The Mathematical Engineering of Deep Learning

Book Draft

Benoit Liquet, Sarat Moka and Yoni Nazarathy

January 3, 2022

Preface - DRAFT

In the last few years deep learning has seen explosive growth and even dubbed as the "new electricity". The field has shown incredible success in automated applications and predictive tasks. Deep learning models are mathematical in nature, and hence to understand deep learning, one needs to understand the mathematical description of the models. This book aims to provide such an understanding via a concise, accessible, and self contained presentation.

Many deep learning resources focus on programming while making an effort to hide the mathematics. Other resources focus on theoretical results without an attempt to disambiguate the terminology of the field. A third breed of resources puts heavy emphasis on historical progression. Each of these viewpoints is important for a specific purpose, however we conjecture that for a mathematical audience, these are all suboptimal ways to learn about deep learning. Hence we created this book.

Our focus is on the basic mathematical constructs that make up the field. We call this **the mathematical engineering of deep learning**. Using the language of equations and algorithms, deep learning objects interface together to make very powerful models. A reader armed with basic familiarity of mathematical notation and knowledge of basic calculus, probability, and linear algebra can go a long way in understanding deep learning quickly. For this, we use simple mathematical notation to outline the mechanisms used for training, execution, and application of deep neural networks.

Deep learning is certainly not soley about mathematics, as it also requires good software, hardware, and data. However, we aim to present the mathematical technology of deep learning without focusing on the implementation aspects that one would consider if trying to use deep learning frameworks in practice. We also aim to focus on the current state of the art as opposed to the historical progression of the field. Finally, we aim to minimize the focus on the human brain and the loose analogies that one can make between

deep artificial neural networks and actual biological neurons. All of these aspects that we downplay are important, but we believe that if in an initial exposure to the field one spends too much time on implementation, history, or bio-neurlaogical analogies, then the simplicity of deep learning is missed.

The book is primarily intended for readers from engineering, signal processing, statistics, physics, econometrics, operations research, quantitative management, pure mathematics, bioinformatics, applied machine learning, or even applied deep learning. A reader with background in one of these domains will be able to get a concentrated and concise description of deep learning. In cases where a mathematical refresher is needed, appendices provide a condensed review, such as for key aspects of multi-variable calculus.

The book can be read sequentially, or alternatively readers may wish to jump between chapters for quick lookup. It is assumed that readers have had exposure to mathematical notation at the level equivalent to at least 3 or 4 university courses. Hence set notation, matrices, basic probability, and calculus are used without apology. However, no explicit knowledge of machine learning, statistics, optimization, or advanced probability is needed or assumed. Our hope is that we strike the right balance so that a mathematically equipped non-expert can easily read the book in a self contained manner.

While the focus of the book is "mathematical engineering", we fully acknowledge the importance of applications and the ability to use software and hardware effectively. For this you may also use the companion website, https://deeplearningmath.org/, where additional examples, links, and software usage details are provided.

Outline of the Contents

The book has 10 chapters, an epilogue, and 4 appendices. Chapters 1-4 introduce the field, outline key concepts from machine learning, overview optimization concepts needed for deep learning, and focus on fundamental models and concepts. Chapters 5-8 deal with the core models and architectures of deep learning, including fully connected networks, convolutional networks, recurrent networks, and outline aspects of model tuning and application. Chapters 9-10 deal with specific domains, namely generative adversarial networks and deep reinforcement learning. The epilogue frames all of the concepts covered in a contemporary perspective. Appendices A-C

provide mathematical support, and appendix D outlines current software and hardware trends. Here is a detailed outlined of the contents.

Chapter 1 – Introduction: In this chapter we overview deep learning, demonstrate key applications, survey the associated ecosystems of high performance computing, discuss big and high-dimensional data, and set the tone for the rest of the book. The chapter discusses key terminology including data science, machine learning and statistical learning, and puts these terms in the context of the book. Key popular datasets such as ImageNet and MNIST digits are also overviewed together with a description of the deep learning culture that emerged.

Chapter 2 – Principles of Machine Learning: Deep learning can be viewed as a sub-discipline of machine learning and hence this chapter provides an overview of key machine learning concepts and paradigms. The reader is introduced to supervised learning, unsupervised learning, and the general concept of iterative based optimization for learning. The concepts of training sets, test sets, and the like, together with principles of cross validation and models selection are introduced. A key object explored in the chapter is the linear model which can be trained also via iterative optimization. This allows to see the basic gradient descent algorithm in action, which is later refined and used heavily in the continuation of the book.

Chapter 3 – Simple Neural Networks: In this chapter we focus on logistic regression for binary classification and the related softmax regression model for multi-class problems. This is where principles of deep learning such as cross entropy loss, decision boundaries, and simple cases of back-propagation are introduced. The chapter also presents a simple non-linear auto-encoder architecture. Aspects of model tuning are also discussed including feature engineering and hyper-parameter choice.

Chapter 4 – Optimization Algorithms: The training of deep learning models involves optimization over the learned parameters. Hence a solid understanding of optimization algorithms is required, as well as an understanding of specialized optimization techniques that work well for deep learning models such as the ADAM algorithm. In this chapter we focus on such techniques as well as on more advanced second order methods that are slowly finding their way into practice. We also study the details of various forms of automatic differentiation and finish with a comparison in the context of logistic regression where both first order and second order methods may be employed.

Chapter 5 – Feed-Forward Deep Networks: This chapter is the heart of the book where the general feed-forward deep neural network is defined and introduced. After exploring the expressive power of deep neural networks we dive into the details of training by understanding the back-propagation algorithm for gradient evaluation and exploring other practical aspects such as weight initialisation, dropout, and batch normalization.

Chapter 6 – Convolutional Neural Networks: Much of the success of deep learning is due to the strength of convolutional neural networks when applied to images and similar data formats. In this chapter we explore the convolution concept and then see it in the context of deep learning models. The concepts of channels, and filter design are introduced, followed by an exploration of common state of the art architectures that have made significant impact and are still in use today. We also explore a few key tasks associated with images such as object localization.

Chapter 7 – Sequence Models: Sequence models are critical for data such as text with applications in natural language processing. In this chapter we get a taste for the key deep learning ideas of the field. We explore recurrent neural networks and their generalizations. These include long short term memory models, gated recurrent units, auto-encoders for end to end language translation, and the attention model with transformers.

Chapter 8 – Tricks of the Trade: After exposure to feed-forward networks, convolutional networks, and various forms of recurrent networks, we now explore common methods for tuning and integrating such models in applications. Key questions involve hyper-parameter choices and techniques for optimizing them. Other questions deal with adaptation of models from one dataset to another via transfer learning, as well as ways to augment datasets. We also discuss applications of transformers for images and various aspects of implementation including a description of what one might expect from a deep learning software framework.

Chapter 9 – Generative Adversarial Networks: In this chapter we survey and explore Generative Adversarial Networks (GANs), which are models that are able to synthesize fake data which appears realistic. The basic GAN construction is based on a game-theoretic setting where a generator model and discriminator model are trained jointly to arrive at a trained system. We discuss several GAN architectures as well as interesting mathematical aspects that arise when adapting loss functions.

Chapter 10 – Deep Reinforcement Learning: In this final chapter we explore principles of deep reinforcement learning, an adaptive control method for dynamic systems. This are is often introduced in the context of agents when considering artificial intelligence systems, however we take a more classical approach and present it in the context of control theory and Markov decision processes. We first lay the foundations for MDPs and Q-learning, and then explore various advances associated with approximating Q-functions via deep neural networks.

With Thanks

We began this project while undertaking instruction at the 2021 AMSI (Australian Mathematical Sciences Institute) summer school. In that course we taught 60 students from over all over Australia for 28 lecture hours. See: https://deeplearningmath.org/amsi-summer-school-course-2021.html. We thank the students for embarking on the journey with us and further appreciate student comments that were useful for creating the book. We also mention that without support from our families and loved ones, this book wouldn't be possible.

We hope you enjoy the book.

Benoit Liquet, Sarat Moka, and Yoni Nazarathy.
December 2021.

Contents