Overview of Machine Learning Algorithms and Its Applications

Dr. K.Padmavathi
Assistant Professor
Department of Computer Science & Engineering
PSG College of Arts & Science, Coimbatore, India

Abstract— Machine Learning is a large field of study that uses most powerful exciting technologies and inherits ideas from many fields such as artificial intelligence and data mining. It main focus of the fields are learning or acquiring knowledge from experience and synthesizing useful concepts from historical data. Machine Learning makes the computer more similar to humans and uses or exposes sets of data, which learn from earlier computations to interpret available data and identify hidden patterns. This involves data analysis and automation of analytical model using numerous Machine Learning algorithms. Machine Learning algorithms enable computers and computing machines to search and identify hidden insights and follow a certain type of data for computations. There are many different types of learning techniques and algorithms encountered as a practitioner in the field of machine learning. This paper gives a gentle introduction about the different types of techniques that can be used in the field of machine learning.

Keywords: Machine Learning, Reinforcement Learning, Supervised Learning, Unsupervised Learning

I. INTRODUCTION

Machine learning is a tool, to build algorithms that receive data or information as input and turning the information as output using statistical analysis. Machine learning techniques are used to find the valuable underlying patterns and information within complex data. These complex hidden patterns and knowledge about a problem can be used to predict future events which are used in decision making. The goal of machine learning generally is to understand the structure of data and fit that data into models that can be utilized by people. A machine learning algorithms allow computers to train on data inputs and use statistical analysis and provide outputs that fall within a specific range. Machine learning facilitates automate decision-making processes based on data inputs. In machine learning techniques are classified into number of broad categories. These categories are classified based on how learning is received or how feedback on the learning is given to the system developed.

II. MACHINE LEARNING TECHNIQUES

The most widely adopted machine learning methods are classified into three sub categories. They are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

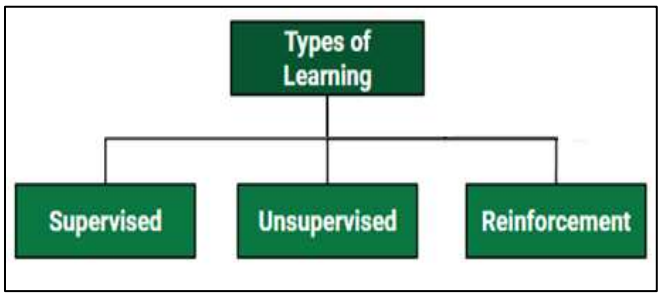

Fig. 1: Types of Machine Learning Techniques

A. Supervised Learning

The supervised learning is used in various important machine learning applications. Supervised learning describes and uses a number of problems that uses a model to learn a mapping between input examples and the target variable. Supervised learning problems use the training data comprises examples of the input vectors along with their corresponding target vectors. In supervised learning, the input variables are (x) and the output variables (Y) and the mapping function of this algorithm from the input to the output is:

$$Y = f(X) \quad \text{----------} \quad (2.1)$$

The mapping function predicts the output variables (Y) based on the new input data (x). Supervised learning uses the training dataset and the algorithm iteratively makes predictions on the training data, and the learning stops when the algorithm achieves an acceptable level of performance. There are two main types of supervised learning problems are:

- Classification
- Regression

Classification is a mostly used supervised learning problem that involves predicting a class label and Regression is another supervised learning problem that involves predicting a numerical label. Both classification and regression problems may have one or more input variables, and input variables may be any data type, such as numerical or categorical. Classification separates the data; Regression fits the data.

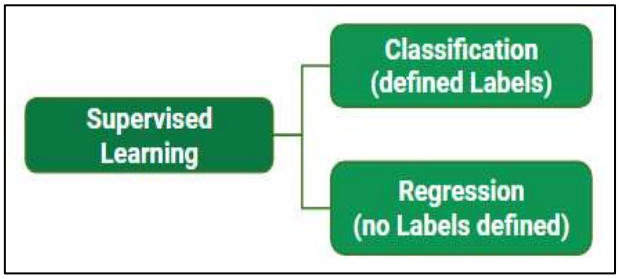

Fig. 2: Types of Supervised Learning Techniques

1) Classification Techniques:

They mainly focused on predicting a qualitative response by analyzing data and recognizing patterns. Classification aims to reproduce class assignments which can be used to predict the response value and the data is separated into "classes".

Example:
- used to classify whether a credit card transaction is fraudulent or not.
- used to classify Gmail mails in more than one classes like social, promotions, updates, forum.

2) Regression Techniques:

It is a supervised learning problem where output is having continuous value. The main aim of this technique is used to predict a value as much closer to actual output value. Regression is used in predicting, forecasting, and finding relationships between quantitative data. It is one of the most popular and earliest learning techniques, which is still widely used in various applications.

Example:
- used to examine the relationship between a company's advertising budget and its sales.

B. Unsupervised learning

Unsupervised learning algorithms process complex processing tasks and it can be more unpredictable compared to supervised learning and other natural learning methods. It deals with the unlabelled data. Unsupervised learning operates the input data without outputs or target variables. Unsupervised learning problems can be classified into the following categories:
- Clustering
- Association problems.

1) Clustering:

It discovers the inherent groupings in the data. Clustering is to create groups of data points in different clusters are dissimilar while points within a cluster are similar. In k-means clustering, data points are combined into *k* groups and larger *k means* smaller groups with more granularity and the lower *k* means larger groups withless granularity. This algorithm creates a set of "labels" assigning each data point to one of the *k* groups. In k-means clustering, the groups are defined is by creating a centroid for each group. The centroids capture and organize the points closest to them and add them to the cluster. The steps involved in k-means clustering:

1) Define the k centroids
2) Find the closest centroid & update cluster assignments
3) Move the centroids to the center of their clusters. The new position of each centroid is calculated as the average position of all the points in its cluster. (Repeat steps 2 and 3 until the centroid stop moving)

Example:
- k-means clustering is used for classifying handwritten digits.

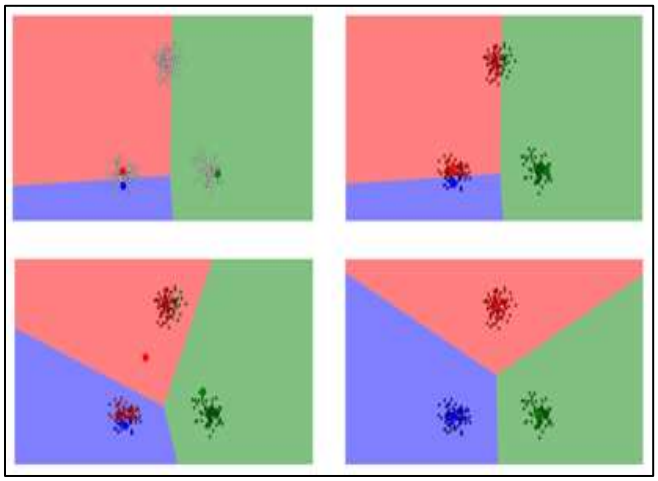

Fig. 3: k-means Clustering Techniques

2) Dimensionality Reduction:

Dimensionality reduction is used to reduce the complexity of the data when the relevant structure as possible. Dimensionality reduction reduces the number of random variables under consideration, by using a set of principal variables. It can be used in feature selection and feature extraction.

Example: simple e-mail classification problem – classification of whether the e-mail is spam or not. This can involve a large number of features, such as whether or not the e-mail has a generic title, the content of the e-mail, whether the e-mail uses a template, etc.

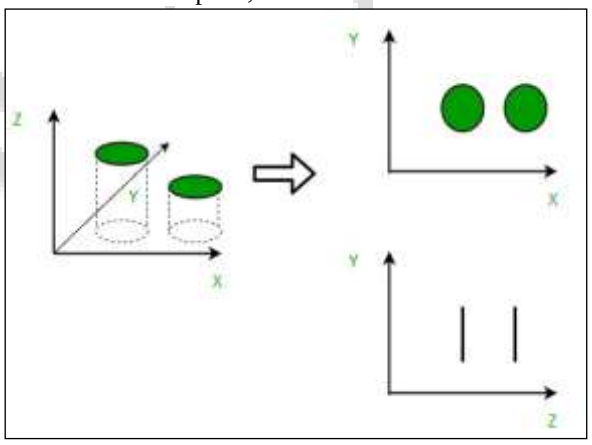

Fig. 4: Dimensionality Reduction

3) Components of Dimensionality Reduction:

There are two different components used in dimensionality reduction:

a) Feature selection:

This is used to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It can be achieved by three ways:
- Filter
- Wrapper
- Embedded

b) Feature extraction:

This is used to reduce the data in a high dimensional space to a lower dimension space.

c) Methods of Dimensionality Reduction:

The different methods used for dimensionality reduction. They are:
- Principal Component Analysis (PCA)

- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA)

Dimensionality reduction may be either linear or non-linear, depending upon the method used.

d) Advantages of Dimensionality Reduction
- used in data compression to reduce storage space.
- reduces computation time.
- removes redundant features, if any.

e) Disadvantages of Dimensionality Reduction
- It may lead to some amount of data loss.
- PCA is used to find linear correlations between variables, which is sometimes undesirable.
- PCA may be fail where mean and covariance are not enough to define datasets.

C. Reinforcement Learning

Reinforcement learning is the training of machine learning models, used to make a decision by sequence of actions. It invokes an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences. There is no fixed training dataset.

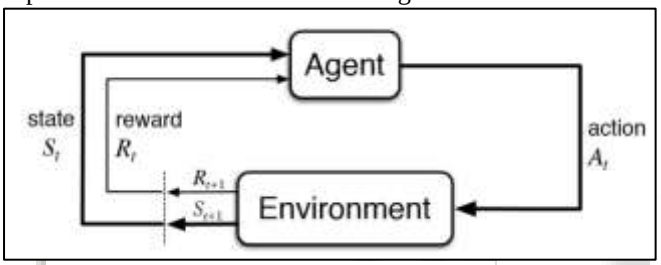

Fig. 5: Reinforcement Learning

In Reinforcement learning,
- The input considers as an initial state of the model.
- It has many possible outputs as there are varieties of solution to a particular problem.
- The training is based upon the input and the model returns a state. The user decides to reward or punish the model based on its output.
- The model keeps continue to learn.
- The best solution is decided based on the maximum reward of number of output values.

a) Types of Reinforcement:

Reinforcement can be classified into two types:
- *Positive:* It is used to define the specific behavior. When an event, occurs due to a particular behavior, increases the strength and the frequency of the behavior. It has a positive effect on the behavior. It Maximizes Performance and sustains Change for a long period of time
- *Negative:* It is used to define the strengthening of a behavior because a negative condition is stopped or avoided. It increases Behavior and provides defiance to minimum standard of performance

b) Applications of Reinforcement Learning are:
- used in robotics for industrial automation.
- used in machine learning and data processing

Generally, RL is used in large environments when a model is known the environment, but an analytic solution is not available.

III. CONCLUSION

The machine learning algorithms generate a new set of rules, based on inferences from the data. This is essential for generating a new algorithm, using different training data. In many complex domains, learning algorithm is used to find the feasible solution to train a program to perform at high levels. This paper has discovered and given a gentle introduction to the different types of learning techniques which may be encountered in the field of machine learning.

REFERENCES

[1] A Tour of the Most Popular Machine Learning Algorithms, https://machinelearningmastery.com/a-tourof-machine-learning-algorithms.
[2] Peter Dayan, "Unsupervised Learning", The MIT Encyclopedia of the Cognitive Sciences, Wilson, RA & Keil, F, Editors, pp. 1-7, 1999.
[3] Machine Learning Algorithms, https://www.dezyre.com/article/top-10-machinelearning-algorithms.
[4] Xin-She Yang, Su Fong Chien, and Tiew on Ting, "Computational Intelligence and Metaheuristic Algorithms with Applications", the Scientific World Journal, Vol.14, pp. 1-4, 2014.
[5] Introduction to Natural Language Processing (NLP): What is NLP?, https://blog.algorithmia .com/ introduction-natural-language-processing-nlp/
[6] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Association for Computational Linguistics, Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 79-86, 2002.
[7] R. P. Ram Kumar, Sanjeeva Polepaka, Lazarus S F, Dasari Vamsi Krishna, "An Insight on Machine Learning Algorithms and its Applications", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-11S2, September 2019.

