Mathematics in Industry 37

Jong Chul Ye

Geometry of Deep Learning

A Signal Processing Perspective

Springer

Mathematics in Industry

Volume 37

Mathematics in Industry focuses on the research and educational aspects of mathematics used in industry and other business enterprises. Books for *Mathematics in Industry* are in the following categories: research monographs, problem-oriented multi-author collections, textbooks with a problem-oriented approach, conference proceedings. Relevance to the actual practical use of mathematics in industry is the distinguishing feature of the books in the *Mathematics in Industry* series.

More information about this series at https://link.springer.com/bookseries/4650

Jong Chul Ye

Geometry of Deep Learning

A Signal Processing Perspective

Springer

Jong Chul Ye
Korea Advanced Institute of Science
and Technology (KAIST),
Daejeon, Republic of Korea

ISSN 1612-3956 ISSN 2198-3283 (electronic)
Mathematics in Industry
ISBN 978-981-16-6045-0 ISBN 978-981-16-6046-7 (eBook)
https://doi.org/10.1007/978-981-16-6046-7

Mathematics Subject Classification: 68T01, 68T07

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

To Andy, Ella, and Joo

Preface

It was a very different, unprecedented, and weird start of the semester, and I did not know what to do. This semester, I was supposed to offer a new senior-level undergraduate class on *Advanced Intelligence* to jointly teach students at the Department of Bio/Brain Engineering and the Department of Mathematical Sciences. I had initially planned a standard method for teaching machine learning, the contents of which are practical, experience-based lectures with a lot of interaction with the students through many mini-projects and term projects. Unfortunately, the global pandemic of COVID-19 has completely changed the world and such interactive classes are no longer an option most of the time.

So, I thought about the best way to give online lectures to my students. I wanted my class to be different from other popular online machine learning courses but still provide up-to-date information about modern deep learning. However, not many options were available. Most existing textbooks are already outdated or very implementation oriented without touching the basics. One option would be to prepare presentation slides by adding all the up-to-date knowledge that I wanted to teach. However, for undergraduate-level courses, the presentation files are usually not enough for students to follow the class, and we need a textbook that students can read independently to understand the class. For this reason, I decided to write a reading material first and then create presentation files based on it, so that the students can learn independently before and after the online lectures. This was the start of my semester-long book project on *Geometry of Deep Learning*.

In fact, it has been my firm belief that a deep neural network is not a magic black box, but rather a source of endless inspiration for new mathematical discoveries. Also, I believed in the famous quote by Isaac Newton, "Standing on the shoulders of giants," and looking for a mathematical interpretation of deep learning. For me as a medical imaging researcher, this topic was critical not only from a theoretical point of view but also for clinical decision-making, because we do not want to create false features that can be recognized as diseases.

In 2017, on a street in Lisbon, I had *Eureka!* moment in understanding hidden framelet structure in encoder-decoder neural networks. The resulting interpretation of the deep convolutional framelets, published in the *SIAM Journal of Imaging*

Science, has had a significant impact on the applied math community and has been one of the most downloaded papers since its publication. However, the role of the rectified linear unit (ReLU) was not clear in this work, and one of the reviewers in a medical imaging journal consistently asked me to explain the role of the ReLU in deep neural networks. At first, this looked like a question that went beyond the scope of the medical application paper, but I am grateful to the reviewer, as during the agony of preparing the answers to the question, I realized that the ReLU determines the input space partitioning, which is automatically adapted to the input space manifold. In fact, this finding led to a 2019 ICML paper, in which we revealed the combinatorial representation of framelets, which clearly shows the crucial connection with the classic compressed sensing (CS) approaches.

Looking back, I was pretty brave to start this book project, as these are just two pieces of my geometric understanding of deep learning. However, as I was preparing the reading material for each subject of deep learning, I found that there are indeed many exciting geometric insights that have not been fully discussed.

For example, when I wrote the chapter on backpropagation, I recognized the importance of the denominator layout convention in the matrix calculus, which led to the beautiful geometry of the backpropagation. Before writing this book, the normalization and attention mechanisms looked very heuristic to me, with no evidence of a systematic understanding that is even more confusing due to their similarities. For example, AdaIN, Transformer, and BERT were like dark recipes that researchers have developed with their own secret sauces. However, an in-depth study for the preparation of the reading material has revealed a very nice mathematical structure behind their intuition, which shows a close connection between them and their relationship to optimal transport theory.

Writing a chapter on the geometry of deep neural networks was another joy that broadened my insight. During my lecture, one of my students pointed out that some partitions can lead to a low-rank mapping. In retrospect, this was already in the equation, but it was not until my students challenged me that I recognized the beautiful geometry of the partition, which fits perfectly with fascinating empirical observations of the deep neural network.

The last chapter, on generative models and unsupervised learning, is something of which I am very proud. In contrast to the conventional explanation of the generative adversarial network (GAN), variational auto-encoder (VAE), and normalizing flows with probabilistic tools, my main focus was to derive them with geometric tools. In fact, this effort was quite rewarding, and this chapter clearly unified various forms of generative model as statistical distance minimization and optimal transport problems.

In fact, the focus of this book is to give students a geometric insight that can help them understand deep learning in a unified framework, and I believe that this is one of the first deep learning books written from such a perspective. As this book is based on the materials that I have prepared for my senior-level undergraduate class, I believe that this book can be used for one-semester-long senior-level undergraduate and graduate-level classes. In addition, my class was a code-shared course for both bioengineering and math students, so that much of the content of the work is interdisciplinary, which tries to appeal to students in both disciplines.

I am very grateful to my TAs and students of the 2020 spring class of BiS400C and MAS480. I would especially like to thank my great team of TAs: Sangjoon Park, Yujin Oh, Chanyong Jung, Byeongsu Sim, Hyungjin Chung, and Gyutaek Oh. Sangjoon, in particular, has done a tremendous job as Head TA and provided organized feedback on the typographical errors and mistakes of this book. I would also like to thank my wonderful team at the Bio Imaging, Signal Processing and Learning laboratory (BISPL) at KAIST, who have produced ground-breaking research works that have inspired me.

Many thanks to my awesome son and future scientist, Andy Sangwoo, and my sweet daughter and future writer, Ella Jiwoo, for their love and support. You are my endless source of energy and inspiration, and I am so proud of you. Last, but not the least, I would like to thank my beloved wife, Seungjoo (Joo), for her endless love and constant support ever since we met. I owe you everything and you made me a good man. With my warmest thanks,

Daejeon, Korea
February, 2021

Jong Chul Ye

Contents