
Preface

This book is an outgrowth of a belief that the mathematics and, in general, the scientific community might be well served by an introduction to deep learning and neural networks in the language of mathematics. To borrow from Churchill, Shaw, and Wilde, mathematics and computer science are two disciplines separated by common notation. We believe that this book might help students, researchers, and practitioners more easily see and explore connections to this increasingly important collection of computational tools and ideas. Over several years of research and teaching in neural networks, the authors have come to the conclusion that

- There are many interesting and open mathematical research questions in the field of deep learning.
- Mathematical maturity gives students and researchers an advantage in thinking about machine learning.

Mathematical thinking innately has unique strengths in generalizing and abstracting ideas and also providing rigorous bounds on complex phenomena. We believe that a greater mathematical presence in the field of deep learning and neural networks can in turn contribute to the larger scientific community.

This book is aimed at advanced undergraduate students and graduate students as well as researchers and practitioners who want to understand the mathematics behind the different deep learning algorithms. The book is composed of two parts. Part 1 contains a mathematical introduction, while Part 2 discusses more advanced mathematical and computational topics, hinting at further research directions. This represents something of a “separation of

scales” in our effort: the basics of deep learning are “microscopic”, while large-scale structural analysis is more “macroscopic”. Our hope is that the combination of both points of view will offer a better comprehension of the topic.

The first part of the book (Part 1) assumes knowledge of basic linear algebra, multivariate calculus, and some statistics and calculus-based probability. This part of the book should be accessible to an advanced mathematics, statistics, computer science, data science, or engineering undergraduate student. Part 1 starts with some classical topics in statistical learning theory (such as linear regression, logistic regression, and kernels), then gradually progresses to deep learning related topics (such as feed forward neural networks, backpropagation, stochastic gradient descent, dropout, batch normalization), and concludes with a broad spectrum of deep learning architectures and models (such as recurrent neural networks, transformers, convolution neural networks, variational inference, and generative models). We have also included sections and chapters on classical statistical topics as regularization, training, validation and testing, and feature importance. The purpose of the earlier chapters in Part 1 is to introduce the reader to the subject of deep learning in a gradual way through classical topics in statistics and machine learning. These early chapters allow us to illustrate known issues that come up in deep learning architectures through easier-to-present concrete settings that often allow explicit computations, the latter being rarely the case for general deep learning architectures.

The second part of the book (Part 2) contains material that is more advanced than Part 1, either mathematically, conceptually, or computationally. This part of the book should be accessible to advanced undergraduate students and graduate students aiming to go deeper in certain topics of deep learning. Certain aspects of the second part of the book (e.g., uniform approximation theorems, convergence theory for gradient and stochastic gradient descent, linear regime and the neural tangent kernel, feature learning regime and mean field field scaling, neural differential equations) would be easier to read given a basic understanding of real analysis and stochastic process. A self-contained appendix with more advanced required background material has been included to aid the reader. Other aspects of the second part of the book, e.g., distributed training and automatic differentiation, require less mathematical background but are more advanced either conceptually or computationally.

Part 1, potentially together with selected topics from Part 2, could serve as standalone material for an advanced undergraduate course or for a first year graduate introduction to the mathematics of deep learning (we have done so in related course offerings in our respective universities).

The topic of deep learning is already huge and is constantly growing. While we have attempted to provide a fairly broad overview, we do not claim to have covered all possible angles. Our aim has been to cover topics that we viewed

as important, foundational, and reasonably well-developed at the time of writing. We have tried to establish a unified and consistent mathematical language, connecting those topics in a comprehensive way and keeping in mind that deep learning is both mathematically interesting and a tool in applied data analysis.

Our efforts have concentrated around the idea of presenting essential ideas as clearly as possible. As such, we may not have always presented the sharpest possible versions of the results, but we have pointed to research articles and other monographs where the interested reader can find more refined results. No attempt has been made to provide comprehensive historical attribution of ideas. We do however give appropriate references which will hopefully provide entry points into the literature.

The book is focused on developing a mathematical language for deep learning and unifying the presentation of concepts and ideas but also maintaining rigorous mathematical results. The goal is not only to understand the mathematical principles behind deep learning algorithms, but also to offer tools to quantify uncertainty in deep learning. Two of the questions that this book is trying to answer are

Why do things work the way they work?

and

How can we guarantee significance and robustness of our conclusions?

Website

Beyond reading the mathematical literature, the readers of this book will hopefully have the opportunity to experiment with the algorithms presented. Deep learning is a tool in data analysis. For the reader's convenience we have included Python code which will hopefully give the reader some appreciation for how deep learning might actually be used in practice. The datasets and Python codes referenced in various chapters of the book can be found and downloaded at the dedicated website for the book

<https://mathdl.github.io/>.

In addition, corrections and errata to the book will be updated there.

In many of the chapters of the book, exercises have been included to aid the reader in better comprehension of the material. Upon request, a solutions manual is available to the instructor of a class using this book.

We hope that this book will help open a door through which the mathematics and research community can pass in order to contribute even more to the ever-growing field of research and applications of deep learning. Whatever goals have motivated us to write these chapters, we admit to partial success only.

Acknowledgments

The authors of the book would like to thank family and friends who contributed in their own unique ways to the completion of the book.

- Konstantinos Spiliopoulos would like to thank first and foremost Anastasia for her constant support and patience, and his children Stavros and Evangelia for being a continuous driving force and source of inspiration!
- Richard Sowers would like to thank his students, who helped him learn deep learning. In particular, Rachneet Kaur, through a series of papers [**HBK⁺20**, **KCM⁺20**, **KKHS20**, **KLM⁺23**, **KMZ⁺19**, **KMSH22**, **KST⁺20**, **KSZ⁺19**, **SKZ⁺19**], helped him bridge the gap from mathematical thinking to applications. Richard Sowers would also like to thank his wife Svetlana, and daughters Rebeka and Veronika. G. B. Shaw once opined that “A happy family is but an earlier heaven.”

Finally, we would like to thank our friends, colleagues and students who looked at earlier versions of the manuscript and made many useful suggestions. In particular, we would like to thank in alphabetical order Nikos Georgoudios, Paul Glasserman, Samuel Isaacson, Markos Katsoulakis, Fotios Kokkotos, Eric Kolaczyk, Scott Robertson, Giorgos Zervas, Benjamin Zhang, and Zhuo Zhao. All of the remaining mistakes and imperfections are our own responsibility.

While working on this book, the authors were supported by the National Science Foundation (NSF) in the USA and the Engineering and Physical Sciences Research Council (EPSRC) in the UK (grant NSF-DMS 2311500).

KONSTANTINOS SPILOPOULOS
Boston, MA, USA

RICHARD SOWERS
Urbana-Champaign, IL, USA

JUSTIN SIRIGNANO
Oxford, UK

May 2025