MACHINE LEARNING ALGORITHMS: OVERVIEW

Article · September 2020

DOI: 10.34218/IJARET.11.9.2020.052

1 author:

Vanshika Rastogi
East Point College of Engineering and Technology
5 PUBLICATIONS **25** CITATIONS

International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 9, September 2020, pp. 512-517, Article ID: IJARET_11_09_052
Available online at http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=9
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.9.2020.052

MACHINE LEARNING ALGORITHMS: OVERVIEW

Vanshika Rastogi
Assistant Professor, ABES Engineering College, Ghaziabad, India

Sugandha Satija
Senior Assistant Professor, ABES Engineering College, Ghaziabad, India

Dr. Pankaj Kumar Sharma
Professor, ABES Engineering College, Ghaziabad, India

Sanika Singh
Assistant Professor, ABES Engineering College, Ghaziabad, India

ABSTRACT

This paper aims to cover the different machine learning algorithms. These algorithms can be used in the different fields of data mining, image processing, predictive analysis and many more.

Key words: Machine learning, algorithms, analysis

Cite this Article: Vanshika Rastogi, Sugandha Satija, Pankaj Kumar Sharma and Sanika Singh, Machine Learning Algorithms: Overview, *International Journal of Advanced Research in Engineering and Technology,* 11(9), 2020, pp. 512-517.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=9

1. INTRODUCTION

Machine Learning is a method to make the computer learn making some predictions with information being provided to it. These predictions can be identifying whether a piece of any fruit in the picture is a watermelon or an apricot, spotting people parking their vehicle, whether the use of the word *kite in a* sentence relates to a bird or a paper, categorize an email as promotional or spam. With the increase in the availability of the data, the demand for the machine learning is also increasing.

The goal of machine learning is to master through the existing information. Multiple researches have been conducted on the various ways to help machines learn on their own [1] [2]. Many programmers and mathematicians have addressed various methods and techniques to realize the solution to this problem. Few of them are shown in Fig. 1.

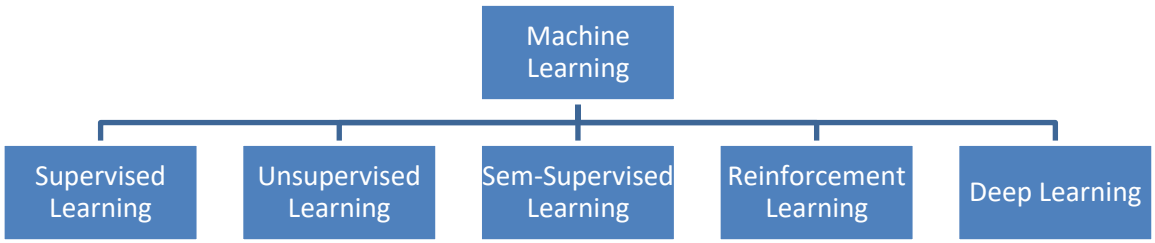

Figure 1 Different Types of Learning

2. SUPERVISED LEARNING

Supervised learning algorithms are developed to grasp by example. The term "supervised" learning arises from the concept that training this kind of algorithm is like having a teacher supervising the complete process. While instructing a supervised algorithm, the data to be trained comprises of inputs coupled with the correct outputs. While training, the algorithm looks for patterns in the provided data that matches with the required outputs. Once the training is done, the algorithm can take in fresh inputs and learn which label the fresh inputs will be classified as depending on the previous training data. All algorithms get trained for some patterns through the trained dataset and try it on the test dataset for the prediction or classification [3]. Few of the most famous algorithms are discussed here.

2.1. Decision Tree

Decision tree is similar to a tree structure where the internal nodes represents some test on an attribute, the leaf node shows the label of the class and branches signifies the coexistence of the attributes that lead to the labelling of the class. Decision tree is used primarily for the objective of performing classification. One of the examples of the decision tree is shown in fig.3. Decision trees perform the classification of the data by sorting them in a top-down approach starting from the root node to the leaf node, where the leaf node provides the label to the class.

2.2. Naïve Bayes

This algorithm makes use of the Bayes Theorem and assumes that features are statistically independent. The theorem makes an assumption that the variables are independent of each other and still proves itself to be a good classifier. Naïve Bayes focuses on the classification of the text and is primarily used for performing clustering and classification [4]. It is a probabilistic classifier and makes use of the probability to classify a given object. The formula for the Bayes Theorem is

$$P(L|M) = (P(M|L)P(L))/P(M)$$

P (L|M) signifies the Posterior Probability, the probability of hypothesis L over the occurred event M.

P (M|L) is the probability of likelihood, Probability of the evidence given that the probability of a hypothesis is true.

P (L) signifies the probability of hypothesis before the evidence is observed.

P (M) signifies the probability of evidence.

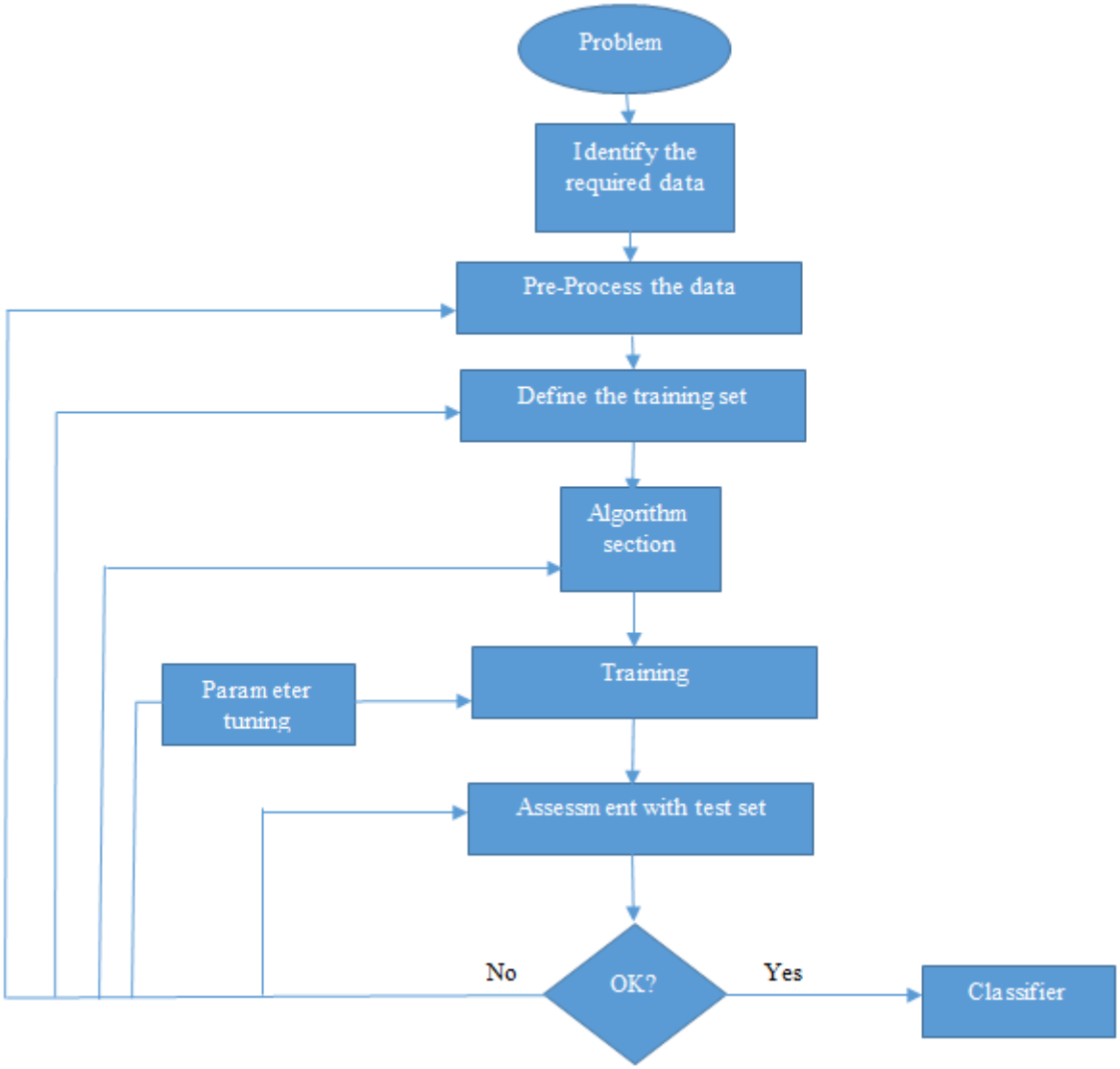

Figure 2 Workflow of Supervised Learning Algorithm

2.3. Support Vector Machine

Another supervised learning algorithm whose primary objective is the classification problem of two-group. The primary purpose of the Support Vector Machine is to find a hyper plane in an n-dimensional space that helps in doing the classification of the data points. The main goal is to search for a hyper plane having the maximum margin. Hyper plane serves as a decision boundary that is used for the classifying the data points. The working of the Support Vector Machine is shown in fig.3.

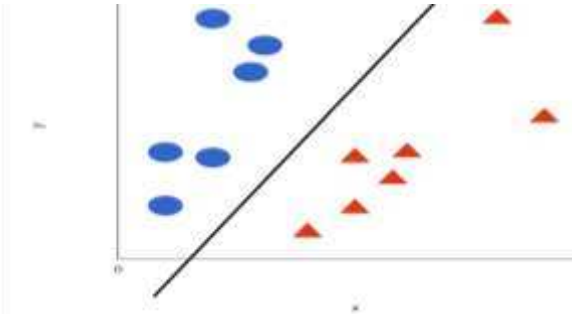

Figure 3 The Support Vector Machine

3. UNSUPERVISED LEARNING

In Unsupervised learning, there are no classes given to the learning algorithm, leaving it to find the pattern or structure in the input. This kind of learning algorithm does not require the user to supervise the algorithm. More complex processing tasks can be performed using unsupervised learning algorithm as compared to the supervised learning algorithms. The primary objective to use this type of learning is to cluster and feature reduction. The work flow of the learning is shown in fig. 4. Few of the important clustering and dimensionality reduction algorithms have been discussed here.

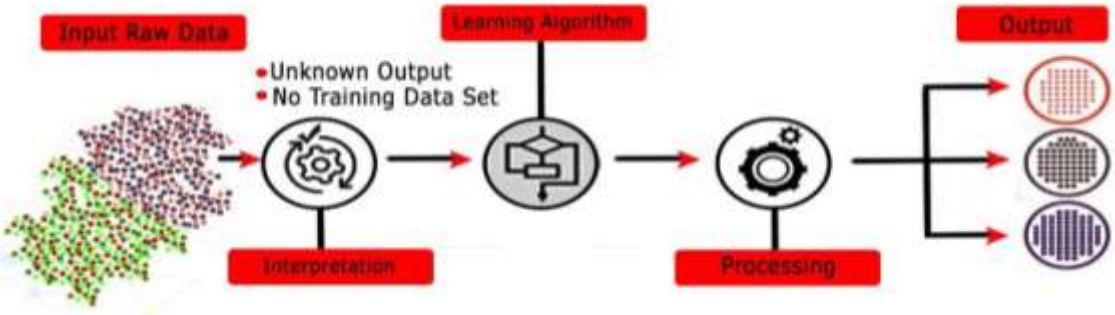

Figure 4 Workflow of Unsupervised Learning Algorithms

3.1. k-means Clustering

The process of grouping the data together based on some similarities or properties is called clustering. The k-means clustering algorithm generates k distinct clusters. To evaluate the center of the cluster the mean of the values of the cluster is used [5]. The aim of the K-Means algorithm is to have the minimum sum of distances between the given points and the respective cluster centroid. The clustered data appears as shown in Fig. 5.

3.2. Principal Component Analysis

Using the Principal Component Analysis, we focus to minimize the dimension of the given data that makes the computation fast and easy. This algorithm preserves the important parts that have more variation of the data and ignore the parts with fewer variations.

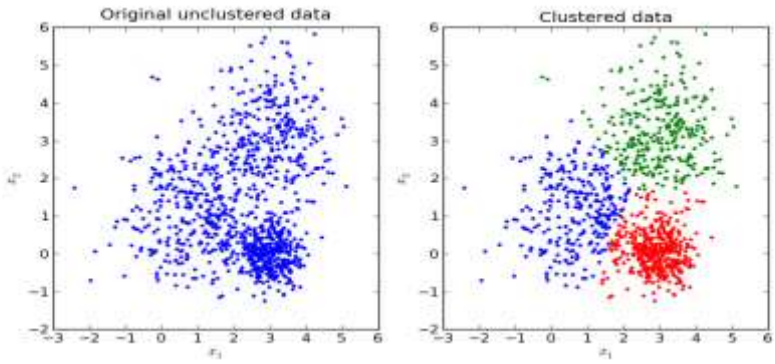

Figure 5 Clustered data using k-means clustering

4. SEMI-SUPERVISED LEARNING

In this kind of learning, the algorithm is instructed over a blend of labelled and unlabelled data. This algorithm is useful in the areas of data mining and machine learning where the unlabelled data is available and to get the labelled data is a difficult task [6].

4.1. Generative Models

Generative models are amongst the classic semi-supervised learning algorithms. It considers a structure like p(l,m) = p(m)p(l|m), where p(l|m) is a mixed distribution. Among the unlabelled data, the mixed components can be identified.

4.2. Self-Training

In this learning algorithm, the classifier gets trained with a content of labelled data. Then some unlabelled data is given to the classifier. The unlabelled data points along with the classified labels are summed up together in the trained set. This process is continuously repeated. As the classifier tries to learn on its own, thus named as self-training.

4.3. Transductive SVM

Transductive support vector machine is an extended version of the SVM. In TSVM, labelled and unlabelled data both are important. It provides the label for the unlabelled data in a way that labelled and unlabelled data have the maximum margin between them.

5. REINFORCEMENT LEARNING

Reinforcement learning is a Machine Learning algorithm that permits the software agents and machines to evaluate the perfect behaviour for a specific context, in order to get the best outcome. Reinforcement Learning is a subset of unsupervised learning and uses the approach of *cause and effect.* Fig.6 shows the working of reinforcement learning.

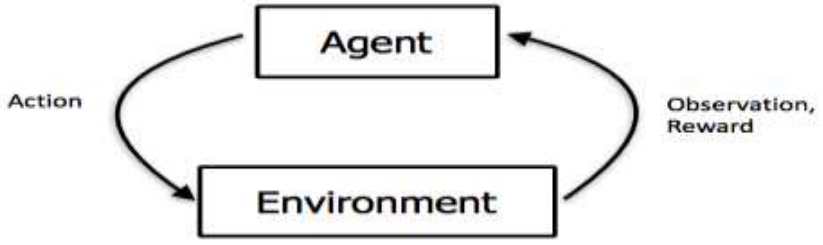

Figure 6 Working of Reinforcement Learning

6. DEEP LEARNING

Deep Learning is an application of Artificial Intelligence that imitates the working of a human brain in order to process the data and take decisions accordingly. This includes the use of statistics and predictive modelling. The advantage of using deep learning is that the program develops the feature set on its own without any supervision.

7. CONCLUSION

This paper tries to cover the different algorithms in the field of machine learning. Today every person is making use of machine learning in some or the other way. This paper gives a brief about the various machine learning algorithms present today.

REFERENCES

[1] M. Welling, "A First Encounter with Machine Learning"

[2] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118- 96174-2

[3] S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249-268

[4] D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation"

[5] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub - Gradient Solver for SVM", Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, 2007.

[6] X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, Vol. 3, No. 1, Pages 1-130

[7] P. Harrington, "Machine Learning in action", Manning Publications Co., Shelter Island, New York, 2012

[8] V. Sharma, S. Rai, A. Dev, "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10, October 2012

[9] S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014

[10] Caruana, R., and Niculescu-Mizil, A., "An empirical comparison of supervised learning algorithms", In Proc. of Int. Conf. on Machine learning, ACM, 2006.

[11] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E., "Machine learning: a review of classification and combining techniques", Artificial Intelligence Review, 26(3), 2006

[12] Das, S., Dey, A., Pal, A. and Roy, N., 2015. "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", International Journal of Computer Applications, 115(9).