Mathematical Engineering of Deep Learning

Book Draft

Benoit Liquet, Sarat Moka and Yoni Nazarathy

February 28, 2024

Contents

Epilogue - DRAFT

Our story was about the **mathematical engineering of deep learning**. Our goal was to describe deep learning ideas in simple mathematical terms. Our goal was not to study implementation of deep learning; it was not to discuss the history and evolution of deep learning; and it was not to dive into subtle mathematical properties of deep learning. We simply wanted to present a basic **mathematical** description, empowering the reader with an understanding of key concepts and terminology. Mathematics is a language of choice.

We focused on the most popular and successful **deep learning** architectures and ideas that emerged over recent years. Somewhat anti-climatically we claim that the popularity and success of these ideas is due to their practical applicability, and not so much due to mathematical elegance. There are many other variants that we did not present here which are interesting and elegant yet have not been as popular from a practical perspective. With this we note that the aspect of **engineering** focusing on the empirical evaluation of architectures was not discussed and studied in the book at all.

Take as an example the *transformer architecture* studied in Section 7.5. This architecture has been pivotal in *large language models*. Indeed, in the same years that we worked on writing this book, 2021–2023, large language models, almost exclusively powered by the transformer architecture, have risen in popularity. Yet it is fair to say that the transformer architecture is quite arbitrary. If a couple of years prior to the development of this architecture, published in 2017 with [410], we the authors would have been presented with a transformer, without empirical trials and experimentation results, we would have no proof that transformers work so well.

It is also important to note that the pace and unpredictability of deep learning developments moves fast. By now, large language models have effectively beaten the Turing test, [38], a goal which seemed yet unattainable in the days when we conceived this book in late 2020. So our humble claim is that while **mathematical engineering** is important, in its own right, without computers, GPUs, software, data, and experimentation, it is void of substance. Nevertheless, we do believe that our presentation approach is succinct and unique, and given that the ideas that we present were previously shown to be winning ideas, the knowledge that you gained by reading this book will be beneficial.

Finally we close by mentioning that while this is a mathematical book, one cannot ignore the vast area of ethical issues associated with deep learning and artificial intelligence. Now, as we are in the third decade of the twenty first century, artificial intelligence is at the center of discussions associated with politics, freedom, social justice, violence, equity, and many other domains. Since this book is not about applications, we as authors had the luxury of ignoring the many ethical issues associated with deep learning in our exposition. Nevertheless, any practitioner using deep learning should at onset make sure to consider what defines responsible use and what not. We certainly want the technology to be used for purposes that do good rather than bad.