

Role of Mathematics in Data Science

Dr.M.Madhavilata

Department of Humanities and Mathematics

G.Narayanamma Institute of Technology and Science(For Women),Shaikpet,Hyderabad,India

ABSTRACT

Mathematics is new Engineering. Mathematics is not only used in theoretical concepts but also applied in areas like image compression, speech recognition. Math builds an ability to look at problems differently and solve them, making it suitable for a career in Analytics. Applied Mathematics, Statistics and Probability are in big demand in Data Science, AI & ML. In this Paper Mathematical Concepts that are required for Data Science are considered.

Keywords— *Data Science, Machine Learning*

I. INTRODUCTION

Data Science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data Science is related to data mining, Machine Learning and Big Data.

Math is foundational science for technology. Many of today's leading digital technologies -artificial intelligence, machine learning, data science, big data, cyber security-need strong foundational knowledge of mathematics. Data Science is based on three skill sets-a background in math or statistics, exposure in computer science, and business or domain knowledge. A math graduate adds value in the ability to understand the mathematics behind the models and innovate on top of that.AI and Deep learning solutions can be implemented using software but mathematics is needed to understand the inner workings of these solutions. Mathematics helps create unique and more effective ML models.

II. MATHEMATICS BEHIND DATA SCIENCE

This paper is about mathematical side of data analytics. Mathematics gives the fundamental ideas that underlie these machine learning algorithms. So the fundamental mathematical topics that are very important from Data Science perspective which are also said to be three pillars of Data science are

1. Linear Algebra
2. Probability and Statistics
3. Optimization

Linear Algebra: To understand the basic ideas in data science it is necessary to have fundamental grounding in math principles. This is an essential branch of Mathematics for understanding how Machine Learning algorithm works on a stream of Data to create insight. Everything from Facebook to Spotify transferring data involves matrices and matrix algebra. Data representation is very important in data science and one way of representing is in matrix form. So concepts in matrices are very important to know Also data contains several variables, out of these one should know how many variables are really used or important .In understanding data it is very important to know matrices and concepts of linear Algebra.

Here are topics in linear algebra which are needed to understand Data Science

- **Matrices:** Matrices can be used to represent the data. Data matrix could be data which is representing the model where coefficient of several equations are there. Next **Rank** is the concept that can be used to identify the number of linear relationships between the attributes purely using data.
- Inner and outer product, matrix multiplication
- Spatial matrices: Square matrix, identity matrix, triangular matrix, symmetric matrix, Hermitian Matrix, Unitary Matrix
- Matrix Factorization concepts: LU decomposition, Gauss Jordan elimination, Solving linear system of equations
- Vector space,basis,span,orthogonality
- Eigen values, Eigen vectors, Diagonalization

- Dimensionality Reduction- Principal component Analysis, Singular value decomposition which is used to achieve a compact dimension representation of data set with fewer parameters .All neural network algorithms use linear algebra techniques to represent and process network structures.

Data science by its nature is not tied to a particular subject area and many deals with phenomena as diverse as cancer diagnosis and social behavioral analysis. This produces the possibility of a dizzy ice away of n-dimensional mathematical objects, statistical distributions, optimization objective functions etc.

Statistics

The importance of having a solid grasp over essential concepts of statistics and probability cannot be overstated. Many practitioners in the field consider classical (non- neural network). Machine learning to be nothing but statistical learning. This subject is vast, and focused planning is critical to cover the most essential concepts:

- Data summaries and descriptive statistics, central tendency, variance, covariance, correlation
- Basic probability: Basic definitions and concepts of probability, Expectation, conditional probability, Bayes' theorem.
- Probability Distributions
- Testing of Hypothesis
- Linear Regression and Multiple Regression
- Time Series Analysis

Optimization

Optimization is defined as a problem where you maximize or minimize a real function by systematically choosing input values from a allowed set and computing the value of the function. So it is always applied to get best solution.

Understanding the basic optimization techniques helps in Machine learning algorithms. Almost all machine learning algorithms can be viewed as solutions to optimization problem and it is interesting that even in cases where the original machine learning technique has a basis derived from other fields for example from biology and so on one could still interpret all of these machine learning algorithms as some solution to an optimization problem.

A basic understanding in optimization helps in :

- More deeply understand the working of machine learning algorithm
- Rationalize the working of algorithm and deep understanding in optimization helps in interpreting result.
- Depending on types of constraints we study optimization problems
 - Constrained optimization problem: in cases where constraints is given there and we have to have solution. Satisfying these constraints we call these constrained optimizing problem
 - Unconstrained Optimization problem: in these cases where constraint is missing we call them as unconstrained optimization problems.

III. CONCLUSION

Hence Mathematics helps create unique and effective models in Machine Learning and Artificial intelligence. Digital Marketing roles require people with knowledge of Math's and Stats .Finally Data Science jobs today require knowledge of both statistics and computing techniques.