IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Overview of Machine Learning

Nitesh Sharma
B. Tech (CSE) Student
Dronacharya College of Engineering, Gurgaon, Haryana, India

Abstract: *The field of machine learning is introduced at a high level of abstraction. The concepts of supervised and unsupervised learning, regression, and classification are all discussed. A major guiding notion of learning is the balance between bias, variance, and model complexity. The neural network (feed-forward and recurrent), support vector machine, random forest, self-organizing map, and Bayesian network are all examples of models that machine learning can develop. The core ideas of partitioning a dataset into training, testing, and validation sets, as well as performing cross-validation, are addressed next. The importance of the domain expert in keeping the project real is discussed next, followed by evaluating the model's goodness. The chapter ends with some practical advice on how to perform a machine learning project.*

Keywords: Natural language processing (NLP), Database, Computer vision, Supervised learning, Unsupervised learning, Reinforcement learning, Neural network, Overfitting.

I. INTRODUCTION

Machine Learning has become a buzzword in recent years, possibly due to the large amount of data produced by applications, the rise in processing power in recent years, and the development of better algorithms. Machine Learning is employed in a variety of ways, from automating mundane chores to providing insightful insights, and companies across the board are attempting to reap the benefits. It's possible that you're already using one. Consider a smart home assistant like Google Home or a wearable fitness tracker like Fitbit. However, there are many more applications of machine learning.

- Prediction — Machine learning can be utilised in prediction systems as well. In the case of the loan, the system will need to classify the available data into categories in order to determine the chance of a fault.
- Image recognition – Machine learning may also be used to recognise faces in photos. Each person in a database of numerous people has their own category.
- Speech Recognition - is the process of converting spoken words into text. It's utilised in voice searches, among other things. Speech dialling, call routing, and appliance control are all examples of voice user interfaces. It can also be used to enter simple data and format documents.
- Medical diagnostics – ML has been taught to spot malignant tissues.

II. SCOPE OF MACHINE LEARNING

We should expect its importance and areas of use to continue to rise, and the more advanced the technology grows, the more widely it will be used. It's already widely utilised in facial recognition and picture recognition in general: nudity filters, document processing, and even surveillance, as well as in content recommendation engines like Instagram, Youtube, Netflix, and Pinterest. As they are given datasets to study and learn, AI systems are increasingly coming up with new antibiotics that are meant to treat specific ailments. It may be possible in the future to further integrate AI into healthcare, using it to diagnose patients and provide treatment for their unique ailment.

We will surely see the rise of self-driving vehicles, since we are now on our approach to entirely automating our transportation systems, with AI systems capable of complicated decision-making much like humans. Machine learning is currently progressing in many parts of our lives, and it will only become more integrated as AI technology advances.

HOW MACHINE LEARNING WORKS

- A Decision Process: Machine learning algorithms are used to produce predictions or classifications in general. Your algorithm will generate an estimate about a pattern in the data based on some input data, which can be labelled or unlabelled.

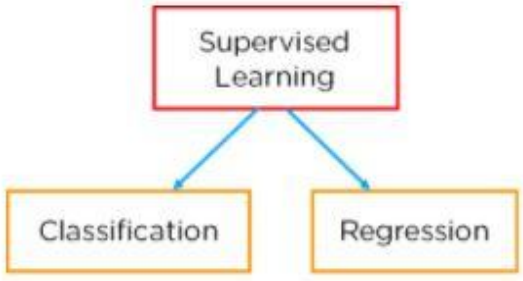
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

- An Error Function: An error function is used to assess the model's prediction. An error function can be used to compare the model's accuracy if there are known examples.
- **An Model Optimization Process**: Weights are modified to lessen the difference between the known example and the model estimate if the model can fit better to the data points in the training set. The algorithm will repeat this assess and optimise process, changing weights on its own until it reaches a certain level of accuracy.

IV. Machine Learning Meathods

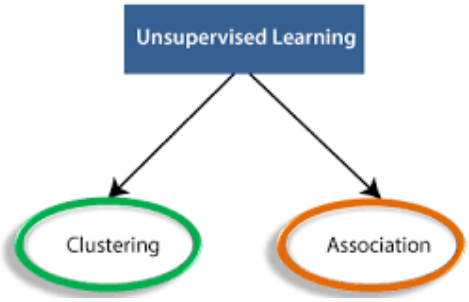
4.1 Supervised Machine Learning

Supervised learning, also known as supervised machine learning, is defined as the use of labelled datasets to train algorithms that reliably categorise data or predict outcomes. As more data is introduced into the model, the weights are adjusted until the model is properly fitted. This happens during the cross validation process to verify that the model does not overfit or underfit. supervised learning can be used to solve a variety of real-world issues at scale, such as spam classification. Create a folder separate from your email .Neural networks, nave bayes, linear regression, logistic regression, random forest, support vector machine (SVM), and other approaches are used in supervised learning.



4.2 Unsupervised Machine Learning

Unsupervised learning, also known as unsupervised machine learning, analyses and clusters unlabelled datasets using machine learning techniques. Without the need for human intervention, these algorithms uncover hidden patterns or data groupings. Its capacity to find similarities and contrasts in data makes it an excellent choice for exploratory data analysis, cross-selling techniques, consumer segmentation, and picture and pattern recognition. Two typical approaches for lowering the number of features in a model through the dimensionality reduction process are principal component analysis (PCA) and singular value decomposition (SVD). Unsupervised learning employs neural networks, k-means clustering, probabilistic clustering techniques, and other algorithms.

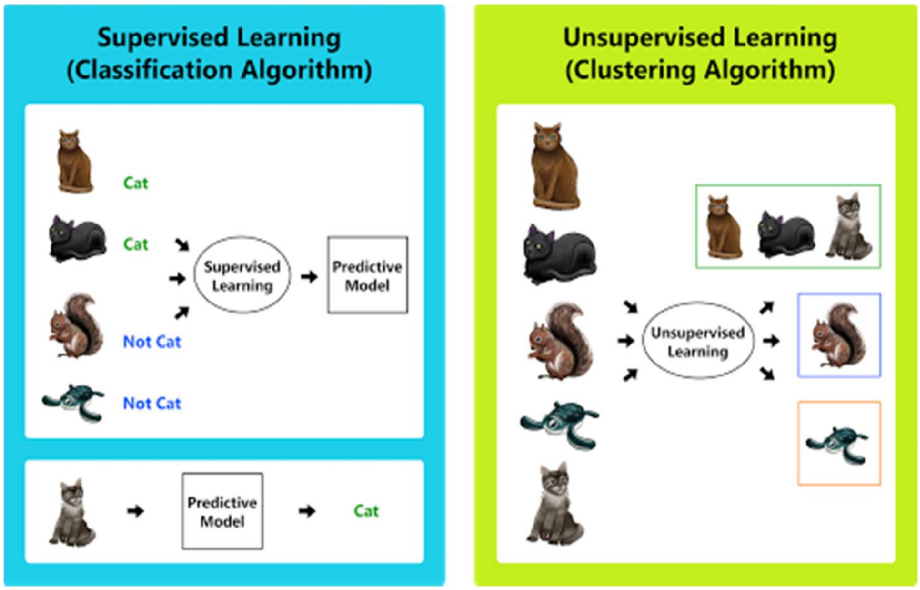


Clustering

A clustering problem is one in which you wish to find the data's inherent groupings, such as classifying customers based on their purchasing habits.

An association rule learning problem is one in which you wish to find rules that explain substantial chunks of your data, such as recommendations for people who like X products and also like Y products that are comparable to X.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252



4.3 Semi-Supervised Learning

Semi-supervised learning is an excellent compromise between supervised and unsupervised learning. During training, it uses a smaller labelled data set to assist categorization and feature extraction from a larger, unlabelled data set. When there isn't enough labelled data to train a supervised learning algorithm, semi-supervised learning can help (or not being able to afford to label enough data).

V. Challenges of Machine Learning

Machine learning technology has surely made our lives easier as it improves. However, incorporating machine learning into enterprises has created a variety of ethical questions about AI technology. Here are a few examples:

5.1 Technical Singularity

While this topic has gotten a lot of press, many experts remain unconcerned about AI exceeding human intelligence in the near or far future. Nick Bostrum describes superintelligence as "any mind that far excels the best human brains in nearly every discipline, including scientific innovation, general knowledge, and social abilities." Although artificial intelligence and superintelligence are not yet a reality in society, they provide some exciting problems when we consider the implementation of autonomous systems such as self-driving cars. It's impossible to expect a self-driving car to never cause a collision, but who is responsible and liable in those situations? Should we continue to seek self-driving cars?

5.2 AI Impact on Jobs

While job loss is a common issue associated with artificial intelligence, this concern can probably be reframed. The market need for specific job roles shifts with each disruptive new technology. In the automotive industry, for example, several manufacturers, such as GM, are moving their focus to electric car production to match with green objectives. The energy business will not go away, but the source of energy will change from fossil fuels to electricity. Artificial intelligence should be seen in the same light, as it is moving job demand to other industries. Individuals will be required to assist in the management of these systems as data increases and changes on a daily basis.

5.3 Privacy

Data privacy, data protection, and data security are often considered in conjunction with privacy, and these issues have allowed policymakers to make progress in recent years. For example, GDPR law was enacted in 2016 to protect people's personal data in the European Union and the European Economic Area, allowing them more control over their data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Individual states in the United States are developing rules, such as the California Consumer Privacy Act (CCPA), that compel firms to notify customers when their data is collected. This new legislation has compelled businesses to reconsider how they retain and use personally identifiable information (PII). As a result, organisations are increasingly prioritising security efforts in order to minimise any weaknesses and potential for spying, hacking, and cyberattacks.

5.4 Components of a Machine Learning System

A. Data Sources

Traditionally, your Software Engineer was in charge of computing an answer (or output) from receiving data. To put it another way, their task has been to create an explicit set of instructions that a computer can execute repeatedly. When it comes to algorithmic development, the Software Engineer may only need a few instances (and possibly a few more "edge-cases") to figure out these procedures. Although machine learning algorithms do this in theory, the approach is more statistical. Data is input into the algorithms, which translate it into a likely outcome. When the prediction is incorrect, they undertake tiny self-corrections. To make matters even more convoluted, features regarded useful for correctly predicting the outcome may come from a variety of places. Understanding the effort required to extract and then combine data from numerous sources can give you an indication of the technical effort required. Data collection is necessary for training, forecasting, and retraining as the outside world changes.

B. Data Storage

Because the available signals for the machine learning system come from a variety of sources and may be reused or processed further, creating an intermediate storage system to keep these extracted "features" is typically a useful solution. Some businesses decide to create a "data lake," which is a storehouse of extracted features. This unstructured data repository can be updated on a regular basis and used to produce features for a wide range of machine learning systems. This structure is advantageous since it is not as firmly connected to a schema as typical relational databases, allowing for easier adaptation of modifications as new features are developed. When a vast amount of data must be ingested and processed. When a huge amount of data needs to be ingested and put together for the machine learning system, it may be beneficial to create a structured data warehouse to store the prepared data.

REFERENCES

[1]. Cohen AM, Hersh WR, Peterson K, et al. Using automated citation classification to save workload in systematic review preparation. J Am Med Inform Assoc. 2006;13(2):206–19. [PMC free article] [PubMed]

[2]. Dalal SR, Shekelle PG, Hempel S, et al. A Pilot Study Using Machine Learning and Domain Knowledge To Facilitate Comparative Effectiveness Review Updating, Methods Research Report. Rockville, MD: Agency for Healthcare Research and Quality; Sep, 2012. Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I). AHRQ Publication No. 12-EHC069-EF. [PubMed]

[3]. Danz M, Rubenstein L, Hempel S, et al. Identifying quality improvement intervention evaluations –is consensus achievable? Qual Saf Health Care. 2010;19(4):279–83. [PubMed]

[4]. Deerwester S, Dumais S, Furnas G, et al. Indexing by Latent Semantic Analysis. J Am Soc Inf Sci. 1990;41(6):391–407.

[5]. EPOC: Cochrane Effective Practice and Organisation of Care Group. Welcome to the EPOC Ottawa Website. [Accessed April 2012]. http://epoc.cochrane.org.