Richard M. Golden

Statistical Machine Learning: A Unified Framework

Contents

Preface

Objectives

Statistical Machine Learning is a multidisciplinary field that integrates topics from the fields of Machine learning, Mathematical Statistics, and Numerical Optimization Theory. It is concerned with the problem of the development and evaluation of machines capable of inference and learning within an environment characterized by statistical uncertainty. The recent rapid growth in the variety and complexity of new machine learning architectures requires the development of improved methods for communicating relevant technological tools for supporting statistical machine learning algorithm analysis and design. The main objective of this textbook is to provide students, engineers, and scientists with practical established tools from mathematical statistics and nonlinear optimization theory to support the analysis and design of both existing and new state-of-the-art machine learning algorithms.

It is important to emphasize that this is a mathematics textbook intended for readers interested in a concise mathematically rigorous introduction to the statistical machine learning literature. For readers interested in non-mathematical introductions to the machine learning literature, many alternative options are available. For example, there are many useful software-oriented machine learning textbooks which support the rapid development and evaluation of a wide range of machine learning architectures (Geron, 2017; Muller and Guida, 2017, Bell, 2015, James et al., 2017). A student can use these software tools to rapidly create and evaluate a bewildering wide range of machine learning architectures. After an initial exposure to such tools, the student will want to obtain a deeper understanding of such systems in order to properly apply and properly evaluate such tools. To address this issue, there are now many excellent textbooks (e.g., Hastie el al., 2001; Bishop et al., Stork et al., Ripley et al.; Hastie and Tibshirani, 2016; Goodfellow and Bengio, 2016) which provide detailed discussions of a variety of machine learning architectures and principles by focusing attention on basic principles. Such textbooks specifically omit particular technical mathematical details under the assumption that students without the relevant technical background should not be distracted, while students with graduate level training in optimization theory and mathematical statistics can obtain such details elsewhere.

However, such mathematical technical details are essential for providing a principled methodology for supporting the communication, analysis, and design of novel nonlinear machine learning architectures. Thus, it is desirable to explicitly incorporate such details into self-contained concise discussions of machine learning applications. Technical mathematical details support improved methods for machine learning algorithm specification, validation, classification, and understanding. Such methods can provide important support for rapid machine learning algorithm development and deployment as well as novel insights into reusable modular software design architectures.

Book Overview

A distinguishing feature of this textbook is that a particular empirical risk minimization framework is introduced for the purpose of analyzing both the asymptotic behavior and generalization performance of commonly encountered machine learning algorithms. In particular, a small set of explicit theorems define a useful pedagogical framework for understanding machine learning algorithms. Explicit examples from the machine learning literature are provided to show students how to properly interpret the assumptions and conclusions of such theorems. Machine learning algorithms that do not conform to this unified framework are easily identified as exceptional cases.

Part 1 is concerned with introducing the concept of machine learning algorithms through examples and providing mathematical tools for specifying such algorithms. Chapter 1 informally shows, by example, that the large class of supervised, unsupervised, and reinforcement learning algorithms which are the focus of this textbook may be interpreted as nonlinear optimization algorithms. Chapter 3 provides a formal description of this large class of nonlinear optimization algorithms and shows how optimization may be semantically interpreted within a rational decision making framework.

Part 2 is concerned with characterizing the asymptotic behavior of deterministic learning machines. Chapter 6 provides sufficient conditions for characterizing the asymptotic behavior of discrete-time and continuous-time time-invariant dynamical systems. Chapter 7 provides sufficient conditions for ensuring a large class of deterministic batch learning algorithms converge to the critical points of the objective function for learning.

Part 3 is concerned with characterizing the asymptotic behavior of stochastic inference and stochastic learning machines. Chapter 11 develops the asymptotic convergence theory for Monte Carlo Markov Chains for the special case where the Markov chain is defined on a finite state space. Chapter 12 provides relevant asymptotic convergence analyses of adaptive learning algorithms for both passive and reactive learning environments.

Part 4 is concerned with the problem of characterizing the generalization performance of a machine learning algorithm. Chapter 13 discusses the analysis and design of semantically interpretable objective functions. Chapters 14, 15, and 16 show how both bootstrap simulation methods (Chapter 14) and asymptotic formulas (Chapters 15, 16) can be used to characterize the generalization performance of the class of machine learning algorithms considered here.

In addition, the book includes self-contained relevant introductions to real analysis (Chapter 2, 5), linear algebra (Chapter 4), measure theory (Chapter 8), and stochastic sequences (Chapter 9) to reduce the required mathematical prerequisites for the analyses presented here.

Targeted Audience

The textbook is written for a multidisciplinary audience with multidisciplinary objectives. It is assumed students taking a course based upon this book have taken lower-division coursework in linear algebra and calculus as well as an upper-division calculus-based probability theory course. Students with these mathematical prerequisites will find this textbook challenging but nevertheless accessible.