Ameet V Joshi

Machine Learning and Artificial Intelligence

Springer

Machine Learning and Artificial Intelligence

Ameet V Joshi

Machine Learning and Artificial Intelligence

Springer

Ameet V Joshi
Microsoft (United States)
Redmond, WA, USA

ISBN 978-3-030-26621-9 ISBN 978-3-030-26622-6 (eBook)
https://doi.org/10.1007/978-3-030-26622-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To everyone who believes in human intelligence to be mature enough to coexist with machine intelligence and be benefitted and entertained by it…

Foreword

I first got exposed to computing about 35 years back. It was the time when the term “microcomputers” was being replaced by the term “personal computers (PC).” Within a space of 5 years, there was this seismic shift that happened during which computing as a superspecialized domain with a total of only a few thousand people across the world working with these magical machines in cooled clean rooms suddenly became readily accessible to hundreds of millions. It dawned on everyone that computing was going to impact every aspect of business and eventually personal lives too. With that came the scramble to learn computing—whether it be about the basics of what a computer is or how to use applications like word processing or spreadsheets or engineers in every part of the world swarming to learn computer architectures and programming, algorithms, and data structures. If you were not learning any of those to some level of competency, you were heavily disadvantaged in your professions—regardless of the profession.

Fast forward 15 years to about 1995. PCs had evolved and become powerful and had taken over business computing with client-server architectures, and most people in developed countries had PCs of their own at work and home. Then, the terms Internet, web, Mosaic, Netscape, browser, webserver, HTML, and HTTP suddenly swept the computing landscape. The seismic shift that happened then entirely reframed how data and information became democratized at a scale several orders of magnitude greater than anything that preceded in the history of humanity. The Internet led to the creation of the smartphones which in turn exponentially increased the scale and reach of the Internet. The computing architectures, especially around data representation, storage, and distribution, went through a massive revolution. With that came the scramble to learn “Internet”—whether it was people of all ages learning the basics of what the Internet is, how to use a browser, and how to communicate with people electronically to authoring content as web consumables (emails, pages, sites, blogs, posts, tweets, etc.). Engineers swarmed to learn Internet architectures, web and mobile application development, distributed data architectures, etc. If you were not learning something along this spectrum, you were not only heavily disadvantaged professionally, you were probably disadvantaged socially.

Fast forward another 15 years. Advances in the Internet and mobile technologies lead to explosive growth in creation and aggregation of data. The massive aggregation of data and the virtually infinite computational capacity that goes with it evolved to being physically distributed and yet logically unified. This became the "Cloud." Organizing and making sense of this scale of data were not only beyond human capabilities, it was even beyond the capabilities of traditional algorithms. Traditional algorithms couldn't even scale to the "searching" through this sea of data let alone make sense of it. Yet, the easy availability of such vast volumes of data finally makes possible the long-held dream of computers transcending from "processing information" to actually "creating intelligence" by making sense of the data. Decades of research in ML and AI became re-energized and have made progress in leaps and bounds with the easy access to massive quantities of data and computation in the cloud. Thus, we are officially now in the age of "AI." Just as has happened in the PC and the Internet revolutions, there is now a realization that AI is going to transform not just computing but society and humanity in unprecedented ways. It is beginning to dawn on people that applicability of AI and machine learning go far beyond mundane things like facial recognition in their Facebook feeds. It is also beyond the flashy, sci-fi scenarios like self-driving cars and digital assistants. Businesses of every kind—health care, manufacturing, hospitality, financial, etc.—are in the process of massive transformation with data and AI at the center of it. In all facets of life, digital information without intelligence will become pointless. The next decade will be transformed by AI. The Internet age is now set to transition to the age of AI.

However, the "swarm" towards understanding AI is just beginning. In that context, I am thrilled that Dr. Ameet Joshi is coming out with this book on Machine Learning and Artificial Intelligence. When a topic is of such broad human interest and impact, one can never have enough sources of knowledge and learning. Compared to the kind of learning content that was created in the PC era or the Internet era, machine learning and AI related content is still sparse and somewhat narrowly targeted. This is an area which is still mysterious even for the technically savvy folks. While there is a need to address the full spectrum of audiences, there is an urgent need to demystify the space for folks who are comfortable with the math and science without necessarily being computer scientists or engineers. These are folks who are willing and capable of getting a good handle on ML and AI. They are folks who understand data and math at fairly advanced levels; they possibly understand traditional computing and algorithms—but they have been left off the island populated by the "priests of data science." They are eager to roll up their sleeves, dive in, and understand what ML and AI from its foundations. ML and AI learning is ripe for democratization. Dr. Joshi's book hits the exact right sweet spot on this. Most people who read this may not become data scientists—but all of them will gain a perspective on how ML and AI work and will be able to take it to whatever domain of work they will be involved in. They will be able to see every application of AI, be it a digital assistant they converse with or a spookily accurate targeting of an advertisement or price fluctuations of airline ticket prices in a very different light and with a much deeper understanding. The shroud of mystery around many things digital happening around us will seem more obvious and matter of fact. More importantly, I fully expect people who read this book to use that learning to look for applications in their future domains of work. For some set of readers, it may even intrigue them enough to become practitioners of ML and AI as a profession. There is definitely a huge demand for such professionals.

Dr. Joshi is a passionate practitioner himself. His passion shows through in this book. He has worked on broad set of ML and AI problems, and that practical experience has helped him identify and explain relatable scenarios and applications in this book. This is one of those books which likely is more than a onetime read. In all likelihood, the readers will want to revisit sections of the book to refresh their understanding or dig deeper in an area when they see real-world applications of the techniques described in this book. I think those are the best kinds of books to own. Even as an industry veteran, I enjoyed reading the book and will keep a copy handy.

Happy reading.

General Manager
Microsoft
Redmond, Washington, USA

Vij Rajarajan

Preface

One of the greatest physicists of all time and Nobel Laureate Dr. Richard Feynman was once asked by his peer to explain a property of Fermi-Dirac statistics that was very recently discovered then. Feynman, quickly said,

> Not only I will explain it to you, but I will prepare a lecture on it for freshman level.

However, quite unusually, after few days, he came back and admitted,

> I could not do it. I just could not reduce the explanation to freshman level. That means we really don't understand it.

It was quite a bold remark coming from even Dr. Feynman. However, apart from the topic of Fermi-Dirac statistics itself, it alludes to a very deep thought about our understanding of things in general. Freshman level here essentially meant something that can be derived directly using the first principles in mathematics or physics. This thought has always made me conscious to try and explain everything that I claim to understand using first principles, try to explain everything conceptually and not only using elaborate set of equations.

The area of *artificial intelligence* and *machine learning* has exploded over the last decade. With the widespread popularity, the core concepts in the field have been sometimes diluted and sometimes reinterpreted. With such exponential growth in the area, the scope of the field has also grown in proportion. A newcomer in the area can quickly find the topic daunting and confusing. One can always start with searching the relevant topics on the web or just start with Wikipedia, but more often than not, every single topic opens a rabbit hole with more and more new and unknown concepts, and one can get lost very easily. Also, most of the concepts in machine learning are deeply rooted in mathematics and statistics. Without solid background in theoretical mathematics and statistics, the sophisticated derivations of the theorems and lemmas can make one feel confused and disinterested in the area.

I have made an attempt here to introduce most fundamental topics in machine learning and their applications to build artificially intelligent solutions with an intuitive and conceptual approach. There would be some mathematical guidance used from time to time, without which the concepts would not be sufficiently clear, but I have tried to avoid complex derivations and proofs to make the content more accessible to readers who are not from strong mathematical background. In the process, as per Dr. Feynman, I also make sure I have understood them myself. As far as general mathematical and statistical requirements go, I would say that typical undergraduate level should suffice. Also, with proliferation and standardization of the machine learning libraries in open-source domain, one does not need to go that deep into mathematical understanding of the theory to be able to implement the state-of-the-art machine learning models, leading to the state-of-the-art intelligent solutions.

One of the main sources of confusion that arises when trying to solve a problem in the given application is the choice of algorithm. Typically, each algorithm presented here has originated from some specific problem, but the algorithm is typically not restricted to solving only that problem. However, choosing the right algorithm for the given problem is not trivial even for a doctoral fellow with strong mathematical background. In order to separate these two areas, I have divided these two areas into separate parts altogether. This will make the topics much easier to access for the reader.

I would recommend the reader to start with *Part I* and then choose *Part II* or *Part III*, depending on the needs. It will be ideal for a student to go sequentially through the book, while a newcomer to the area from professional background would be better suited to start with *Part III* to understand or focus on the precise application at hand and then delve into the details of the theory for the algorithms as needed in *Part II*. *Part IV* and *Part V* should follow afterwards. I have added sufficient references between the two parts to make this transition smooth.

In my mind, unless one can see the models in action on real data that one can see and plot, the understanding is not complete. Hence, following the details of algorithms and applications, I have added another part to cover the basic implementation of the models using free and open-source options. Completion of this part will enable the reader to tackle the real-world problems in AI with state-of-the-art ML techniques!

Redmond, WA, USA Ameet Joshi
March 2019

Acknowledgments

I would like to use this opportunity to acknowledge the people who had deep impact on the creation of this book. This, being the first ever book written by me, was quite a challenging task. There were quite a few times when I had almost given up. However, the continuous support and encouragement from my wife, Meghana, and sons, Dhroov and Sushaan, really helped me continue with the efforts and ultimately complete the book. I would also like to thank my father, Vijay, and brother, Mandar, for their feedback and support.

I would like to thank Mary James of Springer publications for the encouragement and support and understanding as I worked through the completion of the book.

The present book is an exercise in unification of the disparate areas in the field of machine learning to produce artificially intelligent experiences, and hence, the book essentially stands on the pillars of this knowledge that is created by the numerous extraordinary scientists and brilliant mathematicians over several decades in the past. So, I would like to thank them as well.

Last but not the least, I have used multiple quotes from some of the landmark movies produced over the last 60–70 years that depicted various scenarios of AI. These movies, like *2001: A Space Odyssey*, *Star Wars*, *WarGames*, *Terminator 2: Judgment Day*, and *The Matrix*, have showcased the power and scope of what AI can do and how it can change our lives. These cinematic interpretations express these extremely complex and technical concepts to laymen in highly influential manner and prepare them for what is coming in the future. Without them, the technological progress in this area would remain disconnected from people, so I would like to thank the creators of these movies.

Contents

Part I
Introduction

> This is your last chance. After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes. The question is, which pill would you choose?
>
> —Morpheus, "The Matrix"

Part Synopsis

This part introduces the concepts of Artificial Intelligence (AI) and Machine Learning (ML) in the modern context. Various scenarios where these concepts are used will also be discussed. Going further, this part also discusses the understanding, representation and visualization aspects of data that form the foundation to the next topics. This will serve as a good starting point to delve into the details of the techniques, applications and implementations.

Chapter 1
Introduction to AI and ML

1.1 Introduction

It is an understatement to say that the field of Artificial Intelligence and Machine Learning has exploded in last decade. There are some rather valid reasons for the exponential growth and some not so much. As a result of this growth and popularity, there are also some of the most wrongly used words. Along with the enormous growth of the field it has also become one of the most confusing fields to understand the scope of. The very definitions of the terms AI and ML have also become diluted and vague.

We are not going to try and define these entities with word play, but rather are going to learn the context around the origins and scope of these entities. This will make their meaning apparent. The roots of these words originate from multiple disciplines and not just computer science. These disciplines include pure mathematics, electrical engineering, statistics, signal processing, and communications along with computer science to name the top few. I cannot imagine any other area that emerges as a conflation of such wide variety of disciplines. Along with the wide variety of origins, the field also finds applications in even greater number of industries ranging from high-tech applications like image processing, natural language processing to online shopping to nondestructive testing of nuclear power plants and so on.

In spite of being such multi-disciplinary, there are several things that we can conceptually learn collectively and obtain clarity of its scope. That is the primary objective behind this book. I want to make the field accessible to newcomers to the field in the form of graduate level students to engineers and professionals who are entering the field or even actively working in this field.

A. V. Joshi, *Machine Learning and Artificial Intelligence*,
https://doi.org/10.1007/978-3-030-26622-6_1

1.2 What Is AI

Alan Turing defined Artificial Intelligence as follows: "If there is a machine behind a curtain and a human is interacting with it (by whatever means, e.g. audio or via typing etc.) and if the human feels like he/she is interacting with another human, then the machine is artificially intelligent." This is quite a unique way to define AI. It does not directly aim at the notion of intelligence, but rather focusses on human-like behavior. As a matter of fact, this objective is even broader in scope than mere intelligence. From this perspective, AI does not mean building an extraordinarily intelligent machine that can solve any problem in no time, but rather it means to build a machine that is capable of human-like behavior. However, just building machines that mimic humans does not sound very interesting. As per modern perspective, whenever we speak of AI, we mean machines that are capable of performing one or more of these tasks: understanding human language, performing mechanical tasks involving complex maneuvering, solving computer-based complex problems possibly involving large data in very short time and revert back with answers in human-like manner, etc.

The supercomputer depicted in movie 2001: A space odyssey, called HAL represents very closely to modern view of AI. It is a machine that is capable of processing large amount of data coming from various sources and generating insights and summary of it at extremely fast speed and is capable of conveying these results to humans in human-like interaction, e.g., voice conversation.

There are two aspects to AI as viewed from human-like behavior standpoint. One is where the machine is intelligent and is capable of communication with humans, but does not have any locomotive aspects. HAL is example of such AI. The other aspect involves having physical interactions with human-like locomotion capabilities, which refers to the field of robotics. For the purpose of this book, we are only going to deal with the first kind of AI.

1.3 What Is ML

The term "Machine Learning" or ML in short, was coined in 1959 by Arthur Samuel in the context of solving game of checkers by machine. The term refers to a computer program that can learn to produce a behavior that is not explicitly programmed by the author of the program. Rather it is capable of showing behavior that the author may be completely unaware of. This behavior is learned based on three factors: (1) Data that is consumed by the program, (2) A metric that quantifies the error or some form of distance between the current behavior and ideal behavior, and (3) A feedback mechanism that uses the quantified error to guide the program to produce better behavior in the subsequent events. As can be seen the second and third factors quickly make the concept abstract and emphasizes deep mathematical roots of it. The methods in machine learning theory are essential in building artificially intelligent systems.

1.4 Organization of the Book

I have divided this book into six parts, as described below.

Parts of the Book

1. Introduction
2. Machine Learning
3. Building end to end pipelines
4. Artificial Intelligence
5. Implementations
6. Conclusion

1.4.1 Introduction

The first part of the book introduces the key concepts in the field as well as outlines the scope of the field. There are multiple aspects discussed in these chapters that may appear disconnected, but put together they are all extremely important in holistic understanding of the field and inspiring the confidence. This part also discusses the preliminary understanding and processing of the raw data as is typically observed. This sets the foundation for the reader to understand the subsequent parts.

1.4.2 Machine Learning

Second part of the book deals with the theoretical foundation of the machine learning. In this part we will study various algorithms, their origins and their applications as well. Machine learning has united a vast array of algorithms that find their origins in fields ranging from electrical engineering, signal processing, statistics, financial analysis, genetic sciences, and so on. All these algorithms are primarily developed from the principles of pure mathematics and statistics, in spite of being originated in fundamentally different areas. Along with the roots, they also have one more thing in common: the use of computers to automate complex computations. These computations ultimately lead to solving problems that seem so hard that one would believe that they are solved by some intelligent entity, or *Artificial Intelligence*.

1.4.3 Building End to End Pipelines

After learning about the theory of machine learning, one is *almost* ready to start solving problems in real world. However, the small gap that still remains is addressed in this part. Putting together theory and application is the primary objective to be learned from this part. Invariably the machine learning system built to solve a real life problem always involves multiple components being operated in sequence, creating a pipeline of operations. This part discusses all the essential components of such pipeline that takes the raw data at the input and produces interpretable results at the output.

1.4.4 Artificial Intelligence

This is likely the most interesting part of the book that directly deals with problems that machine learning has solved and we experience in our day-to-day life. These applications represent real problems that have been solved in variety of different industries, e.g., face recognition requirements for national security, detecting spam emails for e-commerce providers, deploying drones to carry out tactical jobs in areas where situations are hazardous for humans, etc. These applications need help from the rich collection of machine learning techniques that are described in the previous part. It is important to separate the applications and techniques, as there is a non-trivial overlap between those areas. There are some techniques specifically developed for certain types of applications, but there are some applications that can be benefitted from variety of different techniques, and the relationship between them is many to many. Only when we start looking at the field from these angles separately, lot of confusion starts to melt away and things start to make sense. The seamless solution of these problems and their delivery to millions of people is what creates the feel of artificial intelligence.

1.4.5 Implementations

The completion of studying parts 1 and 2, one can have sufficient confidence in analyzing a given problem in the areas of machine learning and artificial intelligence and come up with an outline of solution. However, this outline is not sufficient in actually implement the solution. In most cases, the theory is quite complex and solution of optimization problem is non-trivial. I am not going to delve into deep mathematical proofs and flowcharts that can be directly translated into computer programs for two reasons:

1. The content can quickly become too mathematical and will lose the interest of most of the target audience.

2. The complete theory behind all the techniques discussed here is so vast, that it is just not within the scope of a single book.

However, what I am going to target here is a conceptual understanding of the techniques, that is sufficient for a professional computer engineer, or a data analyst/scientist to grasp underlying key parameters of the problem.

Fortunately, the open source libraries in machine learning have matured to such level that, with such conceptual understanding of the techniques and applications, one can confidently implement reliable and robust artificially intelligent solutions without need to fully understand the underlying mathematics. This is the target of the part 3 of the book. This part will truly enable the reader to go ahead and build end to end solution of a complex practical problem and demonstrate an artificially intelligent system.

1.4.6 Conclusion

In this last part we summarize the learnings and discuss the next steps for the reader. After completion of this book the reader is well equipped to embark on solving real world problems that are still unsolved and make the world a better place.

Chapter 2
Essential Concepts in Artificial Intelligence and Machine Learning

2.1 Introduction

There are various concepts that are spread all over the place in terms of categorizing them with techniques or applications or implementations. However, they are at the heart of the entire machine learning theory and its applications and need special attention. We will cover these topics in this chapter. We will keep revising them in the book, but aggregating them here will make the reader more equipped to study the subsequent chapters.

2.2 Big Data and Not-So-Big Data

2.2.1 What Is Big Data

Big data processing has become a very interesting and exciting chapter in the field of AI and ML since the advent of cloud computing. Although there is no specific memory size above which the data is called big, the generally accepted definition is as follows: When the size of data is large enough such that it cannot be processed on a single machine, it is called as big data. Based on current (2018) generation of computers this equates to something roughly more than 10 GB. From there it can go to hundreds of petabytes (1 petabyte is 1000 terabytes and 1 terabyte is 1000 gigabytes) and more.

A. V. Joshi, *Machine Learning and Artificial Intelligence*,
https://doi.org/10.1007/978-3-030-26622-6_2

2.2.2 Why Should We Treat Big Data Differently?

Although the big data is separated from the not-so-big-data by simple size restrictions, the handling of the data changes drastically when we move from not-so-big data to big data. For processing not-so-big data, one need not worry about the location where each bit of data is stored. As all the data can be loaded into memory of single machine and can be accessed without additional overhead, all the typical numeric or string operations are executed in predictable time. However, once we enter the realm of big data, all such bets are off. Instead of a single machine, we are dealing with a cluster of machines. Each machine has its own storage and computational resources. For all the numeric and string operations one must carefully split the data into different machines, such that all the operands in individual operation are available on the same machine. If not, there is additional overhead of moving the data across the machines and it typically a very computation intensive operation. Iterative operations are especially very computation intensive for big data. Thus when dealing with big data, one must also carefully design the storage along with the operations.

2.3 Types of Learning

The machine learning algorithms are broadly classified into three types:

1. Supervised learning algorithms
2. Unsupervised learning algorithms
3. Reinforcement learning algorithms

Let's look at each type in detail.

2.3.1 Supervised Learning

For simplicity, let's consider the ML system as a black box that produces some output when given some input. If we already have a historical data that contains the set of outputs for a set of inputs, then the learning based on this data is called as supervised learning. A classic example of supervised learning is classification. Say we have measured 4 different properties (sepal length, sepal width, petal length, and petal width) of 3 different types of flowers (Setosa, Versicolor, and Virginica) [3]. We have measurements for say 25 different examples of each type of flower. This data would then serve as training data where we have the inputs (the 4 measured properties) and corresponding outputs (the type of flower) available for training the model. Then a suitable ML model can be trained in supervised manner. Once the model is trained, we can then classify any flower (between the three known types) based on the sepal and petal measurements.