Simplified Machine Learning

The essential building blocks for Machine Learning expertise

Dr. Pooja Sharma

www.bpbonline.com

First Edition 2024

ISBN: 978-93-55516-145

LIMITS OF LIABILITY AND DISCLAIMER OF WARRANTY

The information contained in this book is true to correct and the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but publisher cannot be held responsible for any loss or damage arising from any information in this book.

All trademarks referred to in the book are acknowledged as properties of their respective owners but BPB Publications cannot guarantee the accuracy of this information.

Dedicated to

My students
and
My family

About the Author

Dr. Pooja Sharma, Assistant Professor, in Computer Science and Engineering has teaching and research experience of more than 17 years. She is a gold medalist in post-graduation and her other academic achievements include a fellowship for a regular PhD from UGC, New Delhi after qualifying UGC NET and JRF, several merit certificates, gold and silver medals in matric, higher secondary, undergraduate and postgraduate levels. She was awarded PhD in 2013 on Content-Based Image Retrieval under the supervision of Dr. Chandan Singh from Punjabi University, Patiala. She has several research publications in peer-reviewed International journals of Springer and Elsevier with significant Thomson Reuters impact factors. She is the author of various research book chapters and has published a book on "Programming in Python". She is the reviewer of various International journals Elsevier, IET (IEEE Computer Society), and Scientific Research and Essays. She has participated in various conferences and workshops. Her areas of specialization include Data Analysis, Machine Learning, Content-Based Image Retrieval, Face Recognition, Pattern Recognition, and Digital Image Processing. She worked and was selected at various eminent Universities and Colleges including Central University. She had been the Head of Department at DAV University for 3 years. Currently, she holds a position as an Assistant Professor in the Department of Computer Science and Engineering, IKG Punjab Technical University, Main Campus, Kapurthala.

Acknowledgement

I am always indebted and grateful to the Almighty for making me capable of writing this book. I extend my gratitude to my parents for their encouragement and support in every sphere of life. I would like to acknowledge my husband Rajkumar for his persistent encouragement and support throughout writing this book and my daughter Angel.

I convey my heartiest thanks to the authorities of IKG Punjab Technical University, Kapurthala and my colleagues in the Department of Computer Science and Engineering for their support and cooperation.

I would like to acknowledge BPB Publications for their direction and expertise in bringing this book to completion. It was a long journey of writing and revising this book, with valuable assistance and guidance from reviewers, technical experts, and editors.

Finally, I would like to thank all the readers who have taken an interest in my book and for their support in making it a reality. Your encouragement has been invaluable.

Preface

Welcome to the world of Machine Learning! This book is designed to be your companion on a journey through the exciting and rapidly evolving field of Artificial Intelligence. Whether you are a student, researcher, or industry professional, this comprehensive guide aims to demystify complex concepts and equip you with practical skills to navigate the vast landscape of Machine Learning.

In recent years, Machine Learning has emerged as a transformative force, revolutionizing industries and shaping our everyday experiences. From personalized recommendations on streaming platforms to advanced medical diagnostics, the applications of Machine Learning are ubiquitous and ever-expanding. As you embark on this learning adventure, our goal is to provide you with a solid foundation in the fundamental principles, algorithms, and techniques that power these innovations.

This book is structured to cater to learners at all levels. We start by laying down the groundwork, explaining core concepts such as supervised and unsupervised learning, regression, classification, and clustering. We then dive into the intricacies of various Machine Learning algorithms, including decision trees, support vector machines, neural networks, and association rule mining. Each chapter is enriched with intuitive explanations, illustrative examples, and hands-on exercises to reinforce your understanding.

Beyond theory, this book emphasizes practical applications. We walk you through real-world case studies, providing insights into deploying Machine Learning models, interpreting results, and addressing ethical considerations. Whether you are interested in healthcare, finance, or natural language processing, you will find actionable insights and project-based learning exercises tailored to diverse domains.

As you embark on this journey into the world of Machine Learning, we invite you to embrace curiosity, embrace challenges, and embrace the possibilities that await. Let us embark on this adventure together and discover the transformative potential of Machine Learning. The Chaperization of the book is given as follows:

Chapter 1: Introduction to Machine Learning – This chapter covers Machine Learning and how it is related to Artificial Intelligence. Various types of Machine Learning along with their applications are also discussed in this chapter. Furthermore, the step-by-step guide is given to install Python Jupyter for implementing Machine Learning algorithms.

Chapter 2: Data Pre-processing – This chapter covers the pre-processing of data before applying any of the Machine Learning algorithms. This chapter covers various types of datasets, the need for data pre-processing, data cleaning, data transformation, data splitting, data normalization and scaling, data integration and aggregation, text processing and more.

Chapter 3: Supervised Learning: Regression - This chapter covers the detailed concept of supervised learning. Various types of regression techniques are discussed such as simple linear regression, multiple linear regression, ridge regression, lasso regression, polynomial regression, and applications of regression are discussed with appropriate programming codes and examples. Various model evaluation methods and errors are also covered in this chapter.

Chapter 4: Supervised Learning: Classification - This chapter covers the fundamental concepts related to the second major technique of supervised learning i.e., classification. This chapter includes logistic regression, K nearest neighbours, decision tree, random forest, naïve bayes classifier, and support vector machines.

Chapter 5: Unsupervised Learning: Clustering - This chapter covers unsupervised learning and the most important concept under it, i.e., clustering. The needs and applications of clustering are also discussed in detail. All types of clustering such as partition-based, hierarchical, and density-based clustering techniques are explored. Principal component analysis and anomaly detection are also covered in this chapter.

Chapter 6: Dimensionality Reduction and Feature Selection – This chapter covers feature engineering using dimensionality reduction, feature selection, and recursive feature elimination method and much more.

Chapter 7: Association Rule Mining - This chapter covers how association is important in Machine Learning. This chapter explains the apriori algorithm in detail with the key concept of association rule mining.

Chapter 8: Artificial Neural Network - This chapter covers the comprehensive concept of artificial neural networks (ANN) which leads to deep learning. All the prominent ANNs viz. perceptron, feedforward, backpropagation, convolutional, recurrent, long short-term memory, gated recurrent, and autoencoders are explored in detail with their programming codes.

Chapter 9: Reinforcement Learning - This chapter covers reinforcement learning with its components and applications. Various methods such as epsilon greedy, softmax, upper confidence bound, and Markov decision process are explored in this chapter.

Chapter 10: Project – This chapter covers real-world problems for a deep understanding of Machine Learning models. A comprehensive practical approach is followed to apply a Machine Learning model with graphical results and measurement of accuracy.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/7be7b5

The code bundle for the book is also hosted on GitHub at
https://github.com/bpbpublications/Simplified-Machine Learning.
In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1
Introduction to Machine Learning

Introduction

Machine Learning (**ML**) is a transformative field within **Artificial Intelligence** (**AI**) that empowers computers to learn and make predictions or decisions without explicit programming. At its core, it simulates the human learning process by using data and mathematical algorithms to identify patterns, make inferences, and improve performance over time. In essence, ML allows computers to generalize from data. It starts with a dataset containing examples and corresponding outcomes, and the ML model learns to recognize underlying patterns or relationships within this data. These patterns can range from recognizing handwritten characters, predicting stock prices and diagnosing diseases from medical scans, to recommending products based on user behavior.

Supervised learning, one of the core branches of ML, involves training a model on labelled data, where the correct outcomes are known. In contrast, unsupervised learning deals with unlabeled data, aiming to discover hidden structures or groupings. Reinforcement learning focuses on training agents to make decisions by interacting with their environment and receiving feedback. Based on this, ML has a broad array of applications across industries, from healthcare and finance to e-commerce and self-driving cars. Its success is driven by advances in computing power, the availability of massive datasets, and improvements in algorithms. Popular ML libraries and frameworks like TensorFlow and sci-kit-learn have democratized the field, enabling researchers and developers to build and deploy powerful models.

Therefore, we can say that ML is the science of teaching computers to learn from data, paving the way for intelligent systems that can automate tasks, make predictions, and continually improve their performance. As it continues to evolve, ML holds immense potential to revolutionize various aspects of our lives and industries.

Structure

The chapter includes the following topics:

- Need for Machine Learning
- Relation between Artificial Intelligence and Machine Learning
- Types of Machine Learning
- Applications of Machine Learning
- Lifecycle of Machine Learning
- Steps to install Anaconda and Python

Objectives

At the end of this chapter, you will be able to understand about basic concept of ML and why it is needed. You will go through all the types of ML and its major applications. Apart from that you will learn about ML lifecycle and steps to install Anaconda for implementing ML algorithms using Python libraries.

Need for Machine Learning

ML is a powerful and versatile field that offers numerous benefits and opportunities, making it a compelling choice for various applications and industries. Here are some key reasons why ML is widely adopted:

- **Data-driven insights**: ML excels at extracting valuable insights and patterns from vast amounts of data. It can uncover trends and relationships that may not be apparent through traditional statistical analysis.
- **Automation**: ML algorithms can automate repetitive and labor-intensive tasks, freeing up human resources for more creative and strategic work. This is particularly valuable in industries like manufacturing, finance, and customer service.
- **Personalization**: ML enables businesses to provide personalized experiences to customers. This includes tailored product recommendations, content suggestions, and targeted marketing campaigns, which can improve customer satisfaction and retention.

- **Efficiency**: ML can optimize processes and resource allocation, leading to cost savings and improved operational efficiency. For example, predictive maintenance in manufacturing can reduce downtime and maintenance costs.
- **Scalability**: ML algorithms can handle large-scale data analysis and decision-making, making them suitable for applications ranging from e-commerce to healthcare.
- **Improved decision-making**: ML models can make data-driven decisions in real-time, which can be invaluable in fields like finance for algorithmic trading or in healthcare for treatment recommendations.
- **Problem solving**: ML can tackle complex problems that may have no straightforward algorithmic solution. This includes tasks like image recognition, language translation, and game playing.
- **Adaptability**: ML models can adapt to changing conditions and new data, allowing systems to remain relevant and effective over time.
- **Innovation**: ML has led to breakthroughs in various domains, including autonomous vehicles, natural language processing, and medical diagnostics, driving innovation across industries.
- **Competitive advantage**: Organizations that harness the power of ML can gain a competitive edge by offering better products, services, and customer experiences.
- **Scientific discovery**: In fields like genomics and materials science, ML accelerates research by analyzing complex datasets and predicting discoveries.
- **Accessibility**: With the availability of open-source ML libraries and cloud-based ML platforms, businesses and researchers have easy access to powerful tools and resources.
- **Sustainability**: ML can be used to optimize resource usage, reduce waste, and support sustainability efforts in areas such as agriculture, energy management, and transportation.
- **Healthcare advancements**: In healthcare, ML assists in disease diagnosis, drug discovery, and personalized treatment plans, potentially saving lives and improving patient outcomes.
- **Cybersecurity**: ML helps organizations detect and respond to cybersecurity threats by identifying anomalies and patterns in network traffic and user behavior.

ML offers the potential to solve complex problems, improve efficiency, and drive innovation across a wide range of fields. Its ability to learn from data and make data-driven decisions makes it a valuable tool for businesses, researchers, and industries looking to harness the power of data and automation to achieve their goals.

Relation between Artificial Intelligence and Machine Learning

AI and ML are closely related fields, with ML being a subset of AI. Here is how they are connected:

- **AI**: It refers to the broader concept of creating intelligent machines that can mimic human-like cognitive functions, such as reasoning, problem-solving, learning, perception, and language understanding. It encompasses a wide range of techniques, including rule-based systems, expert systems, knowledge representation, and more.
- **ML**: It is a subset of AI that focuses on the development of algorithms and statistical models that enable computers to learn from and make predictions or decisions based on data. It is a specific approach within AI that deals with the learning aspect.
- **Learning from data**: ML is a fundamental component of many AI systems. It provides the ability for AI systems to learn patterns, behaviors, and insights from large datasets.

AI systems may use ML techniques to improve their performance or adapt to changing conditions.

Automation

AI and ML often go hand in hand in automating tasks and decision-making. AI systems can use ML models to make informed decisions based on data.

For example, in autonomous vehicles, AI algorithms use ML models to process sensor data and make real-time driving decisions.

Adaptability

ML enables AI systems to adapt and improve their performance over time. AI systems can learn from new data and adjust their behavior accordingly. This adaptability is crucial for AI systems to handle complex and dynamic environments.

Natural language processing

Natural Language Processing (**NLP**) is a subset of AI that deals with human language understanding and generation. ML plays a significant role in NLP, as it is used to build models for tasks like language translation, sentiment analysis, and chatbots.

Computer vision

Computer vision is another AI subfield that focuses on enabling computers to interpret and understand visual information from the world. ML, particularly deep learning, has revolutionized computer vision, allowing AI systems to recognize objects, faces, and scenes in images and videos.

ML is a core component of many AI systems, providing them with the ability to learn, adapt, and make data-driven decisions. While AI encompasses a broader set of goals and techniques, ML is a crucial tool within the AI toolkit, enabling AI systems to perform tasks that involve learning from data and improving their performance over time.

Types of Machine Learning

ML can be categorized into several types, each with its approach and characteristics. The primary types of ML are as follows.

Supervised learning

In supervised learning, the algorithm is trained on a labelled dataset, where each input example is paired with its corresponding output or target. The goal is to learn a mapping from inputs to outputs, allowing the model to make predictions or classifications on new, unseen data. Common algorithms include linear regression, logistic regression, decision trees, support vector machines, and neural networks.

Unsupervised learning

Unsupervised learning deals with unlabeled data, where the algorithm seeks to discover patterns, structures, or relationships within the data without explicit guidance. It is often used for tasks like clustering (grouping similar data points) and dimensionality reduction (reducing the number of features while preserving important information). Common algorithms include k-means clustering, hierarchical clustering, **Principal Component Analysis** (**PCA**), and autoencoders.

Semi-supervised learning

Semi-supervised learning is a combination of supervised and unsupervised learning. It uses both labelled and unlabeled data to improve model performance. This is especially useful when obtaining labelled data is expensive or time-consuming. Techniques may include using a small amount of labelled data to guide the model's learning on the larger unlabeled dataset.