Fundamental Mathematical Concepts for Machine Learning in Science

Umberto Michelucci

Fundamental Mathematical Concepts for Machine Learning in Science

Springer

Umberto Michelucci
TOELT.AI
Dübendorf, Switzerland

ISBN 978-3-031-56430-7 ISBN 978-3-031-56431-4 (eBook)
https://doi.org/10.1007/978-3-031-56431-4

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

To Caterina and Francesca. To Francesca, my love, without whom nothing would be possible. To Caterina, my life, the person I am most proud of in my life.

Preface

In this book, I discuss all the mathematical and methodological aspects that are important when using machine learning techniques in natural sciences (but not only). This book was written with a clear purpose: to explain the complexities of **how to use** machine learning to students and professionals in the natural sciences who may not have a background in computer science. It is a bridge connecting two seemingly disparate worlds and offers a comprehensive guide to understanding and applying the important techniques when applying machine learning in the context of science. I will not discuss algorithms or neural networks, but I will explain all the methods that you need to know to use them properly (e.g. model validation, sampling theory, etc.). The book is structured to gradually build your understanding, starting from fundamental mathematical concepts and progressing to advanced machine learning methods. Each chapter is designed to be self-contained, allowing the reader to focus on specific topics of interest. The chapters on calculus, linear algebra, and statistics are particularly crucial as they lay the foundation for a comprehensive understanding of machine learning algorithms and approaches. Given the breadth and depth of many topics, adequately covering each one would fill many books. My aim has been to cover and explain the core concepts necessary for your journey. I highly recommend further exploring these topics through the additional readings and references that I provide throughout the text. As author, I have striven to present the material in an accessible, yet rigorous manner. The book relies heavily on mathematics as the language of machine learning, ensuring that concepts are conveyed with precision and clarity. Although the book does not delve into programming details, it points out its relevance in machine learning, especially considering Python. What sets this book apart is its focus on methods about **how** to use machine learning, and not on the algorithms themselves filling a gap in the existing machine learning literature. Whether you are a physicist, chemist, biologist, doctor, or involved in any scientific discipline, this book is your guide to harnessing the power of machine learning in your field (if you are a computer scientist, this book is for you too!). This book is a portal to the exciting world of machine learning, written to enrich your scientific research projects, and I hope spark a lifelong interest in this field. If you are an instructor, on the book webpage on Springerlink you will find an instructor guide that will help you to use this book for a course.

Dübendorf, Switzerland — *Umberto Michelucci*
March, 2024

Acknowledgements

This book would not have been possible without the help of many people. At the forefront is my wife and esteemed colleague, Prof. Dr. Francesca Venturini. Her role in this journey has been immeasurable. Francesca is a constant source of support (in life as in projects), always offering insightful critiques and invaluable suggestions that have significantly shaped the content and direction of this book. I extend my heartfelt gratitude to the numerous students who have patiently listened to my lectures. Their participation and feedback have been instrumental in identifying the most challenging concepts in the field, highlighting areas that are often under-represented in standard university curricula. Their curious minds and challenging questions have not only aided my understanding, but also enriched the content of this book, making it more accessible and relevant to learners at all levels.

A special acknowledgement goes to my editor, Paul Drougas. Paul's expertise, patience, and flexibility have been pivotal in bringing this book to fruition. His guidance through the more challenging phases of this project, combined with his unwavering support, has been invaluable. Paul's proficiency and dedication as an editor have greatly contributed to the quality and coherence of this manuscript.

I am also indebted to many others who have offered their insight, participated in thought-provoking discussions, and provided feedback in various forms. Among them, I am particularly grateful to Dr. Michela Sperti, Prof. Dr. Marco Deriu, and Prof. Dr. Els Ortibus. Each of them has contributed to the richness of the content presented in this book.

Although I have attempted to acknowledge as many contributors as possible, I apologise to those I may have inadvertently omitted. Your contributions have not gone unnoticed and I am deeply grateful for your input. Finally, I assume full responsibility for any errors that may remain in the text. The insights and support provided by all mentioned and those unmentioned have been invaluable, but the final product and any shortcomings it may possess are solely my responsibility. This book is not just a reflection of my efforts, but the result of the collective wisdom and support of many brilliant minds. To each and every one of you, I extend my deepest gratitude.

Dübendorf, February 2024

Contents

Acronyms

Acronyms

AE	Autoencoder
AUC	Area Under the Curve
BFE	Backward Feature Elimination
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
DL	Deep Learning
FFNN	Feed Forward Neural Network
FFS	Forward Feature Selection
GAN	Generative Adversarial Networks
GD	Gradient Descent
GBM	Gradient Boosting Machines
kNN	k-Nearest Neighbour
LLM	Large Language Model
MAE	Mean Absolute Error
ML	Machine Learning
MNIST	Modified National Institute of Standards and Technology
MSE	Mean Square Error
NLP	Natural Language Processing
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RFE	Recursive Feature elimintation
RL	Reinforcement Learning
ROC	Receiving Operating Characteristic
t-SNE	t-Distributed Stochastic Neighbour Embeddings
SMOTE	Synthetic Minority Oversampling Techniques
SVM	Support Vector Machine
VAE	Variational Autoencoder