

Preface

This book is about boosting, an approach to machine learning based on the idea of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules. A remarkably rich theory has evolved around boosting, with connections to a wide range of topics including statistics, game theory, convex optimization, and information geometry. In addition, AdaBoost and other boosting algorithms have enjoyed practical success with applications, for instance, in biology, vision, and speech processing. At various times in its history, boosting has been the subject of controversy for the mystery and paradox that it seems to present.

In writing this book, we have aimed to reach nearly anyone with an interest in boosting (as well as an appropriate, but relatively minimal, technical background), whether students or advanced researchers, whether trained in computer science, statistics, or some other field. We specifically hope that the book will be useful as an educational tool, and have therefore included exercises in every chapter. Although centered on boosting, the book introduces a variety of topics relevant to machine learning generally, as well as to related fields such as game theory and information theory.

The main prerequisite for this book is an elementary background in probability. We also assume familiarity with calculus and linear algebra at a basic, undergraduate level. An appendix provides background on some more advanced mathematical concepts which are used mainly in later chapters. The central notions of machine learning, boosting, and so on are all presented from the ground up.

Research on boosting has spread across multiple publications and disciplines over a period of many years. This book attempts to bring together, organize, extend, and simplify a significant chunk of this work. Some of this research is our own or with co-authors, but a very large part of what we present—including a few of the chapters almost in their entirety—is based on the contributions of the many other excellent researchers who work in this area. Credit for such previously published work is given in the bibliographic notes at the end of every chapter. Although most of the material in this book has appeared elsewhere, the majority of chapters also include new results that have never before been published.

The focus of this book is on foundations and algorithms, but also on applications. Following a general introduction to machine learning algorithms and their analysis, the book explores in part I the core theory of boosting, particularly its ability to generalize (that is, make accurate predictions on new data). This includes an analysis of boosting's training error, as well as bounds on the generalization error based both on direct methods and on the margins theory. Next, part II systematically explores some of the other myriad theoretical viewpoints that have helped to explain and understand boosting, including the game-theoretic interpretation, the view of AdaBoost as a greedy procedure for minimizing a loss function, and an understanding of boosting as an iterative-projection algorithm with connections to information geometry and convex optimization. Part III focuses on practical extensions of AdaBoost based on the use of confidence-rated weak hypotheses, and for multiclass and ranking problems. Finally, some advanced theoretical topics are covered in part IV, including the statistical consistency of AdaBoost, optimal boosting, and boosting algorithms which operate in continuous time. Although the book is organized around theory and algorithms, most of the chapters include specific applications and practical illustrations.

Readers with particular interests, or those organizing a course, might choose one of a number of natural tracks through this book. For a more theoretical treatment, part III could be omitted. A track focused on the practical application of boosting might omit chapters 4, 6, and 8, and all of part IV. A statistical approach might emphasize chapters 7 and 12 while omitting chapters 4, 6, 8, 13, and 14. Some of the proofs included in this book are somewhat involved and technical, and can certainly be skipped or skimmed. A rough depiction of how the chapters depend on each other is shown in figure P.1.

This book benefited tremendously from comments and criticisms we received from numerous individuals. We are especially grateful to ten students who read an earlier draft of the book as part of a Princeton graduate seminar course: Jonathan Chang, Sean Gerish, Sina Jafarpour, Berk Kapicioglu, Indraneel Mukherjee, Gungor Polatkan, Alexander

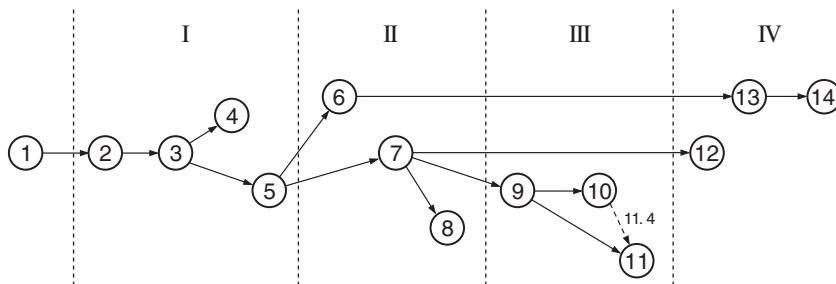


Figure P.1

An approximate depiction of how the chapters of this book depend on each other. Each edge $u \rightarrow v$ represents a suggestion that chapter u be read before chapter v . (The dashed edge indicates that section 11.4 depends on chapter 10, but the other sections of chapter 11 do not.)

Schwing, Umar Syed, Yongxin (Taylor) Xi, and Zhen (James) Xiang. Their close reading and numerous suggestions, both in and out of class, were extraordinarily helpful and led to significant improvements in content and presentation in every one of the chapters.

Thanks also to Peter Bartlett, Vladimir Koltchinskii, Saharon Rosset, Yoram Singer, and other anonymous reviewers of this book for their time and their many constructive suggestions and criticisms. An incomplete list of the many, many others who provided help, comments, ideas, and insights includes: Shivani Agarwal, Jordan Boyd-Graber, Olivier Chapelle, Kamalika Chaudhuri, Michael Collins, Edgar Dobriban, Miro Dudík, Dave Helmbold, Ludmila Kuncheva, John Langford, Phil Long, Taesup Moon, Lev Reyzin, Ron Rivest, Cynthia Rudin, Rocco Servedio, Matus Telgarsky, Paul Viola, and Manfred Warmuth. Our apologies to others who were surely, though certainly not intentionally, omitted from this list.

We are grateful to our past and present employers for supporting this work: AT&T Labs; Columbia Center for Computational Learning Systems; Princeton University; University of California, San Diego; and Yahoo! Research. Support for this research was also generously provided by the National Science Foundation under awards 0325463, 0325500, 0513552, 0812598, and 1016029.

Thanks to all of our collaborators and colleagues whose research appears in this book, and who kindly allowed us to include specific materials, especially figures, as cited and acknowledged with gratitude in the appropriate chapters. We are grateful to Katherine Almeida, Ada Brunstein, Jim DeWolf, Marc Lowenthal, Molly Seamans and everyone at MIT Press for their tireless assistance in preparing and publishing this book. Thanks also to the various editors at other publishers we considered, and to all those who helped with some occasionally thorny copyright issues, particularly Laurinda Alcorn and Frank Politano.

Finally, we are grateful for the love, support, encouragement, and patience provided by our families: Roberta, Jeni, and Zak; Laurie, Talia, and Rafi; and by our parents: Hans and Libby, Ora and Rafi.

