Preface

> *This has necessitated a complete break from the historical line of development, but this break is an advantage through enabling the approach to the new ideas to be made as direct as possible.*
>
> P. A. M. Dirac in the 1930 preface of *The Principles of Quantum Mechanics* [1].

This is a research monograph in the style of a textbook about the theory of deep learning. While this book might look a little different from the other deep learning books that you've seen before, we assure you that it is appropriate for everyone with knowledge of linear algebra, multivariable calculus, and informal probability theory, and with a healthy interest in neural networks. Practitioner and theorist alike, we want all of you to enjoy this book. Now, let us tell you some things.

First and foremost, in this book we've strived for pedagogy in every choice we've made, placing intuition above formality. This doesn't mean that calculations are incomplete or sloppy; quite the opposite, we've tried to provide full details of every calculation – of which there are certainly very many – and place a particular emphasis on the tools needed to carry out related calculations of interest. In fact, understanding how the calculations are done is as important as knowing their results, and thus often our pedagogical focus is on the details therein.

Second, while we present the details of all our calculations, we've kept the experimental confirmations to the privacy of our own computerized notebooks. Our reason for this is simple: while there's much to learn from explaining a derivation, there's not much more to learn from printing a verification plot that shows two curves lying on top of each other. Given the simplicity of modern deep-learning packages and the availability of compute, it's easy to verify any formula on your own; we certainly have thoroughly checked them all this way, so if knowledge of the existence of such plots is comforting to you, know at least that they do exist on our personal and cloud-based hard drives.

Third, our main focus is on realistic models that are used by the deep learning community in practice: we want to study *deep* neural networks. In particular, this means that *(i)* a number of special results on single-hidden-layer networks will not be discussed and *(ii)* the *infinite-width limit* of a neural network – which is equivalent to a zero-hidden-layer network – will be introduced only as a starting point. All such idealized models will eventually be *perturbed* until they correspond to a real model. We certainly acknowledge that there's a vibrant community of deep-learning theorists devoted to

https://doi.org/10.1017/9781009023405.001 Published online by Cambridge University Press

exploring different kinds of idealized theoretical limits. However, our interests are fixed firmly on providing explanations for the tools and approaches used by practitioners, in an effort to shed light on what makes them work so well.

Fourth, a large part of the book is focused on deep multilayer perceptrons. We made this choice in order to pedagogically illustrate the power of the effective theory framework – not due to any technical obstruction – and along the way we give pointers for how this formalism can be extended to other architectures of interest. In fact, we expect that many of our results have a broad applicability, and we've tried to focus on aspects that we expect to have lasting and universal value to the deep learning community.

Fifth, while much of the material is novel and appears for the first time in this book, and while much of our framing, notation, language, and emphasis breaks with the historical line of development, we're also very much indebted to the deep learning community. With that in mind, throughout the book we will try to reference important prior contributions, with an emphasis on recent seminal deep-learning results rather than on being completely comprehensive. Additional references for those interested can easily be found within the work that we cite.

Sixth, this book initially grew out of a research project in collaboration with Boris Hanin. To account for his effort and then support, we've accordingly commemorated him on the cover. More broadly, we've variously appreciated the artwork, discussions, encouragement, epigraphs, feedback, management, refereeing, reintroduction, and support from Rafael Araujo, Léon Bottou, Paul Dirac, Ethan Dyer, John Frank, Ross Girshick, Vince Higgs, Yoni Kahn, Yann LeCun, Kyle Mahowald, Eric Mintun, Xiaoliang Qi, Mike Rabbat, David Schwab, Stephen Shenker, Eva Silverstein, PJ Steiner, DJ Strouse, and Jesse Thaler. Organizationally, we're grateful to FAIR and Facebook, Diffeo and Salesforce, MIT and IAIFI, and Cambridge University Press and the arXiv.

Seventh, given intense (and variously uncertain) spacetime and energy-momentum commitment that writing this book entailed, Dan is grateful to Aya, Lumi, and Lisa Yaida; from the dual sample-space perspective, Sho is grateful to Adrienne Rothschilds and would be retroactively grateful to any hypothetical future Mark or Emily that would have otherwise been thanked in this paragraph.

Eighth, we hope that this book spreads our optimism that it *is* possible to have a general theory of deep learning, one that's both derived from first principles and at the same time focused on describing how realistic models actually work: nearly-simple phenomena in practice should correspond to nearly-simple effective theories. We dream that this type of thinking will not only lead to more *[redacted]* AI models but also guide us toward a unifying framework for understanding universal aspects of intelligence.

As if that eightfold way of prefacing the book wasn't nearly-enough already, please note: this book has a website, deeplearningtheory.com, and you may want to visit it in order to determine whether the error that you just discovered is already common knowledge. If it's not, please let us know.

Dan Roberts & Sho Yaida
Remotely Located

https://doi.org/10.1017/9781009023405.001 Published online by Cambridge University Press