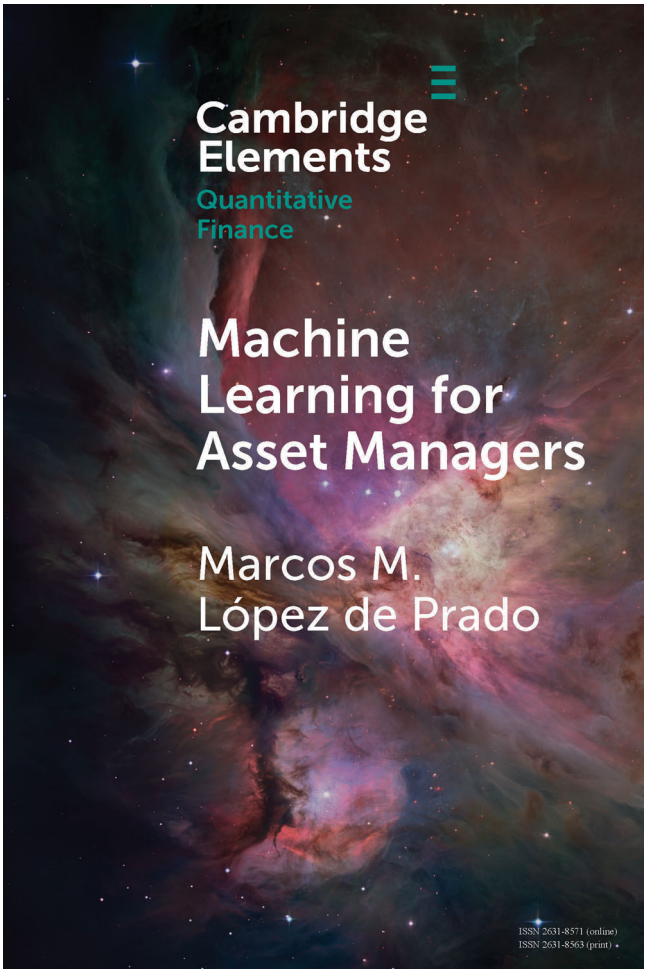
Quantitative Finance, 2020
Vol. 20, No. 11, 1761–1762, https://doi.org/10.1080/14697688.2020.1817534

Routledge
Taylor & Francis Group

Book review



Machine Learning for Asset Managers, by Marcos M. López de Prado, Cambridge University Press (2020). ISBN 9781108792899. Paperback.

Studying machine learning has become a big data problem. The field is characterized by an abundance of techniques, and for each technique, there are dozens of possible implementations. The book 'Machine Learning for Asset Managers' is a cost-efficient entry ticket into the field.

The author of the book, Marcos López de Prado, is among the most influential and highly reputed research-based professionals in the field of machine learning with financial applications. Given his extensive experience in the industry and academic research, lots of learning steps have preceded the writing of this book. The great feature of this book is that you only have to read 141 pages to get a grasp of the machine learning techniques that are useful in asset management.

This book is not a neutral overview book of a research field. It is opiniated, both in terms of selection of topics and writing style. The writing style is a cocktail of five important ingredients: (i) a clear narrative to understand the why and how of the method, (ii) the necessary equations to see the mathematics that are the backbone of the approach, (iii) the Python code snippets that implement the techniques, (iv) an insightful application with a discussion of the results and (v) exercises at the end of each chapter. This combination is excellent and makes the book useful to understand the theory and practice of machine learning in asset management. It motivates the readers to get their hands dirty and use machine learning

for reaching insights that lead to the discovery of financial theories.

The title of this book sells itself short by aiming only at asset managers. I recommend reading this book also to financial data scientists and graduate students in empirical finance. They will learn the importance of having a theory, the danger of p-hacking, and how equations translate into code and how the code leads to results.

The structure of the book is as follows. The book starts with an equation-free introduction about the role of machine learning in asset management. The use of machine learning techniques goes beyond accurate predictions. They are useful as a source of inspiration to discover the theories that explain the cause-effect mechanism of the good prediction performance achieved. As noted by the author, 'without a testable theory that explains your edge, the odds are that you have not an edge at all'.

Besides the introduction, there are seven chapters that help the reader to achieve an edge in terms of practical knowledge to use machine learning techniques in asset management.

Chapter 2 is a technical introduction to specific types of shrinkage approaches to covariance estimation. I recommend reading Chapter 3 first. It introduces the reader to different types of distance measures, that are useful as well to understand the concept of clustering and unsupervised learning approaches in Chapter 4. Chapters 5 and 6 focus on the prediction problem. Chapter 5 describes the process of financial labelling in which the modeller needs to define the relevant transformation of the price data for stock market prediction. Chapter 6 then discusses different approaches for selecting the features. Chapter 7 focuses on portfolio optimization. The bibliography at the end of the book is a great place to start if you are looking for important scientific articles in the field.

In Chapters 6 and 8 the caveats of statistical testing are presented. Chapter 6 presents the use of importance methods as an alternative to using p-values when selecting the relevant features. Chapter 8 shows how one can correct for multiple testing when evaluating a strategy. This is needed since 'when a researcher presents an overfit backtest as the outcome of a single trial, the simulated performance is inflated'. This then leads to false discoveries.

The book has the important merit to encourage the reader to be creative. It describes a careful selection of recipes to use machine learning to discover new theories useful for asset management. It does not provide a magic formula to implement an investment strategy. The discovery of a theory is a must for the asset manager seeking to develop a proprietary investment strategy. That is, in the words of Marcos López de Prado: 'you cannot bake someone else's cake and expect to retain it to yourself'.

Kris Boudt
Ghent University, Belgium

http://orcid.org/0000-0002-1000-5142

Kris Boudt is professor of finance and econometrics at Ghent University, Vrije Universiteit Brussel and Vrije Universiteit Amsterdam. He is an instructor at DataCamp and co-founder of Sentometrics.