

Preface

In the last decade the user base of machine learning has grown dramatically. From a relatively small circle in computer science, engineering, and mathematics departments the users of machine learning now include students and researchers from every corner of the academic universe, as well as members of industry, data scientists, entrepreneurs, and machine learning enthusiasts. The book before you is the result of a complete tearing down of the standard curriculum of machine learning into its most basic components, and a curated reassembly of those pieces (painstakingly polished and organized) that we feel will most benefit this broadening audience of learners. It contains fresh and intuitive yet rigorous descriptions of the most fundamental concepts necessary to conduct research, build products, tinker, and play.

Intended audience and book pedagogy

This book was written for readers interested in understanding the core concepts of machine learning from first principles to practical implementation. To make full use of the text one only needs a basic understanding of linear algebra and calculus (i.e., vector and matrix operations as well as the ability to compute the gradient and Hessian of a multivariate function), as well as some prior exposure to fundamental concepts of computer programming (i.e., conditional and looping structures). It was written for first time learners of the subject, as well as for more knowledgeable readers who yearn for a more intuitive and serviceable treatment than what is currently available today.

To this end throughout the text, in describing the fundamentals of each concept, we defer the use of probabilistic, statistical, and neurological views of the material in favor of a fresh and consistent geometric perspective. We believe that this not only permits a more intuitive understanding of many core concepts, but helps establish revealing connections between ideas often regarded as fundamentally distinct (e.g., the logistic regression and support vector machine classifiers, kernels and feed-forward neural networks). We also place significant emphasis on the design and implementation of algorithms, and include many coding exercises for the reader to practice at the end of each chapter. This is because we

strongly believe that the bulk of learning this subject takes place when one “gets their hands dirty” and code things up for themselves. In short, with this text we have aimed to create a learning experience for the reader where intuitive leaps precede intellectual ones and are tempered by their application.

What this book is about

The core concepts of our treatment of machine learning can be broadly summarized in four categories. *Predictive learning*, the first of these categories, comprises two kinds of tasks where we aim to either predict a continuous valued phenomenon (like the future location of a celestial body), or distinguish between distinct kinds of things (like different faces in an image). The second core concept, *feature design*, refers to a broad set of engineering and mathematical tools which are crucial to the successful performance of predictive learning models in practice. Throughout the text we will see that features are generated along a spectrum based on the level of our own understanding of a dataset. The third major concept, *function approximation*, is employed when we know too little about a dataset to produce proper features ourselves (and therefore must learn them strictly from the data itself). The final category, *numerical optimization*, powers the first three and is the engine that makes machine learning run in practice.

0.1 Overview of the book

This book is separated into three stages, with the latter parts building thematically on each preceding stage.

Stage I: Fundamental tools and concepts

Here we detail the fundamentals of predictive modeling, numerical optimization, and feature design. After a general introduction in Chapter 1, we begin with Chapter 2 by introducing the rudiments of numerical optimization, those critical tools used to properly tune predictive learning models. We then introduce predictive modeling in Chapters 3 and 4, where the regression and classification tasks are introduced, respectively. Along the way we also describe a number of examples where we have some level of knowledge about the underlying process generating the data we receive, which can be leveraged for the design of features.

Stage 2: Tools for fully data driven machine learning

In the absence of useful knowledge about our data we must broaden our perspective in order to design, or learn, features for regression and classification tasks. In Chapters 5 and 6 we review the classical tools of *function approximation*, and see how they are applied to deal with general regression and classification problems. We then end in Chapter 7 by describing several advanced topics related to the material in the preceding two Chapters.

Stage 3: Methods for large scale machine learning

In the final stage of the book we describe common procedures for scaling regression and classification algorithms to large datasets. We begin in Chapter 8 by introducing a number of advanced numerical optimization techniques. A continuation of the introduction in Chapter 2, these methods greatly enhance the power of predictive learning by means of more effective optimization algorithms. We then detail in Chapter 9 general techniques for properly lowering the dimension of input data, allowing us to deflate large datasets down to more manageable sizes.

Readers: how to use this book

As mentioned earlier the only technical prerequisites for the effective use of this book are a basic understanding of linear algebra and vector calculus, as advanced concepts are introduced as necessary throughout the text, as well as some prior computer programming experience. Readers may find a brief tutorial on the Python and MATLAB/OCTAVE programming environments used for completing coding exercises, which introduces proper syntax for both languages as well as necessary libraries to download (for Python) as well as useful built-in functions (for MATLAB/OCTAVE), on the book website.

For self study one may read all the Chapters in order, as each builds on its direct predecessor. However a solid understanding of the first six Chapters is sufficient preparation for perusing individual topics of interest in the final three Chapters of the text.

Instructors: how to use this book

The contents of this book have been used for a number of courses at Northwestern University, ranging from an introductory course for senior level undergraduate and beginning graduate students, to a specialized course on advanced numerical optimization for an audience largely consisting of PhD students. Therefore with its treatment of foundations, applications, and

algorithms this book is largely self contained and can be used for a variety of machine learning courses. For example, it may be used as the basis for

A single quarter or semester long senior undergraduate/beginning graduate level introduction to standard machine learning topics. This includes coverage of basic techniques from numerical optimization, regression/classification techniques and applications, elements of feature design and learning, and feed-forward neural networks. Chapters 1 - 6 provide the basis for such a course, with Chapters 7 and 9 (on kernel methods and dimension reduction/unsupervised learning techniques) being optimal add ons.

A single quarter or semester long senior level undergraduate/graduate course on large-scale optimization for machine learning. Chapters 2 and 6 - 8 provide the basis for course on introductory and advanced optimization techniques for solving the applications and models introduced in the first two-thirds of the book.