Machine Learning in Finance: From Theory to Practice

Paul Bilokon, Matthew Dixon and Igor Halperin

> “Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth.”
>
> Arthur Canon Doyle

Paul Bilokon
Thalesians Ltd, London, e-mail: paul@thalesians.com

Matthew Dixon
Department of Applied Math, Illinois Institute of Technology e-mail: matthew.dixon@iit.edu

Igor Halperin
NYU Tandon School of Engineering and Fidelity Investments, e-mail: igor.halperin@nyu.edu, igor.halperin@fmr.com

Introduction

Machine learning in finance sits at the intersection of a number of emergent and established disciplines including pattern recognition, financial econometrics, statistical computing, probabilistic programming and dynamic programming. With the trend towards increasing computational resources and larger datasets, machine learning has grown into a central computational engineering field, with an emphasis placed on plug-and-play algorithms made available through open-source machine learning toolkits. Algorithm focused areas of finance, such as algorithmic trading have been the primary adopters of this technology. But outside of engineering based research groups and business activities, much of the field remains a mystery.

A key barrier to understanding machine learning for non-engineering students and practitioners is the absence of the well established theories and concepts that financial time series analysis equips us with. These serve as the basis for development of financial modeling intuition and scientific reasoning. Moreover, machine learning is heavily entrenched in engineering ontology, which makes developments in the field somewhat intellectually inaccessible for students, academics and finance practitioners from the quantitative disciplines such as mathematics, statistics, physics and economics. Consequently, there is a great deal of misconception and limited understanding of the capacity of this field. While machine learning techniques are often effective, they remain poorly understood and are often mathematically indefensible. How do we place key concepts in the field of machine learning in the context of more foundational theory in time series analysis, econometrics and mathematical statistics? Under which simplifying conditions are advanced machine learning techniques such as deep neural networks mathematically equivalent to well known statistical models such as linear regression? How should we reason about the perceived benefits of using advanced machine learning methods over more traditional econometrics methods, for different financial applications? What theory supports the application of machine learning to problems in financial modeling? How does reinforcement learning provide a model-free approach to the Black-Scholes-Merton model for derivative pricing? How does Q-learning generalize discrete time stochastic control problems in finance?

This book is written for advanced graduate students and academics in financial econometrics, management science and applied statistics, in addition to quants and data scientists in the field of quantitative finance. We present machine learning as a non-linear extension of various topics in quantitative economics such as financial econometrics and dynamic programming, with an emphasis on novel algorithmic representations of data, regularization, and techniques for controlling the bias-variance tradeoff leading to improved out-of-sample forecasting. The book is presented in three parts, each part covering theory and applications. The first presents supervised learning for cross-sectional data from both a Bayesian and frequentist perspective. The more advanced material places a firm emphasis on neural networks, including deep learning, as well as Gaussian processes, with examples in investment management and derivatives. The second part covers supervised learning for time series data, arguably the most common data type used in finance with examples in trading, stochastic volatility and fixed income modeling. Finally, the third part covers reinforcement learning and its applications in trading, investment and wealth management. We provide python code examples to support the readers' understanding of the methodologies and applications. As a bridge to research in this emergent field, we present the frontiers of machine learning in finance from a researcher's perspective, highlighting how many well known concepts in statistical physics are likely to emerge as research topics for machine learning in finance.

Prerequisites

This book is targeted at graduate students in data science, mathematical finance, financial engineering and operations research seeking a career in quantitative finance, data science, analytics and fintech. Students are expected to have completed upper section undergraduate courses in linear algebra, multivariate calculus, advanced probability theory and stochastic processes, statistics for time series (econometrics), and gained some basic introduction to numerical optimization and computational mathematics. Students shall find the later chapters of this book, on reinforcement learning, more accessible with some background in investment science. Students should also have prior experience with Python programming and, ideally, taken a course in computational finance and introductory machine learning. The material in this book is more mathematical and less engineering focused than most courses on machine learning and for this reason we recommend reviewing the recent book, "Linear Algebra and Learning from Data" by Gilbert Strang as background reading.

Advantages of the book

Readers will find this book useful as a bridge from well established foundational topics in financial econometrics to applications of machine learning in finance. Statistical machine learning is presented as a non-parametric extension of financial econometrics and quantitative finance, with an emphasis on novel algorithmic representations of data, regularization and model averaging to improve out-of-sample forecasting. The key distinguishing feature from classical financial econometrics and dynamic programming is the absence of an assumption on the data generation process. This has important implications for modeling and performance assessment which are emphasized with examples throughout the book. Some of the main contributions of the book are:

- The textbook market is saturated with excellent books on machine learning. However, few present the topic from the prospective of financial econometrics and cast fundamental concepts in machine learning into canonical modeling and decision frameworks already well established in finance such as financial time series analysis, investment science and financial risk management. Only through the integration of these disciplines can we develop an intuition into how machine learning theory informs the practice of financial modeling.

- Machine learning is entrenched in engineering ontology, which makes developments in the field somewhat intellectually inaccessible for students, academics and finance practitioners from quantitative disciplines such as mathematics, statistics, physics and economics. Moreover, financial econometrics has not kept pace with this transformative field and there is a need to reconcile various modeling concepts between these disciplines. This textbook is built around powerful mathematical ideas that shall serve as the basis for a graduate course for students with prior training in probability and advanced statistics, linear algebra, times series analysis and Python programming.
- This book provides financial market motivated and compact theoretical treatment of financial modeling with machine learning for the benefit of regulators, wealth managers, federal research agencies and professionals in other heavily regulated business functions in finance who seek a more theoretical exposition to allay concerns about the “black-box” nature of machine learning.
- Reinforcement learning is presented as a model-free framework for stochastic control problems in finance, covering portfolio optimization, derivative pricing and wealth management applications without assuming a data generation process. We also provide a model-free approach to problems in market microstructure, such as optimal execution, with Q-learning. Furthermore, our book is the first to present on methods of Inverse Reinforcement Learning.
- Multiple-choice questions, numerical examples and more than 80 end-of-chapter exercises are used through out the book to reinforce key technical concepts.
- This book provides Python codes demonstrating the application of machine learning to algorithmic trading and financial modeling in risk management and equity research. These codes make use of powerful open-source software toolkits such as Google’s Tensorflow and Pandas, a data processing environment for Python.

Overview of the book

Chapter 1

provides the industry context for machine learning in finance, discussing the critical events that have shaped the finance industry’s need for machine learning and the unique barriers to adoption. The finance industry has adopted machine learning to varying degrees of sophistication. How it has been adopted is heavily fragmented by the academic disciplines underpinning the applications. We view some key mathematical examples that demonstrate the nature of machine learning and how it is used in practice, with the focus on building intuition for more technical expositions in later chapters. In particular, we begin to address many finance practitioner’s concerns that neural networks are a “black-box” by showing how they are related to existing well established techniques such as linear regression, logistic regression and autoregressive time series models. Such arguments are developed further in later chapters.

Chapter 2

introduces probabilistic modeling and reviews foundational concepts in Bayesian econometrics such as Bayesian inference, model selection, online learning and Bayesian model averaging. We develop more versatile representations of complex data with probabilistic graphical models such as Mixture Models and Hidden Markov Models.

Chapter 3

introduces Bayesian regression and shows how it extends many of the concepts in the previous chapter. We develop kernel based machine learning methods —specifically Gaussian process regression, an important class of Bayesian machine learning methods — and demonstrate their application to "surrogate" models of derivative prices. This chapter also provides a natural point from which to develop intuition for the role and functional form of regularization in a frequentist setting — the subject of subsequent chapters.

Chapter 4

provides a more in depth description of supervised learning, deep learning and neural networks — presenting the foundational mathematical and statistical learning concepts and explaining how they relate to real-world examples in trading, risk management and investment management. These applications present challenges for forecasting and model design and are presented as a reoccurring theme through out the book. This chapter moves towards a more engineering style exposition of neural networks, applying concepts in the previous chapters to elucidate various model design choices.

Chapter 5

presents a method for interpreting neural networks which imposes minimal restrictions on the neural network design. The chapter demonstrates techniques for interpreting a feedforward network, including how to rank the importance of the features. An example demonstrating how to apply interpretability analysis to deep learning models for factor modeling is also presented.

Chapter 6

provides an overview of the most important modeling concepts in financial econometrics. Such methods form the conceptual basis and performance baseline for more advanced neural network architectures presented in the next chapter. In fact, each type of architecture is a generalization of many of the models presented here. This chapter is especially useful for students from an engineering or science background, with little exposure to econometrics and time series analysis.

Chapter 7

presents a powerful class of probabilistic models for financial data. Many of these models overcome some of the severe stationarity limitations of the frequentist models in the previous chapters. The fitting procedure demonstrated is also different — the use of Kalman filtering algorithms for state-space models rather than maximum likelihood estimation or Bayesian inference. Simple examples of Hidden Markov models and particle filters in finance and various algorithms are presented.

Chapter 8

presents various neural network models for financial time series analysis, providing examples of how they relate to well known techniques in financial econometrics. Recurrent Neural Networks (RNNS) are presented as non-linear time series models and generalize classical linear time series models such as $AR(p)$. They provide a powerful approach for prediction in financial time series and generalize to non-stationary data. The chapter also presents Convolution Neural Networks for filtering time series data and exploiting different scales in the data. Finally, this chapter demonstrates how auto-encoders are used to compress information and generalize principal component analysis.

Chapter 9

introduces Markov Decision Processes and the classical methods of dynamic programming, before building familiarity with the ideas of Reinforcement Learning and other approximate methods for solving MDPs. After describing Bellman optimality and iterative value and policy updates before moving to Q-learning, the chapter quickly advances towards a more engineering style exposition of the topic, covering key computational concepts such as greediness, batch learning and Q-learning. Through a number of mini-case studies, the chapter provides insight into how RL is applied to optimization problems in asset management and trading. These examples are each supported with Python notebooks.

Chapter 10

considers real-world applications of reinforcement learning in finance, as well as further advances the theory presented in the previous chapter. We start with one of the most common problems of quantitative finance, which is the problem of optimal portfolio trading in discrete time. Many practical problems of trading or risk management amount to different forms of dynamic portfolio optimization, with different optimization criteria, portfolio composition, and constraints. The chapter introduces a reinforcement learning approach to option pricing that generalizes the classical Black-Scholes model to a data-driven approach using Q-learning. It then presents a probabilistic extension of Q-learning called G-learning, and shows how it can be used for dynamic portfolio optimization. For certain specifications of reward functions, G-learning is semi-analytically tractable, and amounts to a probabilistic version of Linear Quadratic Regulators (LQR). Detailed analyses of such cases are presented, and show their solutions with examples from problems of dynamic portfolio optimization and wealth management.

Chapter 11

provides an overview of the most popular methods of Inverse Reinforcement Learning (IRL) and Imitation Learning (IL). These methods solve the problem of optimal control in a data-driven way, similarly to reinforcement learning, however with the critical difference that now rewards are *not* observed. The problem is rather to learn the reward function from the observed behavior of an agent. As behavioral data without rewards are widely available, the problem of learning from such data is certainly very interesting. The chapter provides a moderate-level description of the most promising IRL methods, equips the reader with sufficient knowledge to understand and follow the current literature on IRL, and presents examples that use simple simulated environments to see how these methods perform when we know the “ground truth" rewards. We then present use cases for IRL in quantitative finance that include applications to trading strategy identification, sentiment-based trading, option pricing, inference of portfolio investors, and market modeling.

Chapter 12

takes us forward to emerging research topics in quantitative finance and machine learning. Among many interesting emerging topics, we focus here on two broad themes. The first one deals with unification of Supervised Learning and Reinforcement Learning as two tasks of perception-action cycles of agents. We outline some recent research ideas in the literature including in particular information theory-based versions of Reinforcement Learning, and discuss their relevance for financial applications. We explain why these ideas might have interesting practical implications for RL financial models, where feature selection could be done within the general task of optimization of a long-term objective, rather than outside of it, as is usually performed in "alpha-research".

The second topic presented in this chapter deals with using methods of Reinforcement Learning to construct models of market dynamics. We also introduce some advanced physics-based approaches for computations for such RL-inspired market models.

Source code

Many of the chapters are accompanied by Python notebooks to illustrate some of the main concepts and demonstrate application of machine learning methods. Each notebook is lightly annotated. Many of these notebooks use TensorFlow. We recommend loading these notebooks, together with any accompanying Python source files and data, in Google Colab. Please see the appendices of each chapter accompanied by notebooks, and the `README.md` in the subfolder of each chapter, for further instructions and details.

Scope

We recognize that the field of machine learning is developing rapidly and to keep abreast of the research in this field is a challenging pursuit. Machine learning is an umbrella term for a number of methodology classes, including supervised learning, unsupervised learning and reinforcement learning. This book focuses on supervised learning and reinforcement learning because these are the areas with the most overlap with econometrics, predictive modeling, and optimal control in finance. Supervised machine learning can be categorized as generative and discriminative. Our focus is on discriminative learners which attempt to partition the input space, either directly, through affine transformations or through projections onto a manifold. Neural networks have been shown to provide a universal approximation to a wide class of functions. Moreover they can be shown to reduce to other well known statistical techniques and are adaptable to time series data.

Extending time series models, a number of chapters in this book are devoted to a introduction to reinforcement learning (RL) and inverse reinforcement learning

(IRL), that deal with problems of optimal control of such time series, and show how many classical financial problems such as portfolio optimization, option pricing and wealth management can naturally be posed as problems for RL and IRL. We present simple RL methods that can be applied for these problems, as well as explain how neural networks can be used in these applications.

There are already several excellent textbooks covering other classical machine learning methods and we instead choose to focus on how to cast machine learning into various financial modeling and decision frameworks. We emphasize that much of this material is not unique to neural networks, but comparisons of alternative supervised learning approaches, such as random forests, are beyond the scope of this book.

Multiple choice questions

Multiple choice questions are included after introducing a key concept. The correct answers to all questions are provided at the end of each chapter with selected, partial, explanations to some of the more challenging material.

Exercises

The exercises that appear at the end of every chapter form an important component of the book. Each exercise has been chosen to reinforce concepts explained in the text, to stimulate the application of machine learning in finance and to gently bridge material in other chapters. Each is graded according to difficulty ranging from (*), which denotes a simple exercise which might take a few minutes to complete, through to (***), which denotes a significantly more complex exercise. Unless specified otherwise, all equations referred in each exercise correspond to those in the corresponding chapter.

Instructor materials

The book is supplemented by a separate Instructor's manual which provides worked solutions to the end of chapter questions. Full explanations for the solutions to the multiple choice questions are also provided. The manual provides additional notes and example code solutions for some of the programming exercises in the later chapters.

Acknowledgements

Peter Carr, Chair of the Department of Financial Engineering at NYU Tandon has been instrumental in supporting the growth of the field of machine learning in finance. Through providing speaker engagements and machine learning instructorship positions in the MS in Algorithmic Finance Program, the authors have been able to write research papers and identify the key areas required by a text book. Miquel Alonso (AIFI), Rama Cont (Oxford), Kay Giesecke (Stanford), Ali Hirsa (Columbia), Sebastian Jaimungal (University of Toronto), Gary Kazantsev (Bloomberg), Morton Lane (UIUC), Jörg Osterrieder (ZHAW), have established various academic and joint academic-industry workshops and community meetings to proliferate the field and serve as input for this book. At the same time, there has been growing support for the development of a book in London, where several SIAM/LMS workshops and practitioner special interest groups, such as the Thalesians, have identified a number of compelling financial applications. The material has grown from courses and invited lectures at NYU, UIUC, Illinois Tech, Imperial College and the 2019 Bootcamp on Machine Learning in Finance at the Fields Institute, Toronto.

Along the way, we have been fortunate to receive the support of Tomasz Bielecki (Illinois Tech), Ali Hirsa (Columbia University) and Brian Peterson (DV Trading). Special thanks to research collaborators and colleagues Kay Giesecke (Stanford University), Diego Klabjan (NWU), Nick Polson (Chicago Booth) and Harvey Stein (Bloomberg), all of whom have shaped our understanding of the emerging field of machine learning in finance and the many practical challenges. We are indebted to Sri Krishnamurthy (QuantUniversity), Saeed Amen (Cuemacro), and Nicole Königstein (ZHAW) for their valuable input on this book. Finally we acknowledge the support of a number of Illinois Tech graduate students who have contributed to the source code examples and exercises: Xiwen Jing, Bo Wang and Siliang Xong. Special thanks to Swaminathan Sethuraman for his support of the code development, and of Volod Chernat and George Gvishiani who provided support and code development for the course taught at NYU and Coursera.

Author biographies

Paul Bilokon

is CEO and Founder of Thalesians Ltd and an expert in electronic and algorithmic trading across multiple asset classes, having helped build such businesses at Deutsche Bank and Citigroup. Before focusing on electronic trading, Paul worked on derivatives and has served in quantitative roles at Nomura, Lehman Brothers, and Morgan Stanley. Paul has been educated at Christ Church College, Oxford, and Imperial College. Apart from mathematical and computational finance, his academic interests include machine learning and mathematical logic.

Matthew Dixon

is an Assistant Professor of Applied Math at the Illinois Institute of Technology. His research in computational methods for finance is funded by Intel. Matthew began his career in structured credit trading at Lehman Brothers in London before pursuing academics and consulting for financial institutions in quantitative trading and risk modeling. He holds a Ph.D. in Applied Mathematics from Imperial College (2007) and has held postdoctoral and visiting professor appointments at Stanford University and UC Davis respectively. He has published over 20 peer reviewed publications on machine learning and financial modeling, has been cited in Bloomberg Markets and the Financial Times as an AI in fintech expert, and is a frequently invited speaker in Silicon Valley and on Wall Street. He has published R packages, served as a Google Summer of Code mentor and is the co-founder of the Thalesians Ltd.

Igor Halperin

is a Research Professor in Financial Engineering at NYU, and an AI Research associate at Fidelity Investments. He was previously an Executive Director of Quantitative Research at JPMorgan for nearly 15 years. Igor holds a Ph.D. in Theoretical Physics from Tel Aviv University (1994). Prior to joining the financial industry, he held postdoctoral positions in theoretical physics at the Technion and the University of British Columbia.

London, Chicago, New York,
December 2019

Paul Bilokon
Matthew Dixon
Igor Halperin

Contents