Role of Mathematics in Data Science - Machine learning

Rajendrakumar Vilas Thorat

Assistant Professor, Department of Computer Science, Sarhad College of Arts, Commerce and Science, Katraj, Pune

Email: rajendrathorat99@gmail.com

ABSTRACT

The mathematical concepts help with identifying patterns and algorithmic development. Mathematics is an important part of many modern scientific areas. It is also a field of major significance in data science. Mainly because of the concepts they propose and the tools they allow, statistics and arithmetic have such an essential role in data science. A specific type of mathematical reasoning is needed for analysing data, in addition to the basic concepts of calculus, discrete mathematics, and linear algebra. To understand how to apply such algorithms in data science, one must have a solid understanding of many different statistical and probability theory ideas. Modern data science methods have a strong mathematical basis for almost all of their techniques, including machine learning.

Keywords: Mathematics for Data Science, Machine Learning, Deep Learning, Linear Algebra, Calculus.

Introduction

The applied disciplines also the fields of statistics, operations research, mathematics, and computing, all have an effect on data science as a branch of science. In the branch of data science, mathematical and statistical concepts are crucial because they act as the fundamental building blocks for all machine learning algorithms. Everything in the world around us, including forms, patterns, colors, sequences, and even the number of petals in a flower, is based on mathematics. The use of mathematics in every area of our everyday lives. Becoming a data scientist requires more than just having an excellent understanding of programming languages, machine learning techniques, and data-driven methods; data science requires more than just these disciplines.

The core of each and every machine learning algorithm is mathematics. Since deep learning is a part of machine learning, the following holds true for all data science techniques, including optimization, supervised learning, deep learning, and unsupervised learning. These algorithms improve our ability to extract information from the data. Artificial intelligence derives from the fact that the computer system that applies these algorithms to the data set looks intelligent. We use a programming language to develop these algorithms (mostly libraries are available). Machine learning provides a model that generates predictions by using an algorithm to learn from data. It aids in finding out how something works and why one model is superior to another. Machine learning is subject to an underlying mathematical limitation.

Mathematics for Data Science: The fundamental study of technology is mathematics. A strong foundation in mathematics is required for many of the modern digital technologies of today, such as cyber security, artificial intelligence, data science, machine learning, and massive data sets. Three skill sets form the base of data science: exposure to computer technology, background in mathematics or statistics and business or related skills. Graduated in mathematics and is able to understand the mathematics behind the models and to develop on top of it is improved.

ML suggests several models that use various detection, prediction, and classification techniques. Predictions may involve what an object in an image looks like, such as what the nation's upcoming fuel price might be or which pair of treatments will work best to treat an illness. This is an essential mathematical requirement for ML. It's used to understand the workings of everything or the superiority of one model over another. Calculus, linear algebra, probability, and statistics are the basis of algorithmic learning.

Even though mathematics is applied everywhere, if I had to single out one area where mathematics has had a big impact on data science, it would be Wing's work to develop a scientific understanding of the elements that go into making learning strategies like deep learning effective and, more importantly, how to keep them that way. We study the properties of deep neural networks, which can be conceptualized as a nonlinear mapping from input space to a reduced representation or a classification.

When comparing data science to computer science, the author (Neethumol K. et.al (2021) argues that most of the importance of math and statistics may be attributed to the concepts they propose and the tools they enable.

Linear Algebra: The area of mathematics known as "linear algebra" deals with linear equations and includes topics like mappings, vector spaces, and the study of lines and planes, all of which are necessary for linear transforms. The most essential mathematical skill in machines is the ability to use linear algebra, which is important to all branches of mathematics. It is an area of mathematics that mainly deals with the concept that matrices and vectors are the language of data and that linear programming acts as the logic of data. The purpose of linear algebra is to solve linear equation problems.

In data science, it is crucial to represent data, and one method of doing so is using matrices. Thus, it is crucial to understand ideas in matrices. Data also includes a number of variables; one should be aware of how many of these are actually utilized or significant. It is crucial to comprehend matrices and the ideas of linear algebra in order to understand data.

Dimensionality reduction is the method of reducing the number of parameters required to obtain a compact dimension representation of a data set through the use of principal component analysis and singular value decomposition. All of the neural network algorithms represent and process network structures using methods from linear algebra.

All of the information can be represented by matrices. Data that represents a model with variables from multiple equations may be called a matrix of data. The idea of future rank can be utilized to determine how many linear correlations there are between the attributes simply by examining the data.

Single value decomposition is a factorization in linear algebra that extends the Eigen decomposition of a square matrix to any other matrix. It can be used for real or complex matrices. Singular Value Decomposition splits matrices into three distinct matrices, allowing for matrix manipulation. Shearing, rotation, and scaling will result in these matrices (Neethumol K.G1 -March-2021).

Calculus: Calculus is a necessary foundation for data science. The main application of calculus is in techniques for optimization. It is extremely difficult to push for a comprehensive understanding of machine learning in the data science field without having a strong foundation in calculus. You may model artificial neural networks mathematically and improve their accuracy and performance by using calculus. Differential calculus and integral calculus are the two categories into which calculus falls. Calculus Differential The study of differential calculus looks at how quickly quantities change. In optimization techniques, when the goal is to find the minimum to attenuate the error function, derivatives are most commonly used to find the maximum and minima of the functions. Understanding that partial derivatives represent another crucial derivative notion.

Calculus is a part of mathematics that studies the rate at which quantities vary. The two areas of calculus are integral and differential calculus. Differential calculus divides a large chunk into smaller portions to see how they change, while integral calculus joins the little pieces to find the entire amount. Databases can be thought of as calculus since relational algebra, commonly referred to as relational calculus, is the foundation of relational database systems.

Derivatives are most frequently employed in optimization approaches to discover the maxima and minima of the functions when it is important to find the minima in order to attenuate the error function. Another important concept related to derivatives is the understanding that partial derivatives are used to construct back propagation in neural networks. The chain rule is another essential concept for backpropagation computation. We combine back propagation and error function minimization with the differential theory of games for generative adversarial neural networks (Neethumol K.G1 -March-2021).

Improving the efficiency of algorithms used for deep learning and machine learning is the main use of calculus. This instrument is used to provide effective and prompt solutions. Algorithms such as Adam, Rms Drop, and Adadelta, as well as optimizers such as Gradient Descent and Stochastic Gradient DescentSGD, use calculus. For the creation of various deep learning and machine learning models, data scientists use calculus very widely.

The main use of calculus by data scientists is to build a large number of deep learning and machine learning models. Calculus helps in data optimization and data output generation by extracting valuable information from data. However, calculus can be used to build pattern-learning models (Dr. Suresh Dara 2022).

Conclusion:

Mathematics is an important field to focus on for those interested in machine learning, and a solid foundation in the topic is needed. All machine learning concepts, whether explicit or tacit, short algorithms you develop, and approaches to problem solving you use have a mathematical foundation. The mathematical ideas used in machine learning are derived from the basic mathematical concepts that are studied in the

eleventh and twelfth grades. Even though we now understand the subject theoretically, machine learning ultimately uses the previously learned mathematics for real-world applications. The ideal technique when implementing a machine learning algorithm is to identify a use case, solve it, and understand the underlying mathematics.

Mathematics plays a role in machine learning and artificial intelligence to help create unique and useful models. Statistics and mathematics skills are required for jobs in digital marketing. Lastly, knowledge of mathematics and computational techniques is required for modern data science roles.

References:

[1]. Neethumol K.G1 , Priya Prakash (2021), The Role of Mathematics and Statistics in the Field of Data Science, International Journal of Scientific & Engineering Research Volume 12, Issue 3,ISSN 2229-5518, PP 8-11.

[2]. Tamara Kolda 1,2,3 , Mathematics: The Tao of Data Science, Harvard Data Science Review • Issue 2.3, Summer 2020.

[3]. Dr. Suresh Dara*1, Subham Surmara*2, Sai Kiran Reddy*3, Kamala Vani*4, Aditya Sai Varma*5, ROLE OF MATHEMATICS IN MACHINE LEARNING, International Research Journal of Modernization in Engineering Technology and Science, e-ISSN: 2582-5208, PP 2543-2548.

[4]. Dr.M.Madhavilata, Role of Mathematics in Data Science, © 2021 JETIR May 2021, Volume 8, Issue 5, (ISSN-2349-5162),PP f691-f692.

[5]. Anjali Nair, Ambika Biradar, Dr. Radhika Menon, MATHEMATICS FOR DATA SCIENCE, Volume - 11 | Issue - 07 | July - 2021 | . PRINT ISSN No 2249 - 555X | DOI : 10.36106/ijar, PP 69-70.

Cite this Article: