A REVIEW STUDY ON VARIOUS ALGORITHMS OF MACHINE LEARNING

[1]Himanshi Bansal, [2]Kapil Sharma
[1]Master of Engineering Student, [2]Assistant Professor
[1]Computer Science,
[1]Guru Nanak Dev Engineering College, Ludhiana, India.

Abstract : Machine Learning is a phenomenon that helps the processor or system to learn from past experiences, train, and practice, without being programmed directly. Rather than writing the whole code, what you do is feed data into the generic algorithm and the algorithm/machine generates the logic, centered on the data supplied. This study has been undertaken to find a brief introduction to the different types of learning you may experience in the machine learning area. Each type of learning provides many algorithms which are suitable for carrying out many processes. The major advantage of using machine learning is that if an algorithm knows what to do with input, it will automatically do its job and generate the outcomes.

IndexTerms **– Machine Learning, Clustering, Classification, Semi-supervised, Classifiers, Ensemble.**

I. INTRODUCTION

Machine learning is a sub-field of artificial intelligence that remains its focus on system designing and allows them to acquire and make some predictions on the basis of data understandings and experiences. Machine learning enables a machine to act and make data driven conclusions rather than being plainly programmed to carry out a particular job. These programs are intended to learn and advance over time when exposed to new data inputs. The idea behind this is that machines should be given full access to data, so they can learn from themselves. It works to extract patterns from the dataset. Meaning machine can not only find the rules for optimum behavior but also can adjust to the variations in the world. Many of the algorithms involved have been identified that can scale immense data sizes. Some basic types of learning are demonstrated in Fig.1 below.

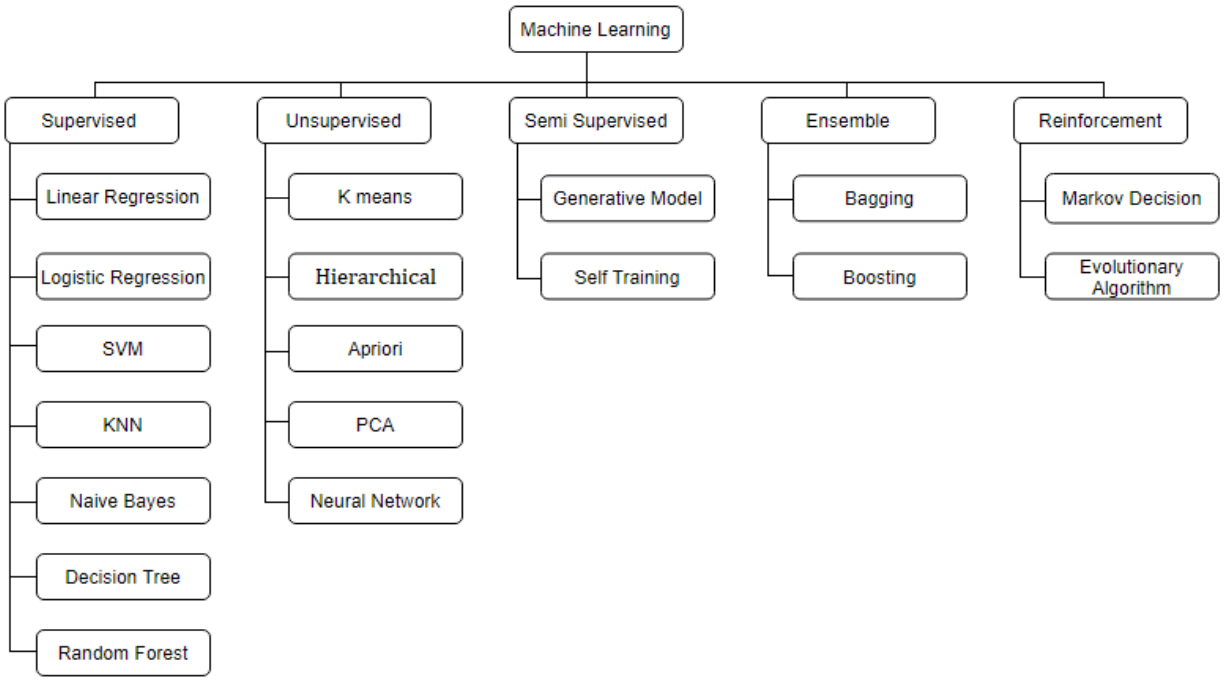

Fig.1 Types of Learning

II. TYPES OF LEARNING

2.1 Supervised Learning

Supervised learning is the learning type which involves a mapping function. This function is understood by an algorithm which maps from the input to the output. The main aim here is mapping function approximation such that whenever you have a new input data you will be able to predict the output variable. Each instance of a training data set is composed of different input attribute and an expected output. The input attribute of a training data set can be of any kind of data like pixel value, histogram of audio frequency, video file etc. For each input instance, an expected output values are associated. Value can be discreet representing a category or real or continuous value. Once the algorithm is strained, it can be used to predict the correct output of a never seen input. The algorithms which follow supervised learning as explained below:

2.1.1 Linear Regression

Linear Regression is a supervised method for machine learning in which the expected output has a steady slope and is continuous. This is used to estimate values within a continuous range instead of classifying them into groups. It is a regression model assuming a continuous relationship between the variables of input and the variables of single output. We develop a relationship between independent and dependent variables by selecting the best fitted line. It can be of two types:

2.1.1.1. Simple Linear Regression

It involves one independent variable and uses traditional slope-intercept form.

$$y = mx + c \qquad (2.1)$$

where m and c are variables that the algorithm learns to produce most accurate predictions, x represents the input data and y represents our prediction.

2.1.1.2. Multiple Linear Regression

It involves more than one independent variable. It looks like a function equation.

$$f(x, y, z) = w_1x + w_2y + w_3z \qquad (2.2)$$

where w represents the coefficients that our model will learn and x, y, z represents attributes we have about each observation.

2.1.2 Logistic Regression

Logistic regression is a method used for assigning results to a specific collection of classes. Logistic regression integrates the performance to produce a probable value that can then be assigned to two or more distinct classes using the logistic sigmoid function. It estimates the probability of an occurrence appearing by fitting the data to a logit function. Since it involves probability, the value of output lies between 0 and 1. It is of three types:

2.1.2.1. Binomial Logistic Regression

The output value may only have two valid types- 0 or 1 which indicate profit/loss, pass/fail, dead/alive, win/lose etc.

2.1.2.2. Multinomial Logistic Regression

The output value may have three or more valid types which are not supervised i.e they do not have abstract value like kind a, b or c.

2.1.2.3. Ordinal Logistic Regression

It deals with output value with ordered categories like a test score can be categorized to the values good, very good, excellent. These categories can be assigned the values 0, 1, 2.

2.1.3 Support Vector Machine

Support vector machine is amongst the most common supervised learning algorithm used for classification and problems with regression. For machine learning, it can be used for classification. The algorithms aim is to build the best line or judgement boundary that can efficiently put the new space into groups such that in the future, we can efficiently put the new piece of data in the right group. This optimal boundary for judgement is called a hyperplane. SVM selects vectors which helps to construct a hyperplane. It is classified into two types as follows:

2.1.3.1. Linear SVM

It can be used for linearly separate and distinct data implying that if a dataset can be divided into two groups with a single straight line, then it is considered linearly separable data and classifier used is Linear SVM.

2.1.3.2. Non-Linear SVM

It can be used for non-linear data which implies that if a sample cannot be categorized by a straight line, the data is considered non-linear data and non-linear SVM is used.

2.1.4 K-Nearest Neighbor

K-Nearest Neighbor is a basic algorithm which stores every eligible case and identifies new cases based on a measure of similarity like distance function. KNN is a process that defines data points depending on the points that most closely match it. This uses test results to create an informed choice of what is to be categorized as an unclassified object. It stores all existing cases and classifies new cases by means of a majority vote by its k neighbors. It refers to labelled input data to learn a feature which when provided new unlabeled data, produces an acceptable output. KNN predicts new instance by checking the most related instances across the entire training set and analyses the performance vector for such k instances(Raj, 2019).

2.1.5 Naïve Bayes

The Naïve Bayes classifier technique is focused on Bayesian theorem and is especially applied when the input dimensionality is large. The Bayesian classifier is able to identify as much output as possible, dependent on data. Adding new data may also be applied to runtime and provide a stronger probabilistic classifier. It assumes that the existence/absence of a specific attribute of the class is irrelevant to the existence or absence of some other attribute when provided with the class variable. It is used for clustering and classification. For a learning problem, NB classifiers are extremely recursive, allowing a variety of linear constraints in the set of variables. Parameter estimation training may be achieved by analyzing a closed

expression that requires linear time rather than expensive iterative inference as found in many other classifier forms(Romana, 2017).

2.1.6 Decision Tree

Decision trees are certain kinds of trees that identify classes by grouping them according to their values. Decision tree is primarily used for the purpose of classification. Every tree is composed of branches and nodes. That node reflects characteristics in a category to be categorized and each division is a meaning the node will take(Dey, 2016). Its internal node reflects the attributes, branches reflect the rules of decision and leaf nodes reflect the output. Some algorithms based on decision trees are(Rokach & Maimon, 2005):

2.1.6.1. ID3

Quinlan suggested an algorithm for the ID3. It's called a really basic algorithm of the decision tree. ID3 uses Criteria for Splitting Data Gain. The developing ends where all instances contribute to a single goal value or where the maximum benefit of knowledge is no greater than zero. No pruning techniques refer to ID3. This treat neither the integer properties nor the absent values.

2.1.6.2. C4.5

C4.5 is an ID3 progression, brought forth by the same author. This makes use of Gain Ratio as basis for separation. If the number of instances to be divided falls beyond a specified level, the separation falls halted. After the rising process error-based pruning is done. C4.5 treats integer attributes. It will induce from a training set containing missing values by utilizing modified benefit ratio parameters.

2.1.6.3. CART

CART means Classification and Regression Trees. Breiman established CART, defined by the fact that it builds binary trees, namely that every internal node has exactly two outgoing sides. The splits are chosen using the Twoing Criterion, and Cost-Complexity Pruning is used to prune the received branch. The costs of misclassification in the tree induction can be regarded when given by CART. It also helps users to offer prior distribution of likelihood.

2.1.7 Random Forest

Random forest algorithm builds decision tree on data sets, then gets the result from each and eventually chooses the best answer by voting. This ensemble approach is stronger than a single decision tree and by integrating the outcome, it reduces the overfit. It is a full set or collection of fast and efficient tree predictors where each tree predictor is self- dependent to produce an answer when provided a range of predictor values. Random forest is a mixture of tree predictors in which each tree operates on the values of an arbitrary selected random variable of the same variable for all forest trees. There are three decisions which are made when using Random forest- how to split the leaves, the predictor type used in every leaf and how randomness will be implemented in the trees.

2.2 Unsupervised Learning

Unsupervised learning is the type of learning in which only input data is available but no corresponding output variable. The aim here is to understand the underlying structure or type of distribution in the dataset in order to learn more about it. In this approach the data instances of a training dataset do not have an expected output associated to them instead unsupervised learning algorithm detects patterns based on characteristics of the input data. Example is clustering. In this, similar data are grouped in order to identify clusters of data. Algorithms are left on their own and present the interesting structure in the data.

2.2.1 K-Means Clustering

K-means is still one of the easiest unsupervised learning algorithms addressing the well-known problem of clustering. The method follows a clear and fast way to identify a specified collection of data into a certain number of defined apriori clusters (assume k clusters). Defining k centers is the key concept, one per each cluster. Such centers should be put in a sly way since they produce various outcomes because of specific locations. And the best option is to separate them from each other as much as feasible. The next move is to take every point that belongs to a collection of data and connect it with the nearest middle. The first phase is done when no stage is pending. We need to re-k new centroids at this point as the resulting barycenter of the clusters from the previous stage. Once we have these k new centroids, a new connection between the same data set points and the nearest new center has to be created. A loop was built. As a part of this process we will find that centers shift their position step by step until no more adjustments are made.

2.2.2 Hierarchical Clustering

Hierarchical clustering, also described as hierarchical cluster analysis, is an algorithm where related items are clustered into groups called clusters. The endpoint is a series of clusters, in which each cluster is different from another cluster, and the items inside each cluster are largely identical. Hierarchical clustering begins by considering each occurrence as a cluster of its own. Next, it performs the following two steps repeatedly: define the two clusters nearest to each other, and combine the two most related clusters. It happens until both of the clusters combine. It uses one of the two approaches:

2.2.2.1. Agglomerative Hierarchical Clustering Technique

In this methodology each data point is initially treated as an independent cluster. The related clusters combine with other clusters at each iteration before they shape one cluster or K cluster.

2.2.2.2. Divisive Hierarchical Clustering Technique

Divisive Hierarchical clustering is exactly the inverse. We treat all the data points as a single cluster and isolate the data points from the cluster in each iteration which are not identical. Any data point that is isolated is viewed as an independent cluster. We are left with n clusters at the center.

2.2.3 Apriori Algorithm

Apriori is a simple algorithm in machine learning that is used to organize knowledge into categories. It defines a certain number of characteristics of a dataset and tries to remember how much the characteristics show up in the set. A popular data attribute is one that happens over the pre-arranged number known as support. Apriori is used specifically for sorting out vast volumes of details. Data processing also occurs according to related rules. Rules tend to demonstrate what feature of various sets of data has in common. Apriori can be used as a foundation for a neural artificial network. This will help the network make sense of massive data reconstitutions and then instantly filter data into groups by frequency.

2.2.4 Particle Component Analysis

The particle component analysis (PCA) attempts to reduce the dimensionality of a data collection composed of many variables strongly or gently correlated with each other and retains the variance present in the dataset to the fullest extent. Likewise, the variables are converted into a new group of variables known as the key components. Therefore, the 1st main component preserves the same variance of the initial components that was available. The central components are the covariance matrix's eigen vectors and are thus orthogonal. It is necessary to scale down the dataset on which the PCA technique is to be used. The findings are sensible of the proportional scaling as well.

2.2.5 Neural Network

Neural Networks are forms of computer design influenced by biological neural networks which are used to estimate functions that are usually unknown and may rely on a wide number of inputs. Artificial neural networks are described as interconnected "neuron" systems which, due to their adaptive existence, can compute values from inputs and are capable of both machine learning and pattern recognition(Nikam, 2017). An artificial neural network works by linking several separate computing elements in a biological brain each corresponding to a single neuron. A digital computer machine will directly create or simulate certain neurons. Every neuron takes several input signals and generates a single output signal dependent on an internal weighting, which is sent to another neuron as feedback.

2.3 Semi Supervised Learning

Semi-supervised learning is the machine learning division dealing with the usage of both labelled and unlabeled data to carry out such learning activities. Conceptually placed between supervised and unsupervised learning, in accordance with usually smaller collections of classified data, it helps to leverage the vast volumes of unlabeled data present in many use cases. Acquisition of labelled data also involves a physical trial for a learning problem. As a consequence, the costs associated with the labelling method will make a completely branded training set unfeasible, while the processing of unlabeled data is fairly inexpensive. Semi-supervised instruction in these cases may be of considerable functional benefit. Since semi-supervised instruction needs fewer human intervention and provides greater precision, both theory and experience are of considerable importance. Some semi supervised learning models are:

2.3.1 Generative Models

Generative models are maybe the oldest semi-supervised form of learning. This assumes a configuration $p(x, y)= p(y)p(x)$ in which $p(x)$ is an observable distribution of mixtures, for example Gaussian mixtures. The mixture components may be defined with huge amounts of unlabeled data; therefore, preferably we will need one labelled example per variable to accurately evaluate the distribution of mixtures(Zhu, 2008).

2.3.2 Self Training

Self-training is a semi-supervised learning method which is widely used. In self-training a classifier with the minimal amount of labelled data is first trained. The classifier is then used for the classification of the unlabeled data. Usually, unlabeled points with the most trust are applied to the training collection along with their projected marks. It retrains the classifier and continues the process. Notice the classifier is educating itself with its own predictions. The treatment is often called self-teaching(Zhu, 2008).

2.4 Ensemble Learning

An ensemble consists of a set of self-trained classifiers (such as neural networks, or decision trees) that combine predictions while classifying specific cases. Previous analysis has demonstrated that an ensemble is always more accurate than either of the ensemble's single classifiers. Bagging and boosting are two common methods for generating specific ensembles.

2.4.1 Bagging

Bagging is a bootstrap ensemble strategy that produces individuals by training each classifier on a random redistribution of the training set for its ensemble. Every classifier's training set is created by random drawing, with substitution, N examples– where N represents the size of original training set; several of the original examples may be replicated in the subsequent training set, while others can be left out. Using a separate random sampling of the training set, each individual classifier within the ensemble is created(Opitz & Maclin, 1999). Bagging is useful on unstable learning algorithms where minor variations in the training set result in significant predictive improvements.

2.4.2 Boosting

Boosting incorporates a network of processes. Such approaches are based on generating a set of classifiers. The training set used by each of the series participants is picked based on the results of the series earlier classifiers. Throughout Boosting, examples which are incorrectly predicted in the sequence by previous classifiers are selected more frequently than examples which have been correctly predicted. So, Boosting aims to generate new classifiers that can help forecast cases on which the efficiency of the current ensemble is low.

2.5 Reinforcement Learning

Reinforcement learning allows software agents and machine to automatically determine the ideal behavior within a specific context to maximize its performance. It is about interaction between two elements, the environments and learning agent. The learning agent leverages to mechanism namely exploration and exploitation. When learning agent acts on trial and error basis, it is called exploration and when it acts based on the knowledge gained from the environment, it is called exploitation. The environment credits the learning agent for taking right actions which is reinforcement trigger. Taking use of the incentives received, the agent develops his knowledge of the environment to choose the next move. Reinforcement learning can be categorized into two categories:

2.5.1 Markov Decision Process

The Markov decision-making method is a paradigm of predicting output. Unlike a Markov chain, the model aims to predict an outcome provided that the current state only provides details. The Markov judgment procedure, integrates the features of the behavior and intentions. The decision maker may choose, at each stage during the process, to take an action required in the present state which results in the model going to the next phase and providing a reward to the decision maker. For an optimization problem a machine learning algorithm may be assigned. The algorithm would try to automate the behavior performed within an environment by leveraging reinforcement learning to increase the future reward. While supervised learning approaches involve proper input / output pairs to construct a pattern, reinforcement learning is utilizing Markov decision processes to establish an appropriate exploration and exploitation combination. Machine learning, where the odds and incentives of a result are uncertain or unclear, can use reinforcement learning using the Markov decision cycle.

2.5.2 Evolutionary Algorithm

It works differently. The aim is to build computer code that will solve a specific problem using an approach to developing computer code that will solve a special problem using a method. Evolutionary algorithm starts completely at random with code generation. Every of such codes were checked to see whether it was meeting the intended goal. Code develops in this way. Overtime, it gets older, and if circumstances are correct over several years, it will become greater than any human coder would build. This consists of sustaining a distribution of network weight values and utilizing parameters extracted from this distribution to operate in tandem with a wide number of agents. These agents behave in their own manner and after a specified number of episodes have been completed, total compensation is returned to the algorithm as a fitness ranking. With this ranking, the distribution of parameters may be shifted to those of the more active operators, and away from that of the ineffective ones.

III. CONCLUSION

Machine learning is a broad area of research that overlaps with theories from other different areas such as artificial intelligence, and inherits them. The field's emphasis is thinking, that is, the acquiring of skills through experience. This involves synthesizing practical ideas from historical records. As such, you will experience several specific forms of learning in machine learning as a practitioner.

Within this paper we offer a short description of what is machine learning and the various forms of algorithms in specific machine learning categories. The methods mentioned can be implemented to different forms of data collection, i.e. medical, financial, etc. This paper provides an introduction to several of the common algorithms for machine learning. Shifting the core problem from how to program a machine to how to enable it to program itself, Machine learning prioritizes the creation of systems that are self-learning and self-reliant and the use of the usable data flow inside the system rather than merely manipulating it.

IV. ACKNOWLEDGMENT

I would like to thank my guide Mr. Kapil Sharma, Computer Science department, GNDEC College, Ludhiana for his continued guidance. Also, I would like to thank my family and friends for their support in my work.

REFERENCES

[1] A. Raj, "A Review on Machine Learning Algorithms," *International Journal for Research in Applied Science and Engineering Technology*, vol. 7, no. 6, pp. 792–796, 2019, doi: 10.22214/ijraset.2019.6138.

[2] A. S. Romana, "A Comparative Study of Different Machine Learning Algorithms for Disease Prediction," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 7, p. 172, 2017, doi: 10.23956 /ijarcsse/v7i7/0177.

[3] A. Dey, "Machine Learning Algorithms: A Review," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016.

[4] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - A survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 4, pp. 476–487, 2005, doi: 10.1109 /TSMCC.2004.843247.

[5] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," *International Journal of Modern Trends in Engineering & Research*, vol. 4, no. 7, pp. 58–63, 2017, doi: 10.21884/ijmter.2017.4211.vxayk.

[6] X. Zhu, "Semi-Supervised Learning Literature Survey Contents," *Sciences New York*, vol. 10, no. 1530, p. 10, 2008, doi: 10.1.1.146.2352.

[7] D. Opitz and R. Maclin, "Popular ensemble learning: an empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999.

[8] K. Alsabti, "Electrical Engineering and Computer Science An efficient k-means clustering algorithm," 1997.

[9] L. Kaelbling, M. Littman, and A. Moore, "Reinforcement Learning? A Survey," *Journal of Artificial Intelligence Research*, vol. 4, no. 1, pp. 237–285, 1996.

[10] D. Lowd and P. Domingos, "Naive Bayes models for probability estimation," *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, no. January 2005, pp. 529–536, 2005, doi: 10.1145/1102351.1102418.

[11] L. Breiman, "ST4_Method_Random_Forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004.

[12] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2019, doi: 10.1007/s10994-019-05855-6.