Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

A Look at Some of the Applications and Methods Used in Machine Learning

Sachin Sharma[1], Rahul Bhatt[2], Gesu Thakur[3]

[1]Assistant Professor, Computer Science & Engineering, School of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[2, 3]Associate Professor, Computer Science & Engineering, School of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Chakrata Road, Manduwala, Naugaon, Uttarakhand 248007

[1]socse.sachin@dbuu.ac.in, [2]socse.rahul@dbuu.ac.in, [3]head.ca@dbuu.ac.in

Article Info
Page Number: 2679-2684
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
Data classification and regression are two applications that often make use of machine learning methods. Learning by machine is a method of data analysis that allows for the automated construction of analytical models. It is an area of artificial intelligence that is predicated on the notion that computers can learn from data, recognize patterns, and make decisions with little to no input from human beings. Machine learning is a subfield of computer science that investigates how computers may learn on their own without being given specific instructions. An application that enables the system to learn and develop from experience without being written to that level is known as machine learning (ML), which is a subset of artificial intelligence (AI). Machine learning relies on data to train itself and provide accurate results.

The significance of machine learning lies in the fact that it makes it possible for businesses to recognize patterns of customer behavior and company operations, in addition to assisting in the development of new goods. Machine learning is an integral part of the operations of a significant number of today's most successful businesses, like Facebook, Google, and Uber, amongst others. Many businesses now consider machine learning to be an essential component of their overall competitive strategy.

Keywords: Machine Learning, also known as ML, Random Forest, also known as RF, Decision Tree Regression, also known as DTR, and Support Vector Regression are all types of keywords (SVR)

1. INTRODUCTION

The study of computer algorithms that make it possible for machines to learn and improve themselves automatically based on their previous experiences is known as machine learning. The majority of the time, it is considered to be a subfield of artificial intelligence. The algorithms that are used in machine learning provide computers the ability to come to their own conclusions without the need for human interaction. The recognition of significant underlying patterns in huge amounts of data is the process that leads to these conclusions.

There are three basic categories of machine learning algorithms: supervised learning, unsupervised learning, and reinforcement learning. These classifications are based on the learning method, the type of data that is input and created, and the sort of problem that is solved by the algorithm. The natural expansion of machine learning problem forms is something that may be accomplished using a few hybrid approaches as well as other standard methodology.

2. ESSENTIAL APPROACHES

2.1 Education Under Supervision

A kind of machine learning known as supervised machine learning occurs when the computers are given specific guidelines to follow. Predict the result by utilizing the data from the training set, which should have clear labels. As can be seen from the labeled data, some of the input data has already been categorized with the appropriate output. The training data that is provided to the computer works as a supervisor in supervised learning, teaching the machines how to appropriately predict the output that will be generated by the computer. In the real world, supervised learning may be put to use for a variety of purposes, including risk assessment, the categorization of pictures, the detection of fraud, the filtering of spam, and so on.

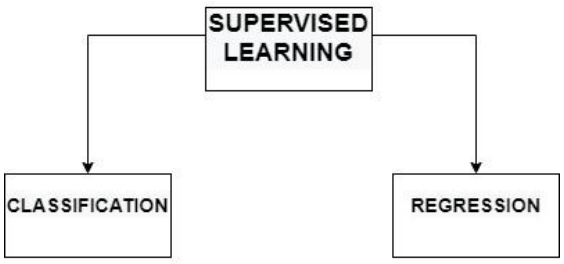

Fig 2.1 Types of Supervised Learning

2.1.1 Classification

Classification is the process of locating a function that aids in the categorization of a dataset based on numerous criteria. This function is often found via the process of searching for a classification function. A computer program is given the opportunity to learn from the training dataset, after which it applies the knowledge it has gained to the task of classifying the data into the appropriate categories. A method of Supervised Learning, the classification algorithm identifies the category of new observations by making use of training data. Classification is the process in which a computer software acquires knowledge from an existing dataset of observations and then uses that knowledge to place new observations into one of a variety of categories or groupings.

A classification algorithm's primary goal is to identify which category a dataset belongs to, and these algorithms are often used to predict the output for categorical data.

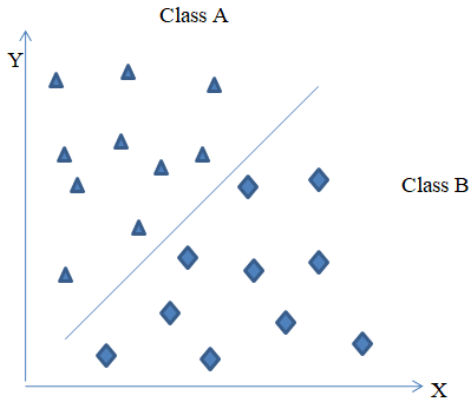

Fig 2.1.1 Classification

2.1.2 Regression

Regression is a statistical technique for identifying the connection between two or more independent variables or features and a dependent variable or outcome. It is a machine learning predictive modeling approach in which an algorithm predicts continuous outcomes.

Regression is a statistical approach used to determine the connection between two or more independent variables or features and a dependent variable or outcome. Outcomes may be anticipated after the relationship between the independent and dependent variables has been estimated. The purpose of regression analysis is to determine how distinct independent variables interact with a dependent variable or result. Models that anticipate or forecast trends and outcomes will be trained using regression methods.

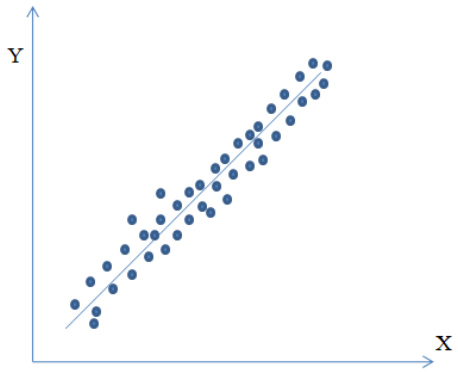

Fig 2.1.2 Linear Regression

2.2 Unsupervised Learning (1.1)

Unsupervised learning is a machine learning approach that creates models without using a training dataset. Models, on the other hand, employ data to identify previously unknown patterns and insights. Unsupervised learning is a kind of machine learning in which models are trained on unlabeled data and then left alone to operate on it.

Unsupervised learning cannot be directly applied to a regression or classification job since,

unlike supervised learning, we have input data but no matching output data. Unsupervised learning seeks to find the underlying structure of a dataset, classify data based on similarities, and present the information in a compact manner.

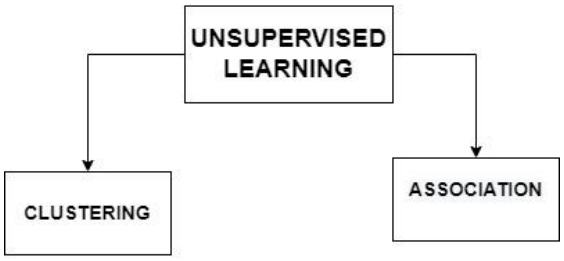

Fig 2.2 Types of Unsupervised Learning

2.2.1 Reinforcement Learning (1.1)

Reinforcement learning is a subfield of machine learning. It all comes down to taking the appropriate actions to maximize your profit in a particular circumstance. A variety of applications and computers utilize it to find the best possible action or course in a given scenario. Reinforcement learning differs from supervised learning in that the answer key is included in supervised learning, allowing the model to be trained with the correct answer, whereas reinforcement learning does not include an answer and instead relies on the reinforcement agent to decide what to do to complete the task. In the absence of a training dataset, it is compelled to learn from its own experience.

1. 1. Algorithms for Supervised Learning
2. 1.1 Regression Using Decision Trees
3. The decision tree is the most fundamental, simple, and widely used classification technique in data mining. This generates a set of basic rules that will be used later to forecast the value of the target variable in a new data point. The produced rules are basic enough for non-experts in data mining to understand. These methods produce models with a graphical tree structure. The model that is built will include decision and leaf nodes. Decision nodes are tree nodes that aid in decision making, while leaf nodes are real decisions. The root node is the tree's highest node. In the provided dataset, the root node is the best predictor variable. The decision tree works with both numeric and nominal datasets.

2.3 Regression using Support Vectors

SVM is a supervised classification technique that predicts the value of a variable. An 'N-dimensional' graph is displayed in this kind of technique, where 'N' is the number of features in a given dataset. Each data record is represented by a plotted plane point. Support vectors are the coordinates of a point. Then it will look for any line that may divide the data points into two distinct groups of data. The line will be drawn such that the distance between the nearest points on both sides of the line in each of the two classes is the greatest.

Random Forest Regression 3.3

2.3.1 A group of decision trees is what we mean when we talk about a Random Forest. In order to categorize a new object based on the attributes it has, each tree is given a classification, and then the tree "votes" for that classification. The classification that received the most votes is the one that is selected by the forest (over all the trees in the forest).

The following procedures are used while planting and cultivating each tree**:•** If the number of instances in the training set is N, a random sample of N cases is chosen. This sample will serve as the tree's training set.

• If there are M input variables, a number mM is supplied so that m variables are chosen at random from the M at each node, and the best split on this m is used to divide the node. Throughout this operation, the value of m is remained constant.

Each tree is grown to the greatest degree feasible. There will be no pruning.

2.3.2 Feature Choice

One of numerous tactics used to enhance the performance of machine learning models is feature selection. It makes it easier to choose just relevant attributes from a list of all available alternatives. When data is gathered, many properties associated with a single data point are obtained. These characteristics may be found in any person, item, document, physical or digital thing. The features in the whole dataset have distinct names. These features are sometimes known as arguments or qualities. Each data point has a unique set of attribute values. In real-world issues, each item in a dataset has a large number of properties associated with it. Manually processing the impact of each characteristic on class value is really difficult. The answer to this problem, referred to as high dimensionality, is feature selection. This method enables algorithms to choose just the most beneficial characteristics that impact prediction accuracy directly or indirectly. With this method, all superfluous characteristics are removed, leaving an easy-to-manage dataset that lowers the time and cost associated with a large number of parameters.

Machine Learning Algorithm Applications

- A decision tree aids in evaluating prospective opportunities for development.
- It aids in the identification of prospective clients via the use of demographic data.
- It is used as a support tool in a variety of professions.
- Face recognition is accomplished using Support Vector Regression.
- SVR is used for text and hypertext classification.
- It aids in picture classification.
- The random forest approach may be used to accomplish both classification and regression tasks.
- Cross validation leads to improved RF accuracy.

- Using the random forest classifier, it is possible to preserve a large level of data accuracy when coping with missing values.
- More trees prevent the model from using over-fitting trees.
- It allows you to work with large, higher-dimensional data sets.

References

1. Alibabaei , Gaspar, Lima, Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling, MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations, 2021, pp. 1- 21.
2. Deepak, Deepika, Dharshibni, Vanathi crop yield prediction based on ensemble model using historical data, International Journal of Advance Research and Innovative Ideas in Education, 2021, pp. 668-673.
3. Suganya, Dayana and Revathi, Crop Yield Prediction using Supervised Learning Techniques, International Journal of Computer Engineering and Technology, 2021, pp. 9-20.
4. Shahhosseini, and Archontoulis SV Forecasting Corn Yield With Machine Learning Ensembles, Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science, 2021, pp. 10-30.
5. Thomas van Klompenburg, AyalewKassahun, CagatayCatal, Crop yield prediction using Machine Learning a systematic literature review, Computers and Electronics an international Journal, 2020, pp. 1-18.
6. Vaishali Pandith, Haneet Kaur, Surjeet Singh, Jatinder Manhas, and Vinod Sharma Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis, Journal of Scientific Research, 2020, pp. 394-398.
7. Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal, Crop Yield Prediction Using Machine Learning Algorithms, Fifth International Conference on Image Information Processing (ICIIP), 2020, pp. 20-40.
8. Eric J. Glover, Gary W. Flake, Steve Lawrence, William P. Birmingham, Andries Kruger, C. Lee Giles, David M. Pennock: Improving Category Specific Web Search by Learning Query Modifications. In Symposium on Applications and the Internet, SAINT 2001, San Diego, California, January 8–12, 2001.
9. J. Kolodner: Case-Based Reasoning. Morgan Kaufmann. 1993.
10. Udo Kruschwitz: A Rapidly Acquired Domain Model Derived from Markup Structure. In Proceedings of the ESSLLI'01 Workshop on Semantic Knowledge Acquisition and Categorisation, Helsinki, 2001.
11. Satoshi Oyama, Takashi Kokubo and Toru Ishida: Domain-Specific Web Search with Keyword Spices. In IEEE Transactions on Knowledge and Data Engineering, 2003 (to appear).
12. Bernhard Sch¨olkopf, Chris Burges, Vladimir Vapnik: Incorporating Invariances in Support Vector Learning Machines. ICANN 96; Springer Lecture Notes in Computer Science, Vol. 1112