Foundations of Data Science

This book provides an introduction to the mathematical and algorithmic foundations of data science, including machine learning, high-dimensional geometry, and analysis of large networks. Topics include the counterintuitive nature of data in high dimensions, important linear algebraic techniques such as singular value decomposition, the theory of random walks and Markov chains, the fundamentals of and important algorithms for machine learning, algorithms, and analysis for clustering, probabilistic models for large networks, representation learning including topic modeling and nonnegative matrix factorization, wavelets, and compressed sensing. Important probabilistic techniques are developed including the law of large numbers, tail inequalities, analysis of random projections, generalization guarantees in machine learning, and moment methods for analysis of phase transitions in large random graphs. Additionally, important structural and complexity measures are discussed such as matrix norms and VC-dimension. This book is suitable for both undergraduate and graduate courses in the design and analysis of algorithms for data.

Avrim Blum is Chief Academic Officer at the Toyota Technological Institute at Chicago and formerly Professor at Carnegie Mellon University. He has over 25,000 citations for his work in algorithms and machine learning. He has received the AI Journal Classic Paper Award, ICML/COLT 10-Year Best Paper Award, Sloan Fellowship, NSF NYI award, and Herb Simon Teaching Award, and is a fellow of the Association for Computing Machinery.

John Hopcroft is the IBM Professor of Engineering and Applied Mathematics at Cornell University. He is a member National Academy of Sciences and National Academy of Engineering, and a foreign member of the Chinese Academy of Sciences. He received the Turing Award in 1986, was appointed to the National Science Board in 1992 by President George H. W. Bush, and was presented with the Friendship Award by Premier Li Keqiang for his work in China.

Ravindran (Ravi) Kannan is Principal Researcher for Microsoft Research, India. He was the recipient of the Fulkerson Prize in Discrete Mathematics (1991) and the Knuth Prize (ACM) in 2011. He is a distinguished alumnus of Indian Institute of Technology, Bombay, and his past faculty appointments include Massachusetts Institute of Technology, Carnegie-Mellon University, Yale University, and the Indian Institute of Science.

Cambridge University Press
978-1-108-48506-7 — Foundations of Data Science
Avrim Blum , John Hopcroft , Ravi Kannan
Frontmatter
More Information

Foundations of Data Science

Avrim Blum

Toyota Technological Institute at Chicago

John Hopcroft

Cornell University, New York

Ravindran Kannan

Microsoft Research, India

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108485067
DOI: 10.1017/9781108755528

First published 2020

Printed in the United Kingdom by TJ International, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Blum, Avrim, 1966– author. | Hopcroft, John E., 1939– author. |
Kannan, Ravindran, 1953– author.
Title: Foundations of data science / Avrim Blum, Toyota Technological Institute at Chicago,
John Hopcroft, Cornell University, New York, Ravindran Kannan, Microsoft Research, India.
Description: First edition. | New York, NY : Cambridge University Press, 2020. |
Includes bibliographical references and index.
Identifiers: LCCN 2019038133 (print) | LCCN 2019038134 (ebook) |
ISBN 9781108485067 (hardback) | ISBN 9781108755528 (epub)
Subjects: LCSH: Computer science. | Statistics. | Quantitative research.
Classification: LCC QA76 .B5675 2020 (print) | LCC QA76 (ebook) | DDC 004–dc23
LC record available at https://lccn.loc.gov/2019038133
LC ebook record available at https://lccn.loc.gov/2019038134

ISBN 978-1-108-48506-7 Hardback

Cambridge University Press
978-1-108-48506-7 — Foundations of Data Science
Avrim Blum , John Hopcroft , Ravi Kannan
Frontmatter
More Information

Contents