International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Use of Machine Learning in the Pattern Finding

Uttama Garg, Anand Kumar Shukla, Harmanjeet Singh, Nisha Sharma, Sandeep Kaur

Abstract*: Today is the generation of Machine Learning and Artificial Intelligence. Machine Learning is a field of scientific study and statistical models to predict the answers of never before asked questions. Machine Learning algorithms use a huge quantity of sample data that is further used to generate model. The higher amount and quality of training set lead to higher accuracy in approximate result calculation. ML is the most popular field to research and also helpful in pattern finding, artificial intelligence and data analysis. In this paper we are going to explain the basic concept of Machine Learning with its various types of methods. These methods can be used according to user's requirement. Machine Learning tasks are divided into various categories . These tasks are accomplished by computer system without being explicitly programmed.

***Keywords* : Classification, clustering, regression, supervised learning**

I. INTRODUCTION

Machine Learning is a field of scientific study that provides a type of capability to computer systems. Using this capability system can [1]learn the things without being explicitly programmed by programmer. Machine Learning is one of the current technologies that is used in almost every aspect of study, research, social media, security and next generation inventions. As the name shows, it makes the computer [7] similar to humans. The phrase "***The ability to learn"*** is the meaning of Machine Learning itself**.** Machine learning is used today so much that one can't expect its uses. A pioneer of Artificial Intelligence and computer gaming named Arthur Samuel is founder of the term "Machine Learning". According to him "Machine Learning is a Field of study that gives computers the capability to learn without being explicitly-programmed". For another definition Machine Learning (ML) can be defined as it is a manner that improve the learning manner of computer system by using it.

Experience or previous data without any human help any code. One need to store a quality data in computer system and then that computer system can provide result for future queries using machine learning algorithms. We need to train

our system with the help of machine learning models. The collection of these algorithms depends on the data we have stored and the task that we are going to perform. One should ensure that appropriate algorithm must be generated through training and learning and feeding-data.

A. Examples

[1]Training of any student during examination. During preparation for the examinations students don't cram the concept but students try to learn and understand the concept right before the exam , they store high quality data in their brain(machine). This data could be any type of questions, answers, diagrams from different books and another sources like online videos, notes etc. Actually by this process they are trying to teach their brain to learn to what type of answer should be given for any question by retrieving that data . During preparation they practice with many sample question papers and then compare their result with answer keys and then make accurate their mistakes and store in brain for next time. Gradually, their performance increases and they get more confidence for future result. That's the way how models are created and then system uses that model to generate algorithm and then that algorithm will be used to produce result for never before questions. Whenever similar question is generated then system remind most suited algorithm and provide result. Researchers are doing research with continuous efforts to improve the quality of algorithms so that these models can perform better.

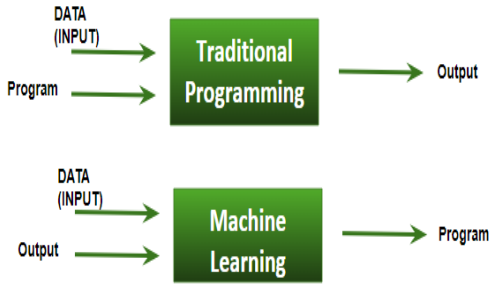


Figure 1: Difference between Machine Learning and Traditional Programming:

B. Traditional Programming

As shown in figure two components DATA (Input) and PROGRAM (logic) are stored by human and then get output.

C. Machine Learning

In this scenario DATA(Input) and Output will be stored, run it on system during training and then machine generates its own program(logic), this program can be used for future results.

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.
* Correspondence Author
Uttama Garg, Assistant Professor and research scholar , University Institute of Computing Chandigarh University, Chandigarh, India.
Anand Kumar Shukla, Associate Professor and Research Coordinator in Chandigarh University, Mohali, Punjab, India
Harmanjeet Singh, Associate Professor and Research Coordinator, Chandigarh University, Mohali, Punjab, India
Nisha Sharma, Associate Professor and Research Coordinator, Chandigarh University, Mohali, Punjab, India
Er. Sandeep Kaur , Associate Professor and Research Coordinator, Chandigarh University, Mohali, Punjab, India

D. Problem in Machine Learning

There are various types of problem in Machine Learning[14]. Here, we discuss some of them. Supervised learning: Labels are given in supervised learning. We can say that examples will be having proper input and output with names. Some of them are: Image Classification: We train the system with images/labels. So that when we enter an image we can get their name.

Regression: We can train the system with historical data and ask the computer system to predict the new price of that product in the future.Unsupervised learning: No labels are given with examples that are to be stored in system. Machine will find its own label on products based on features by making different groups. It is used in discovering hidden patterns.[4]

Clustering: We ask the system to group similar data into clusters.

High Dimension Visualization: We use the system to help in solving problems of high dimension data.

Here are two diagrams describing Supervised and Unsupervised learning.

Here we can see Supervised learning is labelled and unsupervised learning is unlabelled. Some another types of learning like Semi supervised and Reinforcement learning are given below:

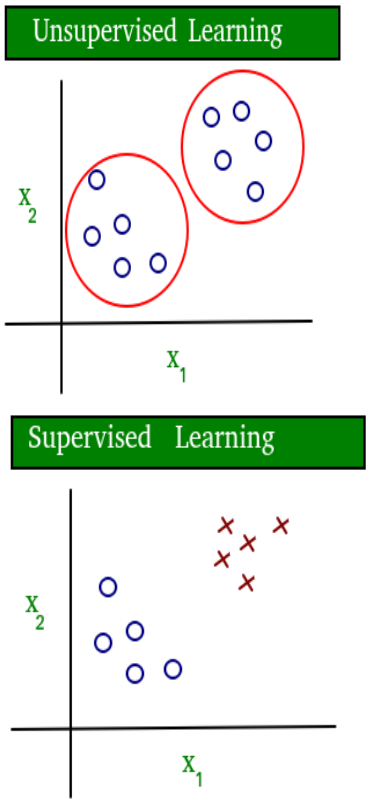


Fig 2: Supervised and Unsupervised Learning

Semi-supervised learning: When we have a large amount of data and some of them are labelled and some of them are not then these problems are in between both supervised and unsupervised learning and known as semi supervised learning.Reinforcement learning: A computer program called agent must be there to intract in dynamic environment and feedback must be given for every round in terms of reward.

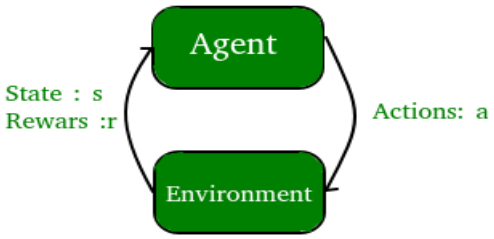


Fig 3: Reinforcement Learning

II. SOME MACHINE LEARNING TECHNIQUES:

A. Classification

System will produce a model on the bases of examples where examples are in form of groups. By filtering based on similar features system can put the output in single category. For example whether an email should be kept in spam or not.

B. Regression

It is a supervised learning problem, where output is based on historic data. When we need a continuous answer rather than discrete in classification, it comes under regression. Example stock price, house price etc[12].Following figure is showing the difference between regression and classification.

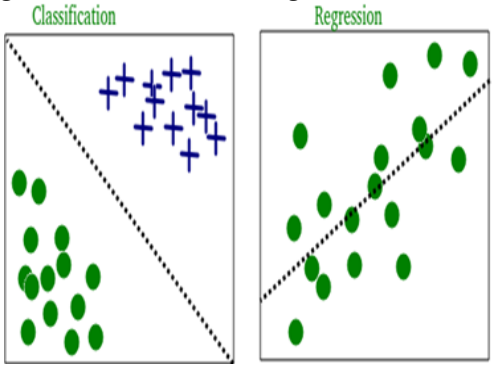


Fig 4: Classification and Regression

C. Clustering

In this, a set of inputs will be given and is organised in groups. But the data is not labelled here. So this is clearly unsupervised machine learning problem. Here an example is given where various datapoints are grouped into three categories which can be identifying in three colors red, blue and green..

Fig 5: Clustering

These are the types of problems which can be solved using machine learning. Some of the famous algorithms or problems are Linear Regression,

Logistic Regression, Decision Tree, Support Vector Machine, Naive Bayes Algorithm, K nearest neighbours, K-Means Algorithm, Random Forest Problem, etc.

III. TERMINOLOGIES OF MACHINE LEARNING:

A. Model

It is also known as hypothesis. This is the representation that is learned from data by applying any algorithm.

B. Feature

Feature is the property of any example or the input. For example to predict a fruit its shape, color, smell, taste are its features. Based on these features any model can be prepared. We can employ a feature extracter to extract the features from any input.

C. Label

A label is the value or result to be predicted by system model[15]. For above example if features are given then we can predict the fruit's name like banana, apple etc.

D. Training

Training is a process before that an input and an output are given and after processing a hypothesis is generated to predict the result.

E. Prediction

Once the hypothesis or model is ready, we can predict the output for never before asked question.
The figure shown below explains the above terms:

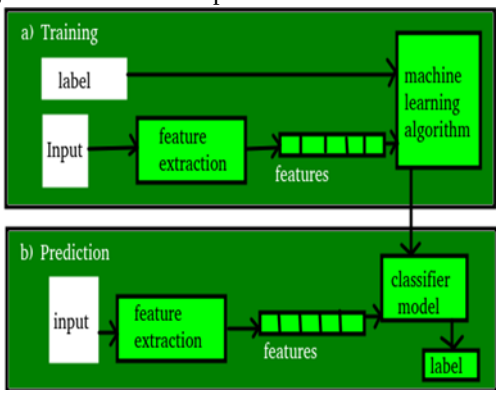


Fig 6: Terms Of Machine Learning

IV. SUPERVISED MACHINE LEARNING:

Maximum problems comes under supervised learning in ML field. Supervised learning problems are those where we have input variable (x) and an output variable (Y) and an algorithm that is used to create the mapping function Y = f(X) . The goal of this learning is to give best approximate solution in terms of when we enter input variable X.
Techniques that comes under [3] Supervised Machine Learning algorithms are Linear Regression, Logistic Regression, Classification, Support Vector Machines (SVM) and Decision Trees. The data that is stored in supervised learning must be labelled with correct answer to get best approximate solution. It can be divided into regression and classification.

A. Regression

When problem is continuous or real then regression is used. We need to feed historic data into system with quality and quantity. The large amount of data will lead to more accurate approximation. "Salary:, "weight", "price" are the problems that use regression model. There are many different models to use but the simplest is the linear regression. It tries to generate a straight line that goes through maximum points. This line is known as hyperplane.

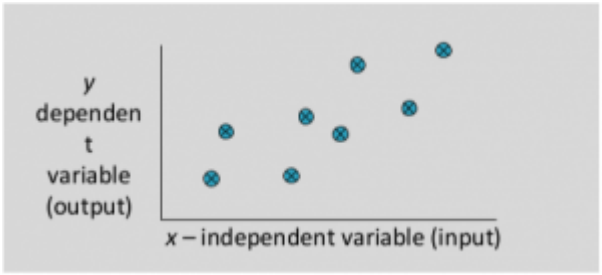


Fig 7: Regression

Types of Regression Models:

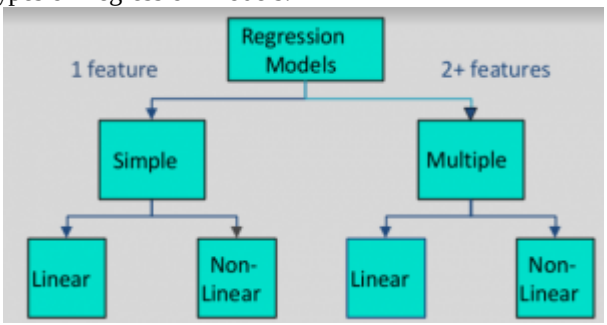


Fig 8: Types of Regression

B. Logistic Regression

Logistic regression is also a supervised machine learning algorithm. It comes under classification problem where the output Y can give only discrete value like yes or no, 0 or 1 etc. for given input X.Logistic model builds a regression model that predict the chances or probability for any input to belong to category valued 1. In linear regression we find a straight linbe with linear function and in logistic model we find S shaped curve with sigmoid function[11].

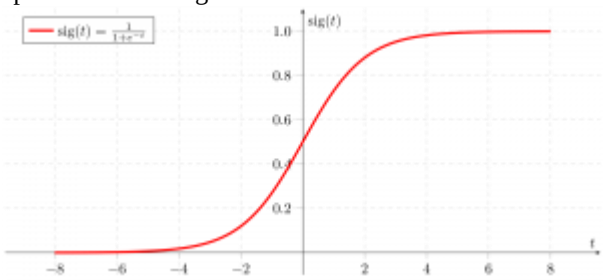


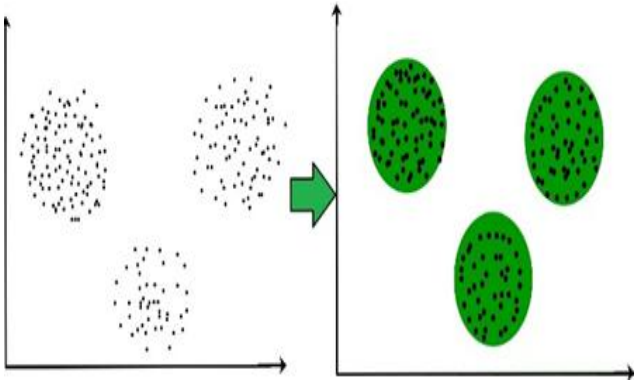
Fig 9: Logistic Regression

Logistic regression also becomes a classification problem when inputs are given into the images. The setting of the inputs and outputs are very important aspect of Logistic model.

V. UNSUPERVISED MACHINE LEARNING

A. Clustering

Clustering comes under unsupervised machine learning technique . An unsupervised learning technique is a method where we feed inputs without labels[13]. It is used to find hidden patterns, making groups and differentiate between various data points. [9]Clustering is used to create meaningful structures by a set of examples. Clustering is a task that divide data points into multiple groups where same data points remain in single group. These data points have more dissimilarity with another groups . so similar objects will remain in single group. Example– Here a graph is shown in



which 3 clusters are given, and data points are clustered together in single group.

Fig 10: Clustering

It is not compulsion for clusters to be spherical always. Such as following graph, here data points are clustered by measuring the distance of data points from the center of cluster. Various distance finding methods and techniques can be used to find out the distance of the outliers. There are no criteria to define good clustering. It is based on user's need. For example, user could be interested in clusters that are homogeneous groups or in data reduction, in finding useful and suitable groupings, in finding unusual data objects or in finding "natural clusters" .

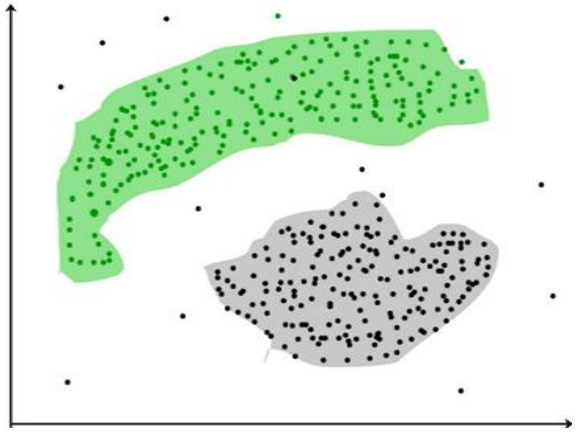


Fig 11: Non spherical Clustering

Clustering Algorithms:

K-means clustering algorithm – It is one of the most simple technique that comes under unsupervised machine learning algorithms to solve clustering problems. K-means algorithm do partitioning of n data points/observations into k clusters such that each data point or each observation belongs to that cluster which is having nearest mean. It could be prototype of that particular cluster[10]. There may be many more ways to make clusters but K Means uses only mean of data points as shown in following figure.

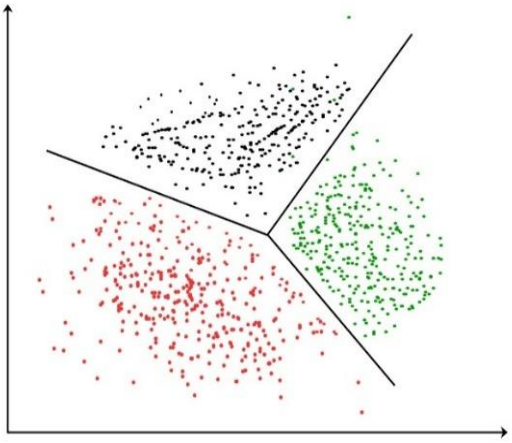


Fig 12: K-means Clustering

B. Decision Tree

The next method to predict any output is Decision Tree. It is one of the most powerful and useful tool in machine learning for classification and making prediction. We can make more accurate prediction using this technique. A Decision tree is a hierarchical type structure having a number of nodes, where each internal [2] node represents a test or constraint on an attribute, each branch indicates a result or outcome of that particular test, and each a leaf node (terminal node) holds a discrete value or class label.

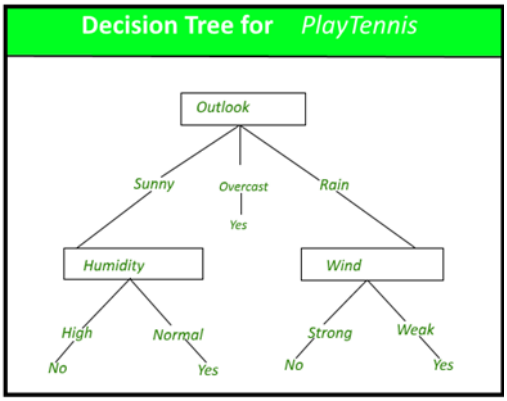


Fig 13: Example of Decision Tree

C. Construction of a Decision Tree

A tree can be constructed by dividing the given inputs or data points into different subsets based on features and constraints. This process can be repeated for each subset in a recursive manner known as [8] "*recursive partitioning*". Decision tree can provide good accuracy. We need not any domain knowledge or any type of parameter setting to construct the decision tree. And also it is easy to implement and understand because of its flowchart like structure.

D. Decision Tree Representation

Decision trees can be represented by sorting data points from root to the leaf node which provide the classification of many data points.

Every data point will be tested from root and then check on branch level then comes to nest level and then compare with attribute value and so on. The decision tree in above figure predict the weather of a particular morning according to its conditions, to check whether it is suitable to play tennis or not at that morning using discrete values YES or NO.

Let us take an example: *(Outlook = Sunny, Humidity = Normal, Wind = Weak)*. This instance would be sorted down from leftmost branch to next right branch. This will lead to a positive instance or suitable for playing tennis. In other words, it can be said that decision tree represents a Disjunction of Conjunctions (DNF) of constraints on the attribute values of various instances.

(Outlook = Rain ^ Humidity = High) v (Outlook = Sunny ^ Wind = Strong)

VI. RESULT

With this study and analysis that is clear that; Pattern recognition is the automated recognition of patterns and regularities in data. Pattern recognition is deeply concerned with artificial intelligence and machine learning, together with applications such as data mining and knowledge discovery in databases (KDD), and is often used interchangeably with these terms

VII. CONCLUSION

By reviewing many web links and published papers it is observed that Machine Learning is all based on training data. Training data with high quality and large amount will lead to a model with more accuracy. It depends on problem statement that which model should be choosen whether it is regression, classification or clustering. Each model of Machine Learning provides a result that approximately same to actual result if we are given with accurate training data.

REFERENCES

1. https://www.geeksforgeeks.org/
2. Shu-Ching Chen. "Effective Feature Space Reduction with Imbalanced Data for Semantic Concept Detection", 2008 IEEE International Conference on Sensor Networks Ubiquitous and Trustworthy Computing (sutc 2008), 06/2008 pythonistaplanet.com
3. Vadym Pasko. "Chapter 3 Prediction of Orbital Parameters for Undiscovered Potentially Hazardous Asteroids Using Machine Learning", Springer Science and Business Media LLC, 2018
4. ekababisong.org
5. edithlaw.ca
6. vitorhugooficial.com.br
7. https://in.zapmetasearch.com
8. V. Kavitha, M. Punithavalli (2010) Clustering Time Series Data Stream – A Literature Survey, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 1, pp. 289-294.
9. Sunita Jahirabadkar and Parag Kulkarni (2013) Clustering for High Dimensional Data: Density based Subspace Clustering Algorithms, International Journal of Computer Applications (0975 – 8887) Vol 63– No.20, pp. 29-35
10. A Statistical Learning Approach to Modal Regression: Yunlong Feng, Jun Fan, Johan A.K. Suykens; (2):1−35, 2020.
11. R.T. Ng, J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining", *Proc. 20th Int'l Conf. Very Large Databases*, pp. 144-155, Sept. 1994.
12. Martin Grohe, Christof Löding, and Martin Ritzert. Learning MSO-definable hypotheses on string. In S. Hanneke and L. Reyzin, editors, Proceedings of the 28th
13. International Conference on Algorithmic Learning Theory, volume 76 of Proceedings of Machine Learning Research, 2017
14. M. Grohe and M. Ritzert. Learning first-order definable concepts over structures of small degree. In Proceedings of the 32nd ACM-IEEE Symposium on Logic in Computer Science, 2017.
15. https://www.omicsonline.org/logistic-regression-journals.php

AUTHORS PROFILE

Uttama Garg is UGC Net qualified**,** working as an Assistant Professor and research scholar in the University Institute of Computing **Chandigarh University, Chandigarh**, India. He has published more than 05 National and International publications.

Dr Anand Kr Shukla is having 15+ years of related work experience, working as an Associate Professor and Research Coordinator in Chandigarh University, Mohali, Punjab, India. He is the solo author of 5 books and more than 30 international publications. He is Life Member (259607) of IAENG (International Association of Engineers), **Hong Kong** and also the member of Advisory Boards and Technical Program Committees of several **IEEE and Elsevier's** International Conferences and supervising 05 Ph.D. Research Scholars. He was also the convener of an International Conference. He is the expert of Java & .Net having sound knowledge of Data Sciences, web engineering, and R language. He has exposure of Software development and have delivered seminars, expert lectures in many Institutions and Universities all over the India.

Mr Harmanjeet Singh, working as an Assistant Professor and research scholar in the University Institute of Computing **Chandigarh University, Chandigarh**, India. He has published more than 05 National and International publications. He is having 8+ years of experience. He was also the academic coordinator of BCA,

.

Ms. Nisha Sharma, working as an Assistant Professor and research scholar in the University Institute of Computing **Chandigarh University, Chandigarh**, India. She has published more than 04 National and International publications. He is having 7+ years of experience.

.

Ms Sandeep Maan, working as an Assistant Professor and Project coordinator in the University Institute of Computing **Chandigarh University, Chandigarh**, India. She has published more than 05 National and International publications. He is having 8+ years of experience.

.