The Principles of Deep Learning Theory

This textbook establishes a theoretical framework for understanding deep learning models of practical relevance. With an approach that borrows from theoretical physics, Roberts and Yaida provide clear and pedagogical explanations of how realistic deep neural networks actually work. To make results from the theoretical forefront accessible, the authors eschew the subject's traditional emphasis on intimidating formality without sacrificing accuracy. Straightforward and approachable, this volume balances detailed first-principle derivations of novel results with insight and intuition for theorists and practitioners alike. This self-contained textbook is ideal for students and researchers interested in artificial intelligence with minimal prerequisites of linear algebra, calculus, and informal probability theory, and it can easily fill a semester-long course on deep learning theory. For the first time, the exciting practical advances in modern artificial intelligence capabilities can be matched with a set of effective principles, providing a timeless blueprint for theoretical research in deep learning.

Daniel A. Roberts was cofounder and CTO of Diffeo, an AI company acquired by Salesforce; a research scientist at Facebook AI Research; and a member of the School of Natural Sciences at the Institute for Advanced Study in Princeton, NJ. He was a Hertz Fellow, earning a PhD from MIT in theoretical physics, and was also a Marshall Scholar at Cambridge and Oxford Universities.

Sho Yaida is a research scientist at Meta AI. Prior to joining Meta AI, he obtained his PhD in physics at Stanford University and held postdoctoral positions at MIT and at Duke University. At Meta AI, he uses tools from theoretical physics to understand neural networks, the topic of this book.

Boris Hanin is an assistant professor at Princeton University in the Operations Research and Financial Engineering Department. Prior to joining Princeton in 2020, Boris was an assistant professor at Texas A&M in the Math Department and an NSF postdoc at MIT. He has taught graduate courses on the theory and practice of deep learning at both Texas A&M and Princeton.

Prepublication praise

"In the history of science and technology, the engineering artifact often comes first: the telescope, the steam engine, digital communication. The theory that explains its function and its limitations often appears later: the laws of refraction, thermodynamics, and information theory. With the emergence of deep learning, AI-powered engineering wonders have entered our lives — but our theoretical understanding of the power and limits of deep learning is still partial. This is one of the first books devoted to the theory of deep learning, and lays out the methods and results from recent theoretical approaches in a coherent manner."

– **Prof. Yann LeCun**, *New York University and Chief AI Scientist at Meta*

"For a physicist, it is very interesting to see deep learning approached from the point of view of statistical physics. This book provides a fascinating perspective on a topic of increasing importance in the modern world."

– **Prof. Edward Witten**, *Institute for Advanced Study*

"This is an important book that contributes big, unexpected new ideas for unraveling the mystery of deep learning's effectiveness, in unusually clear prose. I hope it will be read and debated by experts in all the relevant disciplines."

– **Prof. Scott Aaronson**, *University of Texas at Austin*

"It is not an exaggeration to say that the world is being revolutionized by deep learning methods for AI. But why do these deep networks work? This book offers an approach to this problem through the sophisticated tools of statistical physics and the renormalization group. The authors provide an elegant guided tour of these methods, interesting for experts and non-experts alike. They write with clarity and even moments of humor. Their results, many presented here for the first time, are the first steps in what promises to be a rich research program, combining theoretical depth with practical consequences."

– **Prof. William Bialek**, *Princeton University*

"This book's physics-trained authors have made a cool discovery, that feature learning depends critically on the ratio of depth to width in the neural net."

– **Prof. Gilbert Strang**, *Massachusetts Institute of Technology*

The Principles of Deep Learning Theory

An Effective Theory Approach
to Understanding Neural Networks

DANIEL A. ROBERTS
MIT

SHO YAIDA
Meta AI

based on research in collaboration with

BORIS HANIN
Princeton University

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

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

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/9781316519332
DOI: 10.1017/9781009023405

First published 2022

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

Library of Congress Cataloging-in-Publication Data
Names: Roberts, Daniel A., 1987– author.
Title: The principles of deep learning theory : an effective theory approach to understanding neural networks / Daniel A. Roberts and Sho Yaida based on research in collaboration with Boris Hanin.
Description: New York : Cambridge University Press, 2022. |
Includes bibliographical references and index.
Identifiers: LCCN 2021060635 (print) | LCCN 2021060636 (ebook) |
ISBN 9781316519332 (hardback) | ISBN 9781009023405 (epub)
Subjects: LCSH: Deep learning (Machine learning) |
BISAC: SCIENCE / Physics / Mathematical & Computational
Classification: LCC Q325.73 .R63 2022 (print) | LCC Q325.73 (ebook) |
DDC 006.3/1–dc23/eng20220215
LC record available at https://lccn.loc.gov/2021060635
LC ebook record available at https://lccn.loc.gov/2021060636

ISBN 9781316519332 Hardback

Additional resources for this publication at www.cambridge.org/deeplearningtheory