UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

Informatics Institute of Technology

Trends in Computer Science

4COSC008C

Machine Learning

2c. Overview of Machine Learning. Describe and compare two different machine learning techniques.

Student Full Name: Nilwakke Kankanamalage Havindu Vijan Wijewardana

UOW Number: w2053121

IIT Number: 20220895

Group Members:

Student Name	UOW ID	IIT ID
Vijan Wijewardana	W2053121	20220895
Navindu Fernando	W2053138	20220965
Ashen Panduka	W2053116	20220966
Malith Tissera	W2052092	20220792
Aathiq Ahamed	W2053136	20220942

I. ACKNOWLEDGEMENT

My lecturer, Mrs. Sulochana Rupasinghe, for her support, and encouragement. Her support and advice have been crucial in helping me stay motivated and accomplish my objectives. I also want to thank again to Mrs. Sulochana Rupasinghe and Mrs. Mohanadas Jananie my tutorial lecturers, They give many lot of supports and knowledge and motivation to make me as successful person. I would like to thank Dr M. Chondrogianni , Module leader for 4COSC008C Trends in Computer and my parents for their support.

II. TABLE OF CONTENT

III. LIST OF FIGRES

IV. LIST OF TABLES

1. INTRODUCTION

1.1. What is machine learning?

Machine learning is an evolving branch of computational algorithms that designed are to emulate human intelligence by learning from the surrounding environment.

(El Naqa, I., Murphy, M.J. (2015)

1.2. How Machine learning works?

Machine learning allows computers to memorize from information, mirroring the way people secure information through involvement.

Machine learning allows computers to learn from experience, similar to how humans do. By analysing data like images or text, it identifies patterns and uses them to make predictions on new data, improving its performance without explicit instructions. This is akin to learning a language; the more examples you see, the better you understand its structure and can form your own sentences. (Mitchell, T. M.,1997)

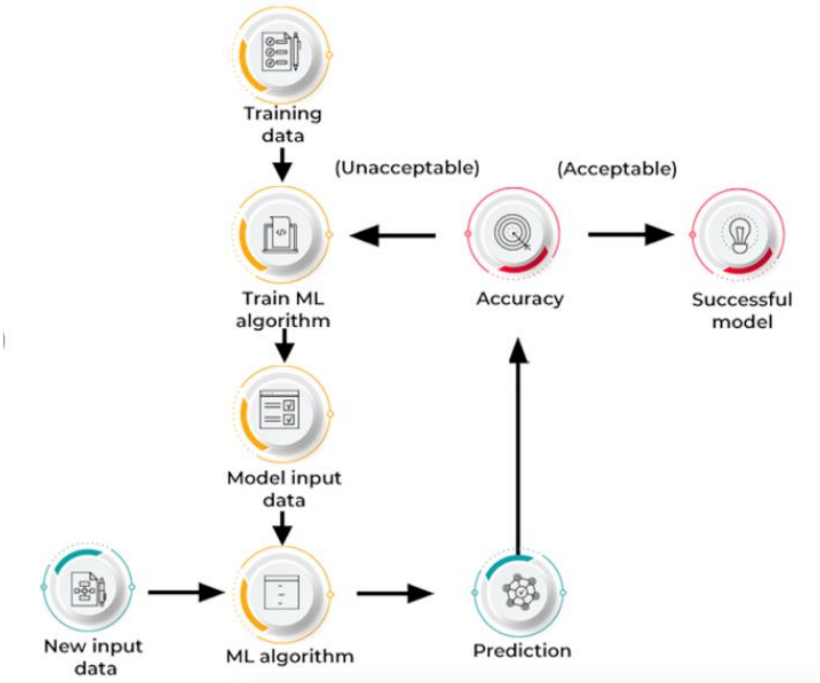

Figure 1:How machine learning works. (researchgate.net)

1.3. When can we used machine learning?

The first step to determine if machine learning methods to use is specifying one's research question. Using Hernán and colleagues' framework, there are three major data science research tasks: description, prediction, and causal inference (Hernán et al., 2019).

Machine Learning can be used to these three tasks. But statistical methods may prove to be sufficient and depending on specific research.

1.3.1 Description

Using data to produce a quantitative summery of factors is the goal of description tasks. Description tasks very useful in unsupervised machine learning methods because, their aim is to identify relations in data without having a measured outcome.

1.3.2 Prediction

Second major data science task is prediction because, it involve a huge range of goals. In traditional methods can be used to test hypothesis by statistical significance of association between predictors and the outcome.

Example: Odds Ratio

1.3.3 Causal Inference

Causal inference involves for estimating effects by comparing outcomes among exposed the counterfactual outcomes if they had instead not been exposed.

Example: estimating the suicide rate

In this type of question, clinical psychologists use a RCT (Randomized Controlled Trail) design and traditional statistical methods.

These all Above research tasks can be proved using Hernan and colleagues' framework and there are some common uses of machine learning listed below.

1. For automate a process
2. For monitor future events
3. For gathering knowledge from data

1.4. Types of machine learning

Machine learning algorithms are typically classified into three broad categories.

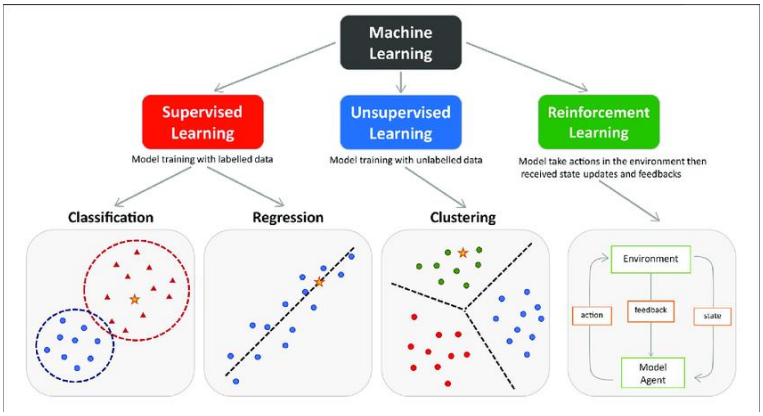

Figure 2:Types of Machine Learning (researchgate.net,2021)

For supervised learning problems, the training data comprises examples of the input vectors along with their corresponding target vectors.

For unsupervised learning problems, no targets are defined so that the training data consist of only a set of input vectors.

Reinforcement learning is to learn how to act or behave in a given situation for given reward or penalty signals.

2. SUPERVISED LEARNING

We've been using a more recent language for machine learning. The inputs are often referred to in statistical literature as predictors, a term that we will use interchangeably with inputs, and more classically as independent variables. The term features, as well as those that we are using, have been given preference in the literature for pattern recognition. The outputs are called the responses, or classically the dependent variables.

Supervised learning learns patterns and relations between input and output data, it defined as its use of labeled data. A labeled data is a dataset with many examples. Algorithms for supervised learning use the dataset the relationship between features and target. We call this process Fitting or training.

There are two types of supervised learning algorithms.

1. Classification
2. Regression

2.1. Classification

Classification means the algorithms learn from the data to predict an outcome or event in future.

Example: Bank Credit History

Image and speech recognition

Credit scoring

2.2. Regression

Regression means where algorithms learn from the data to predict continuously.

Example: House lot size

Estimate house's price

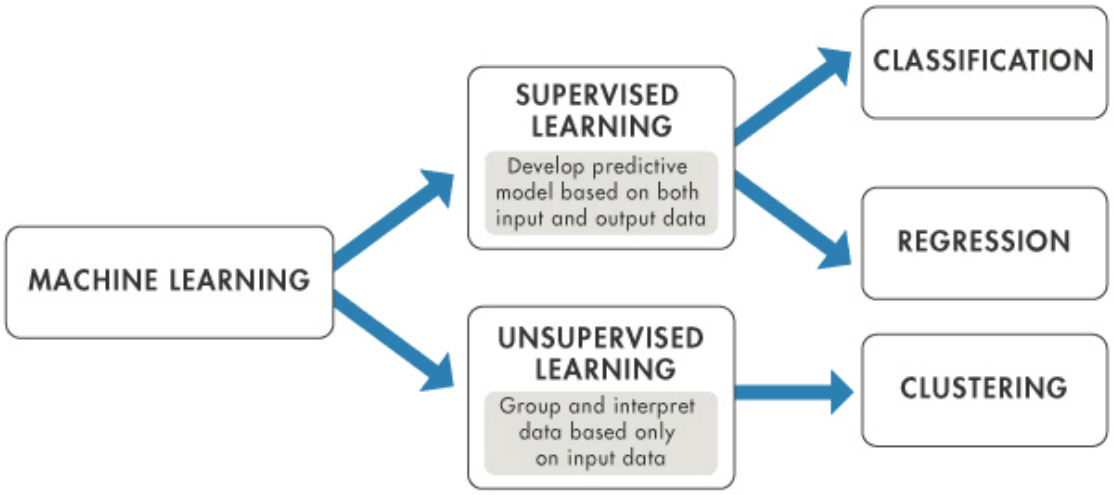

Figure 3:Supervised Learning and unsupervised learning Algorithm Types (Mathworks.com.2019).

3. UNSUPERVISED LEARNING

Unsupervised learning examines how an unlabeled dataset can be used to infer a function that defines an underlying pattern. The purpose of unsupervised learning aims to model the data's underlying structure or distribution to get further insight into the data. Clustering is an unsupervised learning technique. (Matloob et al., 2021).

Unsupervised learning is learning without a teacher. One basic thing that you might want to do with data is to visualize it. Sadly, it is difficult to visualize things in more than two (or three) dimensions, and most data is in hundreds of dimensions (or more). Dimensionality reduction is the problem of taking high dimensional data and embedding it in a lower dimension space. Another thing you might want to do is automatically derive a partitioning of the data into clusters. (Hal Daume III ,2017).

3.1. Clustering

Clustering is the most common unsupervised learning technique, it used for exploratory data analysis to unmask hidden data pattern or grouping in data.

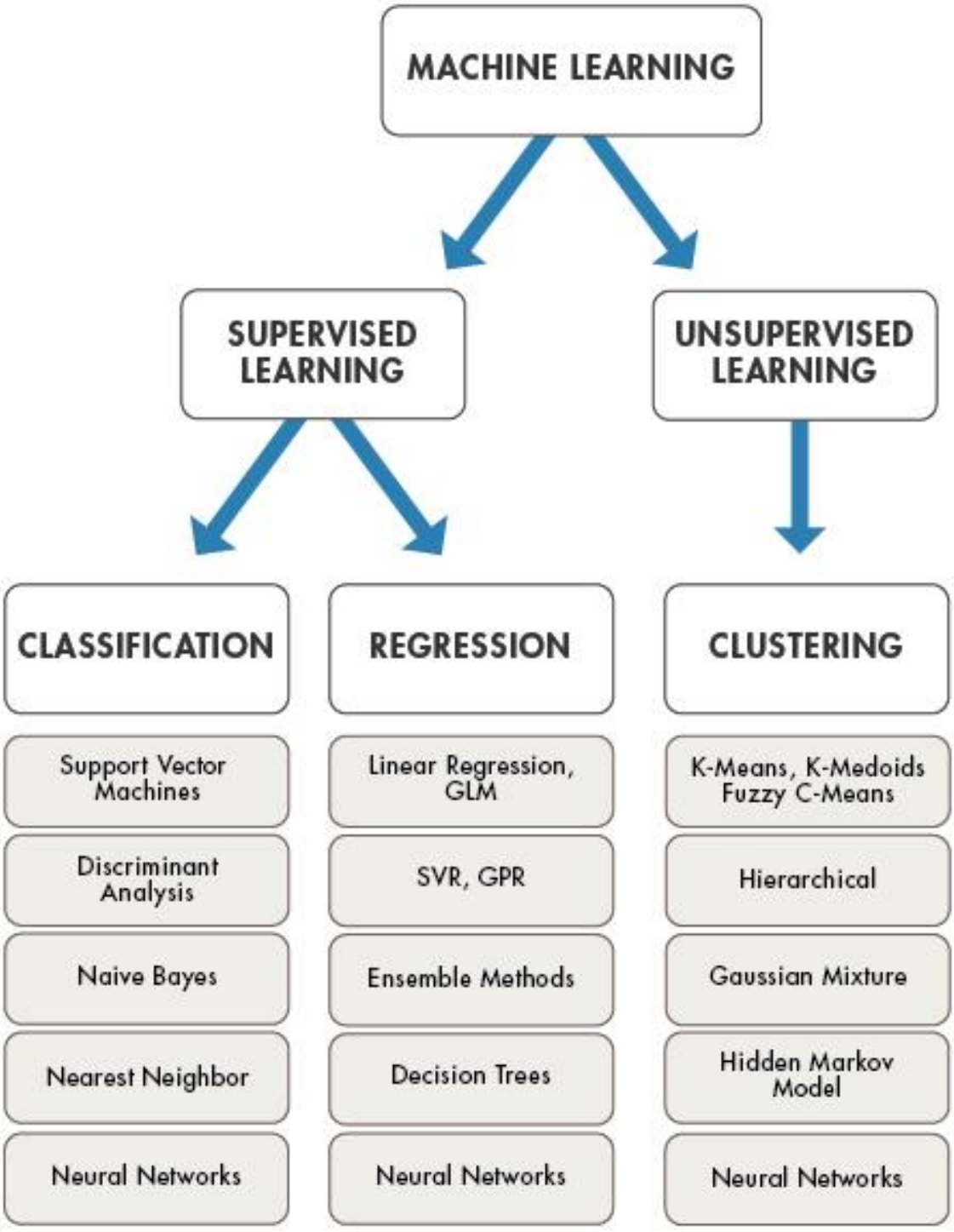

Figure 4: Supervised and Unsupervised Types Models (Mathworks.com. 2019).

4. SUPERVISED LEARNING VS UNSUPERVISED LEARNING

Supervised Learning	Unsupervised Learning
Labeled dataset	Unlabeled dataset
Highly Accurate	Less Accurate
It is time-consuming, and the labels for input and output variables need knowledge	Unless human interaction is used to check the output variables, unsupervised learning approaches can produce radically erroneous findings.
Simpler Method	Computationally Complex
No of classes known	No of classes unknown
Requires human involvement	Functions without human assistance

Table 1: Supervised Learning Vs Unsupervised Learning

5. ADVANTAGES AND DISADVANTAGES OF SUPERVISED & UNSUPERVISED LEARNING

Supervised learning		**Unsupervised learning**	
Pros	**Cons**	**Pros**	**Cons**
Predictive Power	Limited Generalization	Discovering Hidden Patterns	Lack of Ground Truth
Availability of Labeled Data	Dependence on labelled data	Unlabeled Data	Difficulty in Interpretation
Flexibility	Difficulty in Handling High-Dimensional Data	Data Preprocessing	Scalability
Iterative Improvement	Lack of Adaptability to Changing Data	Scalability and Adaptability	Overfitting and Noise Sensitivity

Table 2: Pros and Cons of Supervised & Unsupervised learning

CONCLUSION

In conclusion finally an evaluation is conducted to cover an overview of machine learning as well as the strategies that have been used in this field and the two types of primary methods such as supervised learning and unsupervised learning.

This research examined about the differences between supervised and unsupervised learning. Each approach has advantages and disadvantages.

The future of machine learning demonstrates that it will become progressively applicable across various industry verticals.

In future, machine learning could be able to help us with infinite supply of accurate insights and automated knowledge.

REFERENCES

El Naqa, I., Murphy, M.J. (2015). What Is Machine Learning?. In: El Naqa, I., Li, R., Murphy, M. (eds) Machine Learning in Radiation Oncology. Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1 [Accessed 20 Feb. 2024].

Mitchell, T. (1997). *Machine Learning*. [online] Available at: https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf [Accessed 20 Feb. 2024].

Jiang, T., Gradus, J.L. and Rosellini, A.J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, 51(5), pp.675–687.: https://doi.org/10.1016/j.beth.2020.05.002[Accessed 21 Feb. 2024].

Matloob, I., Khan, S.A., Hussain, F., Butt, W.H., Rukaiya, R. and Khalique, F. (2021). Need-Based and Optimized Health Insurance Package Using Clustering Algorithm. *Applied Sciences*, 11(18), 8478. Available from https://www.mdpi.com/2076-3417/11/18/8478 [Accessed 21 Feb. 2024].

Hastie, T., Tibshirani, R., Friedman, J. (2009). Overview of Supervised Learning. In: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7_2 [Accessed 22 Feb. 2024].

Mathworks.com. (2019). *VisibleBreadcrumbs*. [online] Available at: https://www.mathworks.com/help/stats/machine-learning-in-matlab.html [Accessed 22 Feb. 2024].

Mathworks.com. (2016). [Image] Available at: https://www.mathworks.com/help/stats/machinelearningtypes.jpg [Accessed 23 Feb. 2024].

Mathworks.com. (2024). [Image] Available at: https://www.mathworks.com/help/stats/machinelearning_supervisedunsupervised.png

[Accessed 23 Feb. 2024].

Detecting loneliness in people using technology - Scientific Figure on ResearchGate.[Image] Available from: https://www.researchgate.net/figure/How-does-machine-learning-work-Sourced-from-10_fig1_366160395 [Accessed 23 Feb. 2024]

www.oreilly.com. (n.d.). *1. Introduction to Clustering - The Unsupervised Learning Workshop [Book]*. [online] Available at: https://learning.oreilly.com/library/view/the-unsupervised-learning/9781800200708/B15923_01_Final_RK.xhtml#_idParaDest-25.

[Accessed 26 Feb. 2024]

logicmojo.com. (n.d.). *Supervised and Unsupervised Learning - Logicmojo*. [online] Available at: https://logicmojo.com/supervised-and-unsupervised-learning#:~:text=Supervised%20and%20Unsupervised%20learning%20play

[Accessed 28 Feb. 2024].