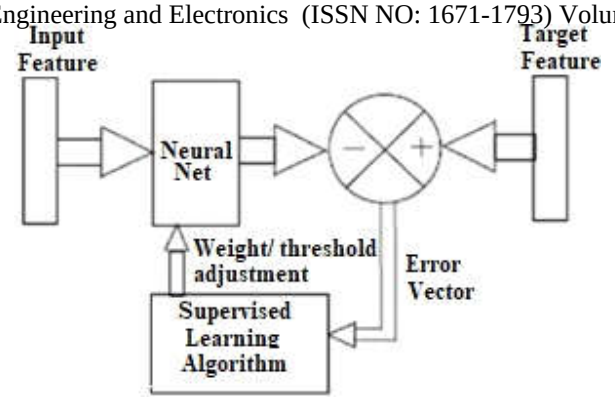


Figure: Supervised Neural Network

Unsupervised Neural Network

The input and output are unknown to the neural network beforehand. Classifying data based on similarities is the network's primary function. After grouping different inputs, the neural network determines if they are correlated.

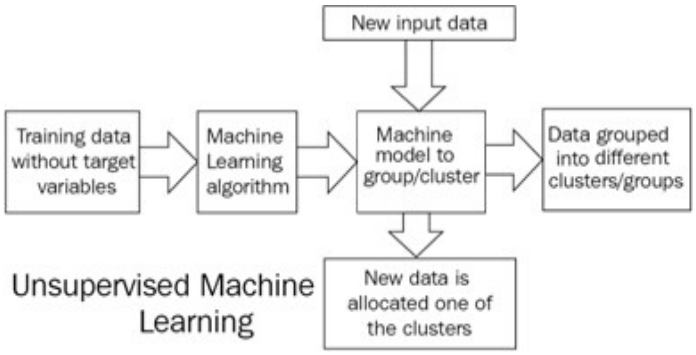


Figure: Unsupervised Neural Network

Reinforced Neural Network

Reinforcement learning is the term used to describe goal-oriented algorithms that learn how to maximise along a specific dimension over a large number of steps, for as maximising the number of points earned in a game over a large number of moves, or how to achieve a difficult aim (goal). They can begin with nothing and, given the correct circumstances, perform at superhuman levels. These algorithms are reinforced when they make the correct choices and punished when they make the wrong ones, much like a toddler who is rewarded with candy and spankings.

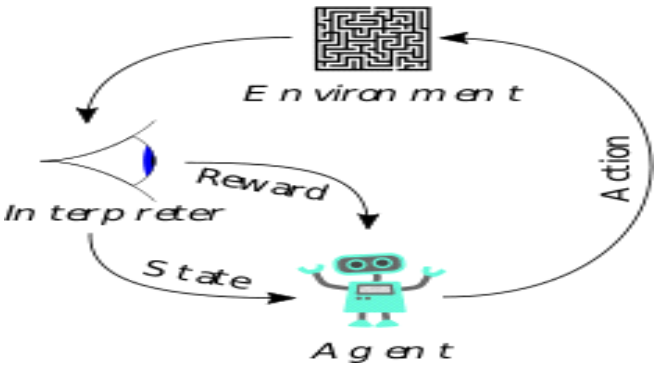


Figure: Reinforced Neural Network

Instance-Based Learning

A class label or prediction is generated via instance-based learning, a family of classification and regression techniques, depending on how similar the query is to its nearest neighbour or neighbours in the training set. Instance-based learning algorithms explicitly differ from other approaches like decision trees and neural networks in that they do not generate an abstraction from individual instances. Instead, they merely store all the information and obtain an answer at query time by analysing the query's nearest neighbour (s).

K-Nearest Neighbor

One easy supervised machine learning method for handling regression and classification issues is the k-nearest neighbours (KNN) algorithm. It is simple to use and comprehend, but it has a key flaw in that it becomes more slower as the amount of data being used increases.

Machine Learning Application and Task.

Statistical arbitrage, learning associations, classification, prediction, medical diagnosis, digital image processing (image recognition) [5], large data analysis [4], speech recognition, and many other applications employ machine learning techniques.

Generally, machine learning tasks fall into one of three major groups based on the type of "signal" or "feedback" that a learning system can use. Reinforcement learning, unsupervised learning, and supervised learning are these.

Conclusion

One can choose between supervised and unsupervised machine learning. Supervised learning is the better option if you have fewer data and clearly marked training data. For huge data sets, unsupervised learning will typically perform better and produce better outcomes. Choose deep learning methods if you have a large data collection at your disposal. Additionally, you have studied Deep Reinforcement Learning and Reinforcement Learning. Neural networks, their uses, and their limits are now clear to you. This study examines a number of machine learning algorithms. Nowadays, whether they realise it or not, everyone uses machine learning. using social networking sites to update images or receiving product recommendations when buying online. An overview of the majority of well-known machine learning algorithms is provided in this publication.

References

[1] W.Richert, L. P. Coelho, “Building Machine Learning Systems with Python”, Packt Publishing Ltd., ISBN 978-1-78216-140-0

[2] J. M. Keller, M. R. Gray, J. A. Givens Jr., “A Fuzzy K- Nearest Neighbor Algorithm”, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4,August 1985 "http://www.geeksforgeeks.org/machine-learning/.

[3] S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.M. Bkassiny, Y. Li, and S. K. Jayaweera, “A survey on machine learning techniques in cognitive radios,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3,pp. 1136–1159, Oct. 2012. https://en.wikipedia.org/wiki/Instance-based_learning

[4] R.S. Sutton, “Introduction: The Challenge of Reinforcement Learning”, Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992

[5] P. Harrington, “Machine Learning in action”, Manning Publications Co., Shelter Island, New York, 2012 Sharma, D., Pabby, G. and Kumar, N., Challenges Involved in Big Data Processing & Methods to Solve Big Data Processing Problems.IJRASET,5(8),pp.841- 844.

[6] Kumar, N. and Gupta, S., 2016. Offline Handwritten Gurmukhi Character Recognition: A Review. International Journal of Software Engineering and Its Applications, 10(5), pp.77-86.

[7] Talwar, A. and Kumar, Y., 2013. Machine Learning: An artificial intelligence methodology. International Journal of Engineering and Computer Science, 2, pp.3400-3404.

[8] Muhammad, I. and Yan, Z., 2015. Supervised Machine Learning Approaches: A Survey. ICTACT Journal on Soft Computing, 5(3).

[9] Singh, S., Kumar, N. and Kaur, N., 2014. Design Anddevelopment Of Rfid Based Intelligen SecuritySystem. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume, 3.

[10] D. Pelleg, A. Moore (2000): "X-means: Extending K-means with Efficient Estimation of the Number of Clusters"; ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning, pp. 727-734.

[11] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R , "Prediction of Crop Yield using Machine Learning", International Research Journal of Engineering & Technology, Vol 5, Issue 2, Feb-2018.

[12] C. Phua, V. Lee, K. Smith, R. Gayler (2010); "Comprehensive Survey of Data Mining-based Fraud Detection Research", ICICTA '10 Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation Volume 1, pp. 50-53.