IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 12, Issue 1, December 2021
Impact Factor: 5.731

Machine Leaning Fundamentals

Dhruvi Ashwin Tank
Department of Computer Engineering
Shri Bhagubhai Mafatlal Polytechnic, Mumbai, Maharashtra
dhruvitank2403@gmail.com

Abstract: *Machine Learning is one of the fastest-growing fields which has witnessed exponential growth in the technical world. But in this fast-growing field, the question is how to get started? Briefly this paper introduces various languages popular for machine learning and next it introduces a few math concepts that helps us understand what exactly is happening and to improve our model further we need to understand that. And finally, there is an overview of various IDEs that can be used to implement these languages for machine learning.*

I. INTRODUCTION

Machine learning is a very vast topic. As a beginner it is important to understand the building blocks of Machine Learning and how does it work. The learning process in machine learning is automated and improved based on the experiences of the machines throughout the process. Good quality data is fed to the machines, and different algorithms are used to build ML models to train the machines on this data. Depending on the type of data at hand and the activity that we have to automate, the machine learning algorithm is selected.

In traditional programming, we used to feed the input data and a well written and tested program into a machine to generate output whereas in machine learning, input file alongside its output is fed into the machine during the learning phase, and it generates a program for itself.

Machine learning has grown exponentially, and it will continue to grow in the future. This article will help you understand various methods and tools one can choose to learn machine learning. Each method requires you to know a programming language (e.g. python) and various math concepts. Almost all of ML is about applying concepts from statistics and computer science to data.

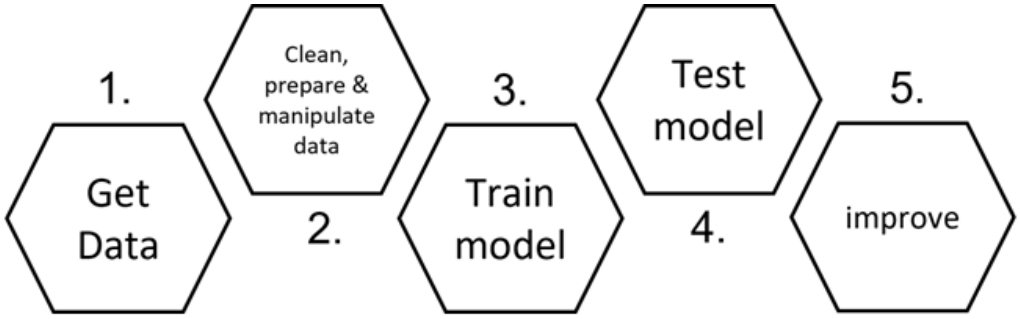

Figure 1: Steps in Machine Learning

II. CORE CONCEPT

There are three concepts that are at the core of machine learning: **data, a model, and learning.** Since machine learning is inherently data driven, data is at the core of machine learning. The goal of machine learning is to design general purpose methodologies to extract valuable patterns from data, ideally without much domain-specific expertise. To achieve this goal, we design models that are typically related to the process that generates data, similar to the dataset we are given. A model is said to learn from data if its performance on a given task improves after the data is taken into account. The goal is to find good models that generalize well to yet unseen data, which we may care about in the future. Learning can be understood as a way to automatically find patterns and structure in data by optimizing the parameters of the model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 12, Issue 1, December 2021
Impact Factor: 5.731

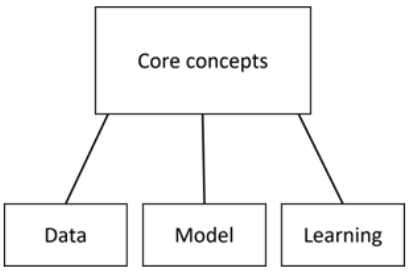

Figure 2

III. PROGRAMMING LANGUAGES

Python, R, C++, java, and JavaScript are the top 5 languages used by developers for machine learning. There is no best language for Machine Language as such it all depends on what we want to build. If you are a beginner than python can be more convenient because of rich set of libraries and ease of use. Most of the times developers use the same language that they are working with into machine learning, particularly in case they are to utilize it in projects adjacent to their past work, such as engineering projects for C/C++ developers or web visualizations for JavaScript developers.

3.1 Python

More than 60% of machine learning developers use python for development because it is easy to use, scalable and open source. Python has many packages and core libraries which makes it easy to use and helps the machines to learn effectively.

- Keras, TensorFlow, and Scikit-learn for machine learning
- NumPy for high-performance scientific computing and data analysis
- SciPy for advanced computing
- Pandas for general-purpose data analysis
- Seaborn for data visualization

3.2 R language

R is one of the most powerful machine learning platforms and it is used by the top data scientists in the world. A survey says around 31% developers prefer using R for machine learning. It is an open-source software. R is an interpreter. It provides a REPL environment where you can type in commands and see the output immediately. It can used to create and display graphics, to save and load state and to interface with other systems.

3.3 C++

C++ is an object-oriented programming language with faster run-time as compared to other programming languages hence is suitable for machine learning since fast and reliable feedback is essential in machine learning. However, it would be difficult to implement machine learning using C++ without understanding the algorithms first.

3.4 Java

Java is an extraordinarily valuable, fast, and dependable programming language that can be incredibly useful in ML for various reasons. Most of the well-known frameworks and tools utilized for Big Data are normally written in Java. This incorporates Flink, Hadoop, Hive, and Spark.

Designers consider the Java Virtual Machine as perhaps the best stage for ML and information science, as it empowers the engineer to compose code that is identical across various platforms. It additionally permits them to make custom instruments at a quicker speed and features a load of IDEs that help to improve overall productivity levels. Java has numerous libraries and tool accessible for Data Science and Machine Learning. For instance, Weka 3 is a completely Java-based workbench famously utilized for Algorithms in ML.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 12, Issue 1, December 2021
Impact Factor: 5.731

3.5 Javascript

JavaScript is supported by all modern mobile and desktop browsers. This implies JavaScript ML applications are ensured to run on most modern mobile and desktop browser. There are already various JavaScript machine learning libraries. An example is TensorFlow.js, the JavaScript version of TensorFlow machine learning and deep learning library.

One more significant use for JavaScript ML is model customization. Python support in mobile operating systems is still in the preliminary stages on the other hand there is already a rich set of cross-platform JavaScript mobile app development tools such as Cordova and Ionic.

IV. MATHS CONCEPTS

There are various reasons why the mathematics concepts are important for Machine Learning. There are so many high-level machine learning libraries that allow you to run a machine learning model with one line of code. All you would need is data. Machine learning is all about mathematics which helps to create an algorithm that can learn from the data provided to make an accurate prediction. The thing is mathematics really helps you understand what is happening and if you want to improve your model further you need to know what is happening.

There are 4 important concepts for machine learning those are linear algebra, statistics, probability, and calculus. Statistical concepts are a core part of every model and calculus helps us to learn and optimize a model, linear algebra is used to deal with a large dataset and probability helps to predict the livelihood of events that might occur.

4.1 Linear Algebra

Linear Algebra is the mathematical foundation that is used to resolve the problem of representing data and computations in machine learning models. All major phases of developing a model have linear algebra running in the background.

Linear algebra is used for data representation, the building blocks of ML models, that is data, needs to be converted into arrays before it can be feed to the models. Operations like matrix multiplication are performed on these arrays. Now, this returns the output which is also represented as a transformed matrix of numbers. Then next is word embeddings, it is nothing but representing large dimensional data with a smaller dimensional vector.

4.2 Statistics

Raw observations are data, but they are not knowledge or information. To understand the data used to train the machine learning model and to interpret the results of testing different machine learning models, we require statistics. Statistics is concerned with the collection, organization, analysis, interpretation, and presentation of data.

A. Types

1. **Descriptive Statistics:** It deals with understanding, analyzing, summarizing the data in the form of numbers and graphs.
2. **Inferential Statistics:** It makes an inference form a sample about the population. Its main aim is to draw some conclusions from the sample and generalize them for the population data.

4.3 Probabillity

Machine learning is all about developing predictive models from uncertain data. Working with imperfect or incomplete data means uncertainty. Probability represents the certainty factor. The extent of certainty about an uncertain event is measured by probability. Probability is important in approximating the analysis since machine learning is based on probable but not mandatory situations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 12, Issue 1, December 2021
Impact Factor: 5.731

4.4 Calculus

We can think of calculus as simply a set of tools for analyzing the relationship between functions and their inputs. Often, in machine learning, we are trying to find the inputs which enable a function to best match the data. Calculus plays an integral role in understanding the internal workings of machine learning algorithms, for e.g., the gradient descent algorithm that minimizes an error function based on the computation of the rate of change.

V. INTEGRATED DEVELOPMENT ENVIRONMENT (IDE)

These are few of the IDEs which support various languages used to implement machine learning. They are open-source IDEs.

Name	Supported languages	Platform	Special Features
Jupyter Lab	40+ programming languages (Including python, R, Julia, Scala)	Web based	• Share notebooks: Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer. • Interactive output: code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types. • Big data integration: Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.
Visual Code	These include C++, HTML, Java, JavaScript, Python, etc.	• Windows 7, 8, 9, 10 • Linux (Ubuntu, Debian, Fedora) • macOS 10.11+	• Syntax highlighting and bracket matching • Smart completions (IntelliSense) • Linting and corrections • Code navigation (Go to Definition, Find All References) • Debugging • Refactoring
Spyder	Python	• Windows • Linux • macOS • Web browser (using binder)	• An editor with syntax highlighting, introspection, code completion • Support for multiple IPython consoles • The ability to explore and edit variables from a GUI • A debugger linked to IPdb, for step-by-step execution • Available plugins: • Spyder notebook • Spyder Terminal • Spyder unittest.
TensorFlow	• Python • C++ • JavaScript • Java • R • Ruby, etc.	• Ubuntu 16.04 or later • Windows 7 or later • macOS 10.12.6 (Sierra) or later (no GPU support)	• Open-source Library. It is an open-source library that allows rapid and easier calculations in machine learning. • Easy to run. • Fast Debugging. • Effective. • Scalable.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 12, Issue 1, December 2021
Impact Factor: 5.731

			• Easy Experimentation. • Abstraction. • Flexibility
Eclipse	• Java • C/C++ • COBOL • JavaScript • C#, etc	• Windows • macOS • Linux • Web based	• Coding Shortcuts • Autocorrection • Refactoring • Diffing Files • Organizing Imports • Formatting Source Code

Table 1: Integrated Development Environment

VI. CONCLUSION

This study explored the basics of Machine Learning that will help you get started. This is just the tip of the iceberg, using this knowledge various models can be developed to make reliable predictions. There are various fields in which ML is most efficiently such as computer vision, business analytics, deep learning, working with big data etc. There are many other tools except for the ones mentioned above but which one to be selected relies on the user application and the needs.

REFERENCES

[1]. Brownlee J., Basics of Linear Algebra for Machine Learning. Machine Learning Mastery: Australia, 2018.
[2]. Sharma, S. and Sharma, S.K., A study on machine learning tools. machine learning, 2018. p.13.
[3]. Raschka, S., Patterson, J. and Nolet, C., Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. Information, 11(4), 2020, p.193.
[4]. Dangeti, P., Statistics for machine learning, 2017.
[5]. Unpingco, J., Python for probability, statistics, and machine learning, 2016.
[6]. Brownlee, J., Probability for machine learning: Discover how to harness uncertainty with Python, 2019