Charu C. Aggarwal

Linear Algebra and Optimization for Machine Learning

A Textbook

A frequent challenge faced by beginners in machine learning is the extensive background requirement in linear algebra and optimization. This makes the learning curve very steep. This book, therefore, reverses the focus by teaching linear algebra and optimization as the primary topics of interest, and solutions to machine learning problems as applications of these methods. Therefore, the book also provides significant exposure to machine learning. The chapters of this book belong to two categories:

1. Linear algebra and its applications: These chapters focus on the basics of linear algebra together with their common applications to singular value decomposition, similarity matrices (kernel methods), and graph analysis. Numerous machine learning applications have been used as examples, such as spectral clustering, kernel-based classification, and outlier detection.

2. Optimization and its applications: Basic methods in optimization such as gradient descent, Newton's method, and coordinate descent are discussed. Constrained optimization methods are introduced as well. Machine learning applications such as linear regression, SVMs, logistic regression, matrix factorization, recommender systems, and K-means clustering are discussed in detail. A general view of optimization in computational graphs is discussed together with its applications to backpropagation in neural networks.

Exercises are included both within the text of the chapters and at the end of the chapters. The book is written for a diverse audience, including graduate students, researchers, and practitioners.

About the Author

Charu C. Aggarwal is a Distinguished Research Staff Member (DRSM) at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his undergraduate degree in Computer Science from the Indian Institute of Technology at Kanpur in 1993 and his Ph.D. in Operations Research from the Massachusetts Institute of Technology in 1996. He has published more than 400 papers in refereed conferences and journals, and has applied for or been granted more than 80 patents. He is author or editor of 19 books, including textbooks on data mining, neural networks, machine learning (for text), recommender systems, and outlier analysis. Because of the commercial value of his patents, he has thrice been designated a Master Inventor at IBM. He has received several internal and external awards, including the EDBT Test-of-Time Award (2014), the IEEE ICDM Research Contributions Award (2015), and the ACM SIGKDD Innovation Award (2019). He has served as editor-in-chief of the ACM SIGKDD Explorations, and is currently serving as an editor-in-chief of the ACM Transactions on Knowledge Discovery from Data. He is also an editor-in-chief of ACM Books. He is a fellow of the SIAM, ACM, and the IEEE, for "contributions to knowledge discovery and data mining algorithms."

ISBN 978-3-030-40343-0

▶ springer.com

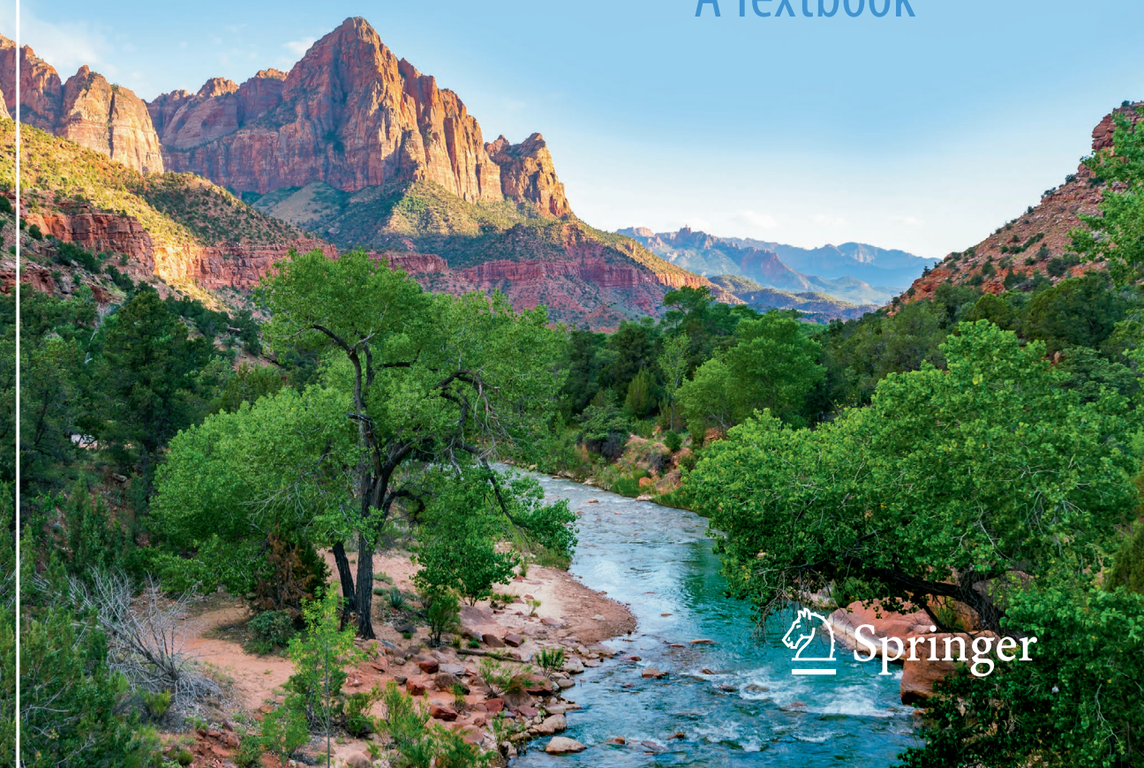
Aggarwal

Linear Algebra and Optimization for Machine Learning

Charu C. Aggarwal

Linear Algebra and Optimization for Machine Learning

A Textbook



Springer

Linear Algebra and Optimization for Machine Learning

Charu C. Aggarwal

Linear Algebra and Optimization for Machine Learning

A Textbook

Springer

Charu C. Aggarwal
Distinguished Research Staff Member
IBM T.J. Watson Research Center
Yorktown Heights, NY, USA

ISBN 978-3-030-40343-0 ISBN 978-3-030-40344-7 (eBook)
https://doi.org/10.1007/978-3-030-40344-7

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

To my wife Lata, my daughter Sayani,
and all my mathematics teachers

Contents

Preface

"Mathematics is the language with which God wrote the universe."– Galileo

A frequent challenge faced by beginners in machine learning is the extensive background required in linear algebra and optimization. One problem is that the existing linear algebra and optimization courses are not specific to machine learning; therefore, one would typically have to complete more course material than is necessary to pick up machine learning. Furthermore, certain types of ideas and tricks from optimization and linear algebra recur more frequently in machine learning than other application-centric settings. Therefore, there is significant value in developing a view of linear algebra and optimization that is better suited to the specific perspective of machine learning.

It is common for machine learning practitioners to pick up missing bits and pieces of linear algebra and optimization via "osmosis" while studying the solutions to machine learning applications. However, this type of unsystematic approach is unsatisfying, because the primary focus on machine learning gets in the way of learning linear algebra and optimization in a generalizable way across new situations and applications. Therefore, we have inverted the focus in this book, with linear algebra and optimization as the primary topics of interest and solutions to machine learning problems as the *applications* of this machinery. *In other words, the book goes out of its way to teach linear algebra and optimization with machine learning examples.* By using this approach, the book focuses on those aspects of linear algebra and optimization that are more relevant to machine learning and also teaches the reader how to apply them in the machine learning context. As a side benefit, the reader will pick up knowledge of several fundamental problems in machine learning. At the end of the process, the reader will become familiar with many of the basic linear-algebra- and optimization-centric algorithms in machine learning. Although the book is not intended to provide exhaustive coverage of machine learning, it serves as a "technical starter" for the key models and optimization methods in machine learning. Even for seasoned practitioners of machine learning, a systematic introduction to fundamental linear algebra and optimization methodologies can be useful in terms of providing a fresh perspective.

The chapters of the book are organized as follows:

1. *Linear algebra and its applications:* The chapters focus on the basics of linear algebra together with their common applications to singular value decomposition, matrix factorization, similarity matrices (kernel methods), and graph analysis. Numerous machine learning applications have been used as examples, such as spectral clustering,

kernel-based classification, and outlier detection. The tight integration of linear algebra methods with examples from machine learning differentiates this book from generic volumes on linear algebra. The focus is clearly on the most relevant aspects of linear algebra for machine learning and to teach readers how to apply these concepts.

2. *Optimization and its applications:* Much of machine learning is posed as an optimization problem in which we try to maximize the accuracy of regression and classification models. The "parent problem" of optimization-centric machine learning is least-squares regression. Interestingly, this problem arises in both linear algebra and optimization and is one of the key connecting problems of the two fields. Least-squares regression is also the starting point for support vector machines, logistic regression, and recommender systems. Furthermore, the methods for dimensionality reduction and matrix factorization also require the development of optimization methods. A general view of optimization in computational graphs is discussed together with its applications to backpropagation in neural networks.

This book contains exercises both within the text of the chapter and at the end of the chapter. The exercises within the text of the chapter should be solved as one reads the chapter in order to solidify the concepts. This will lead to slower progress, but a better understanding. For in-chapter exercises, hints for the solution are given in order to help the reader along. The exercises at the end of the chapter are intended to be solved as refreshers after completing the chapter.

Throughout this book, a vector or a multidimensional data point is annotated with a bar, such as $\overline{X}$ or $\overline{y}$. A vector or multidimensional point may be denoted by either small letters or capital letters, as long as it has a bar. Vector dot products are denoted by centered dots, such as $\overline{X} \cdot \overline{Y}$. A matrix is denoted in capital letters without a bar, such as R. Throughout the book, the $n \times d$ matrix corresponding to the entire training data set is denoted by D, with n data points and d dimensions. The individual data points in D are therefore d-dimensional row vectors and are often denoted by $\overline{X}_1 \ldots \overline{X}_n$. Conversely, vectors with one component for each data point are usually n-dimensional column vectors. An example is the n-dimensional column vector $\overline{y}$ of class variables of n data points. An observed value y_i is distinguished from a predicted value $\hat{y}_i$ by a circumflex at the top of the variable.

Yorktown Heights, NY, USA Charu C. Aggarwal

Acknowledgments

I would like to thank my family for their love and support during the busy time spent in writing this book. Knowledge of the very basics of optimization (e.g., calculus) and linear algebra (e.g., vectors and matrices) starts in high school and increases over the course of many years of undergraduate/graduate education as well as during the postgraduate years of research. As such, I feel indebted to a large number of teachers and collaborators over the years. This section is, therefore, a rather incomplete attempt to express my gratitude.

My initial exposure to vectors, matrices, and optimization (calculus) occurred during my high school years, where I was ably taught these subjects by S. Adhikari and P. C. Pathrose. Indeed, my love of mathematics started during those years, and I feel indebted to both these individuals for instilling the love of these subjects in me. During my undergraduate study in computer science at IIT Kanpur, I was taught several aspects of linear algebra and optimization by Dr. R. Ahuja, Dr. B. Bhatia, and Dr. S. Gupta. Even though linear algebra and mathematical optimization are distinct (but interrelated) subjects, Dr. Gupta's teaching style often provided an integrated view of these topics. I was able to fully appreciate the value of such an integrated view when working in machine learning. For example, one can approach many problems such as solving systems of equations or singular value decomposition either from a linear algebra viewpoint or from an optimization viewpoint, and both perspectives provide complementary views in different machine learning applications. Dr. Gupta's courses on linear algebra and mathematical optimization had a profound influence on me in choosing mathematical optimization as my field of study during my PhD years; this choice was relatively unusual for undergraduate computer science majors at that time. Finally, I had the good fortune to learn about linear and nonlinear optimization methods from several luminaries on these subjects during my graduate years at MIT. In particular, I feel indebted to my PhD thesis advisor James B. Orlin for his guidance during my early years. In addition, Nagui Halim has provided a lot of support for all my book-writing projects over the course of a decade and deserves a lot of credit for my work in this respect. My manager, Horst Samulowitz, has supported my work over the past year, and I would like to thank him for his help.

I also learned a lot from my collaborators in machine learning over the years. One often appreciates the true usefulness of linear algebra and optimization only in an applied setting, and I had the good fortune of working with many researchers from different areas on a wide range of machine learning problems. A lot of the emphasis in this book to specific aspects of linear algebra and optimization is derived from these invaluable experiences and

collaborations. In particular, I would like to thank Tarek F. Abdelzaher, Jinghui Chen, Jing Gao, Quanquan Gu, Manish Gupta, Jiawei Han, Alexander Hinneburg, Thomas Huang, Nan Li, Huan Liu, Ruoming Jin, Daniel Keim, Arijit Khan, Latifur Khan, Mohammad M. Masud, Jian Pei, Magda Procopiuc, Guojun Qi, Chandan Reddy, Saket Sathe, Jaideep Srivastava, Karthik Subbian, Yizhou Sun, Jiliang Tang, Min-Hsuan Tsai, Haixun Wang, Jianyong Wang, Min Wang, Suhang Wang, Wei Wang, Joel Wolf, Xifeng Yan, Wenchao Yu, Mohammed Zaki, ChengXiang Zhai, and Peixiang Zhao.

Several individuals have also reviewed the book. Quanquan Gu provided suggestions on Chapter 6. Jiliang Tang and Xiaorui Liu examined several portions of Chapter 6 and pointed out corrections and improvements. Shuiwang Ji contributed Problem 7.2.3. Jie Wang reviewed several chapters of the book and pointed out corrections. Hao Liu also provided several suggestions.

Last but not least, I would like to thank my daughter Sayani for encouraging me to write this book at a time when I had decided to hang up my boots on the issue of book writing. She encouraged me to write this one. I would also like to thank my wife for fixing some of the figures in this book.

Author Biography

Charu C. Aggarwal is a Distinguished Research Staff Member (DRSM) at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his undergraduate degree in Computer Science from the Indian Institute of Technology at Kanpur in 1993 and his Ph.D. from the Massachusetts Institute of Technology in 1996.

He has worked extensively in the field of data mining. He has published more than 400 papers in refereed conferences and journals and authored more than 80 patents. He is the author or editor of 19 books, including textbooks on data mining, recommender systems, and outlier analysis. Because of the commercial value of his patents, he has thrice been designated a Master Inventor at IBM. He is a recipient of an IBM Corporate Award (2003) for his work on bioterrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and a recipient of two IBM Outstanding Technical Achievement Awards (2009, 2015) for his work on data streams/high-dimensional data. He received the EDBT 2014 Test of Time Award for his work on condensation-based privacy-preserving data mining. He is also a recipient of the IEEE ICDM Research Contributions Award (2015) and the ACM SIGKDD Innovation Award (2019), which are the two highest awards for influential research contributions in data mining.

He has served as the general cochair of the IEEE Big Data Conference (2014) and as the program cochair of the ACM CIKM Conference (2015), the IEEE ICDM Conference (2015), and the ACM KDD Conference (2016). He served as an associate editor of the IEEE Transactions on Knowledge and Data Engineering from 2004 to 2008. He is an associate editor of the IEEE Transactions on Big Data, an action editor of the Data Mining and Knowledge Discovery Journal, and an associate editor of the Knowledge and Information Systems Journal. He serves as the editor-in-chief of the ACM Transactions on Knowledge Discovery from Data as well as the ACM SIGKDD Explorations. He serves on the advisory board of the Lecture Notes on Social Networks, a publication by Springer. He has served as the vice president of the SIAM Activity Group on Data Mining and is a member of the SIAM Industry Committee. He is a fellow of the SIAM, ACM, and IEEE, for "contributions to knowledge discovery and data mining algorithms."