

A Study on Machine Learning: Approaches and Applications

Nisha Charaya

Assistant Professor, Department of Electronics & Communication Engineering, Amity University Haryana,
Gurugram, India

ABSTRACT: Machine learning has emerged and is continuously growing as an efficient tool in computer processing. With having numerous applications it has proved as a boon to data analysts. Machine learning has different approaches for different circumstances. In this paper, the machine learning approaches and its real life applications are presented.

KEYWORDS: Learning phase, Reinforcement, Supervised, Tracking, Training phase, Unsupervised.

I. INTRODUCTION

We are witnessing the most remarkable period of human history. The period when computing moved from large mainframes to PCs to cloud. The democratization of various tools and techniques has made it exciting and thrilling at the same time. But the best is yet to come in coming years[1]. The field of machine learning has gained popularity in years since Data Scientist job has been named the Sexiest job by an article in Harvard Business Review [2]. As Big Data is the hottest trend in the tech industry at the moment, machine learning is incredibly powerful to make predictions or provide calculated suggestions based on large amounts of data. Some of the most common examples of machine learning are Netflix's algorithms to make movie suggestions based on movies watched in the past or Amazon's algorithms that recommend books based on books bought before[3].

Machine learning algorithms are the programs that can learn from data and improve from experience, without human intervention. Learning tasks may include learning the function that maps the input to the output, learning the hidden structure in unlabeled data; or 'instance-based learning', where a class label is produced for a new instance by comparing the new instance (row) to instances from the training data, which were stored in memory. 'Instance-based learning' does not create an abstraction from specific instances.

In simple language, Machine Learning algorithm is an evolution of the regular algorithm. It makes the programs "smarter", by allowing them to automatically learn from the data provided [5]. Machine Learning made it possible to build software that can understand images, sounds, and language and to learn more about technology each day [7].

II. MACHINE LEARNING ALGORITHM PHASES

The general machine learning algorithm is mainly divided into two phases:

- **Training Phase:**

In the training phase, the model/ machine is made to learn from the portion of dataset called as training data. The training data is used to fit the model. This training data will help the machine to connect the patterns in the training data to the right answer. For an example, while buying some random apples from the market based on its characteristics like color, size, shape, vendor, origin etc and feed it (regression/ classification) as a specimen to the model along with its output features as sweetness, juiciness and ripeness. It will act as a training data and the model will learn from it. It fits a model of the correlation between an average apple's physical characteristics, and its quality.

The portion of complete dataset selected as training data is not fixed and is totally up to the programmer. Generally 70% of total data is taken as training data, but it is not essential that 70% of the data has to be for training and rest for testing. It completely depends on the dataset being used and the task to be accomplished. For example, if training data is taken as 50% and remaining 50% as test data, the precision will be different from training it on 90% or so. This is mostly because in Machine Learning, the bigger the dataset to train, better is the learning and hence accuracy of the model [5].

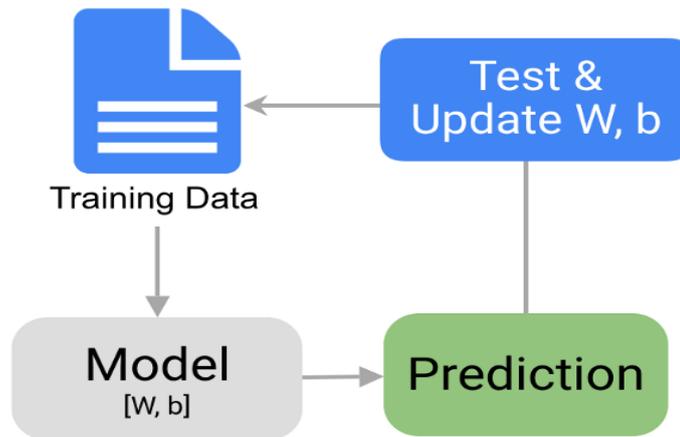


Fig 1: Phases of Machine Learning Algorithm [4]

- **Testing phase**

Once the algorithm/model is trained, it is used to predict the unknown results of the test data. The accuracy and precision of the model is calculated from there. In continuation with previous example, while buying apples again, the characteristics of the apples which are available for buying (test data) will be measured and feed it to the Machine Learning algorithm. It will use the model which was computed earlier to predict if the apples are sweet, ripe and/or juicy. The algorithm may internally use the rules, similar to the one manually written earlier (for eg, a **decision tree**). Finally, the apples can be bought with great confidence, without worrying about the details of how to choose the best apples. The best part of machine learning is that, the same algorithm can be used to train different models. We can create one each for predicting the quality of mangoes, grapes, bananas, or any other fruit.

This is crucial to choose the right parameters for the estimator. Sometimes, the third set of data i.e. cross-validation data is used to ensure better accuracy and efficiency of the algorithm used to train the machine. The training dataset is divided into train set and validation set. Based on the validation test results, the model can be trained (for instance, changing parameters, and classifiers). This will help to get the most optimized model. The test data is used to see how well the machine can predict new answers based on its training [5].

III. APPROACHES

The machine learning algorithms can be classified in two ways:

(a) grouping of algorithms by their learning style.

(b) grouping of algorithms by their similarity in form or function (like grouping similar animals together)[6].

Grouping of algorithms by their learning style is preferred and convenient among these. Based on the learning style, algorithms can be classified into three categories: Supervised Learning, Unsupervised Learning and Reinforcement learning. Figure 2 shows the classification of machine learning algorithm based on its learning style.

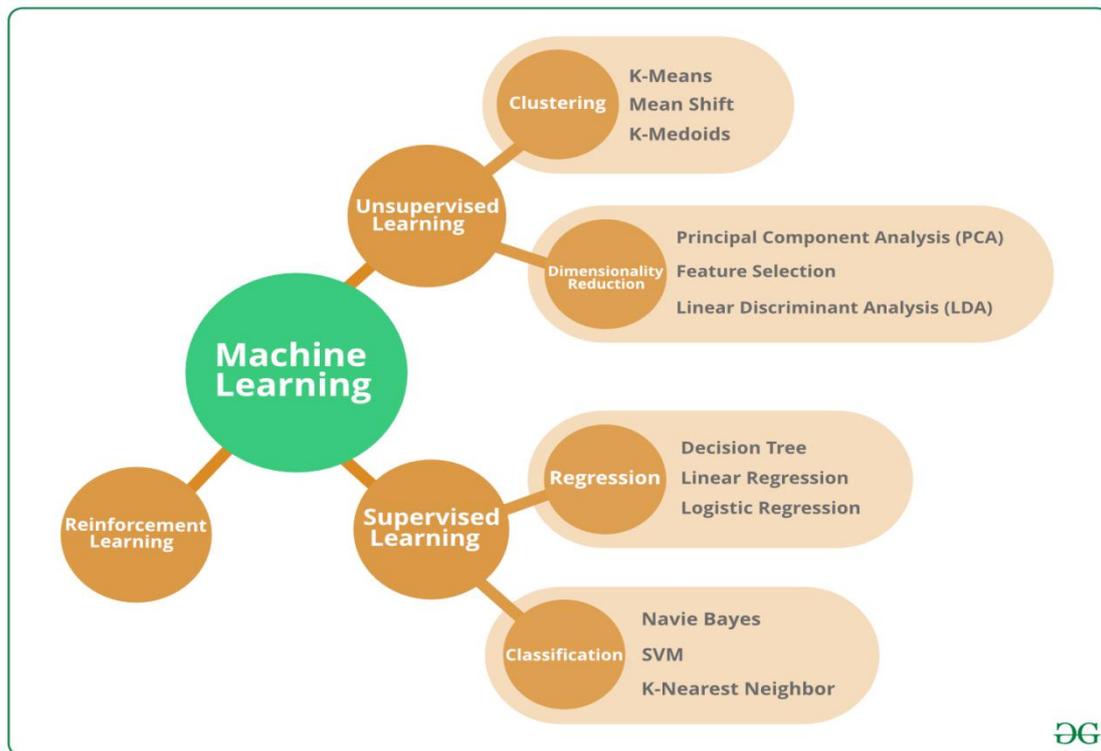


Fig 2: Machine Learning Algorithm Approaches [10]

1. Supervised Learning

This category is termed as supervised learning because the process of learning from the training dataset can be thought of as a teacher teaching his students. The algorithm continuously predicts the result on the basis of training data and is continuously corrected by the teacher. The learning continues until the algorithm achieves an acceptable level of performance.

In Supervised machine learning algorithm, every instance of the training dataset consists of input attributes and expected output. The training dataset can take any kind of data as input like values of a database row, the pixels of an image, or even an audio frequency histogram.

For example, in biometric attendance the machine is trained with inputs of user’s biometric identity that can be thumb, iris or ear-lobe, etc. Once the machine is trained it can validate the future input and can easily identify the user [5].

It uses labeled training data to learn the mapping function that turns input variables (X) into the output variable (Y). In other words, it solves for f in the following equation:

$$Y = f(X)$$

This allows the machine to accurately generate outputs when given new inputs.

There are two major types of supervised learning: classification and regression.

Classification is used to predict the outcome of a given sample when the output variable is in the form of categories. A classification model might look at the input data and try to predict labels like “sick” or “healthy.”

Regression is used to predict the outcome of a given sample when the output variable is in the form of real values. For example, a regression model might process input data to predict the amount of rainfall, the height of a person, etc.

Linear Regression, Logistic Regression, CART, Naïve-Bayes, and K-Nearest Neighbors (KNN) are examples of supervised learning.

Ensembling is another type of supervised learning. It means combining the predictions of multiple machine learning models that are individually weak to produce a more accurate prediction on a new sample. Bagging with Random Forests, Boosting with XGBoost are examples of ensemble techniques [2]

2. Unsupervised Learning Algorithms:

This category of machine learning is known as unsupervised because unlike supervised learning there is no teacher. Algorithms are left on their own to discover and return the interesting structure in the data.

The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

In this approach, the sample of a training dataset does not have an expected output associated with them. Using the unsupervised learning algorithms, patterns can be detected based on the typical characteristics of the input data. Clustering can be considered as an example of a machine learning task that uses the unsupervised learning approach. The machine then groups similar data samples and identify different clusters within the data.

For example, fraud Detection is probably the most popular use-case of unsupervised learning. Utilizing past historical data on fraudulent claims, it is possible to isolate new claims based on its proximity to clusters that indicate fraudulent patterns [5].

Unsupervised learning models are used when there are no corresponding output variables with the input variables (X). They use unlabeled training data to model the underlying structure of the data.

There are three types of unsupervised learning:

(a) Association is used to discover the probability of the co-occurrence of items in a collection. It is extensively used in market-basket analysis. For example, an association model might be used to discover that if a customer purchases bread, he/she is 80% likely to also purchase eggs.

(b) Clustering is used to group samples such that objects within the same cluster are more similar to each other than to the objects from another cluster.

(c) Dimensionality Reduction is used to reduce the number of variables of a data set while ensuring that important information is still conveyed. Dimensionality Reduction can be done using feature extraction methods and feature selection methods. Feature selection selects a subset of the original variables. Feature extraction performs data transformation from a high-dimensional space to a low-dimensional space. Principal component analysis (PCA) algorithm uses a feature extraction approach.

Apriori, K-means and PCA are examples of unsupervised learning [2].

3. Reinforcement learning:

Reinforcement learning is a type of machine learning algorithm that allows an agent to decide the best next action based on its current state by learning behaviors that will maximize a reward [2].

Reinforcement learning can be thought of like a hit and trial method of learning. The machine gets a reward or penalty point for each action it performs. If the option is correct, the machine gains the reward point or gets a penalty point in case of a wrong response.

The reinforcement learning algorithm is all about the interaction between the environment and the learning agent. The learning agent is based on exploration and exploitation.

Exploration is when the learning agent acts on trial and error whereas exploitation is when it performs an action based on the knowledge gained from the environment. The environment rewards the agent for every correct action, which is the reinforcement signal. With the aim of collecting more rewards obtained, the agent improves its environment knowledge to choose or perform the next action [5].

For example, a video game in which the player needs to move to certain places at certain times to earn points. A reinforcement algorithm playing that game would start by moving randomly but, over time through trial and error, it would learn where and when it needed to move the in-game character to maximize its point total [2].

Supervised learning is useful in cases where a property (label) is available for a certain dataset (training set), but is missing and needs to be predicted for other instances. Unsupervised learning is useful in cases where the challenge is to discover implicit relationships in a given unlabeled dataset (items are not pre-assigned). Reinforcement learning falls between these 2 extremes i.e. there is some form of feedback available for each predictive step or action, but no precise label or error message[3].

IV. APPLICATIONS

The most common real life applications of machine learning are:

Virtual Personal Assistants: Google Now, Siri, Alexa are popular examples of virtual personal assistants. They assist the user in finding information, when asked over voice. Machine learning is an important part of these personal assistants as they collect and refine the information on the basis of user's previous involvement with them. Later, this set of data is utilized to generate results as per user preferences.

Predictions while Commuting: GPS navigation services are a must while driving on the roads. It saves the current locations and speeds of all users at a central server for managing traffic. This data is then used to build a map of current

traffic. This helps in preventing the traffic and does congestion analysis. While for the cars without GPS, Machine learning plays its role by estimating the regions where congestion can be found on the basis of daily experiences.

Web Search Engine: The search engines like google, bing etc rank pages through a complex learning algorithm.

Photo tagging Applications: The social media app like facebook uses a face recognition algorithm that runs behind the application that makes the posted photo more interesting and memorable.

Spam Detector: All mail agents like Gmail or Hotmail classifies the mails and moves the spam mails to spam folder by using a spam classifier which runs at the back end of mail application [9].

Marketing and Sales: The online shopping apps use machine learning technology to analyze the purchase history of their customers and make personalized product recommendations for their next purchase.

Financial Services: Machine learning technology is helpful in analyzing the financial data and hence preventing the financial frauds. The technology is also used to identify opportunities for investments and trade. The cyber surveillance helps in identifying those individuals or institutions which are prone to financial risk, and take necessary actions in time to prevent fraud.

Healthcare: ML is becoming a fast-growing trend in healthcare with the advent of wearable sensors and devices that use data to access health of a patient in real time. These sensors provide real-time patient information, such as overall health condition, heartbeat, blood pressure and other vital parameters. These data can be further used by medical experts to analyze the health condition of an individual and predict the occurrence of any ailments in the future. The technology also helps to identify trends that facilitate better diagnoses and treatment [10].

V. CONCLUSION

Machine learning is a boon in computer processing to handle the complex problems with ease. It has been implemented in various areas and is especially helpful in handling big data. It has solved complicated problems and made our life easy in many areas of its applications. However, it can be further combined with other technologies to improve its performance in terms of accuracy and time-complexity.

REFERENCES

- [1] <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>
- [2] <https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/>
- [3] <https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html>
- [4] <https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html>
- [5] <https://www.edureka.co/blog/machine-learning-algorithms/>
- [6] <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
- [7] <https://towardsdatascience.com/top-10-algorithms-for-machine-learning-beginners-149374935f3c>
- [8] <https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0>
- [9] <https://www.geeksforgeeks.org/machine-learning-introduction/>
- [10] <https://www.geeksforgeeks.org/top-10-algorithms-every-machine-learning-engineer-should-know/>