

Cambridge Elements

Elements in Quantitative Finance
edited by
Riccardo Rebonato
EDHEC Business School

MACHINE LEARNING FOR ASSET MANAGERS

Marcos M. López de Prado
Cornell University



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-1-108-79289-9 — Machine Learning for Asset Managers
Marcos M. López de Prado
Frontmatter
[More Information](#)

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108792899

DOI: 10.1017/9781108883658

© True Positive Technologies, LP 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-79289-9 Paperback

ISSN 2631-8571 (online)

ISSN 2631-8563 (print)

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Machine Learning for Asset Managers

Elements In Quantitative Finance

DOI: 10.1017/9781108883658
First published online: April 2020

Marcos M. López de Prado
Cornell University

Author for correspondence: ml863@cornell.edu

Abstract: Successful investment strategies are specific implementations of general theories. An investment strategy that lacks a theoretical justification is likely to be false. Hence, an asset manager should concentrate her efforts on developing a theory rather than on backtesting potential trading rules. The purpose of this Element is to introduce machine learning (ML) tools that can help asset managers discover economic and financial theories. ML is not a black box, and it does not necessarily overfit. ML tools complement rather than replace the classical statistical methods. Some of ML's strengths include: (1) a focus on out-of-sample predictability instead of in-sample variance adjudication; (2) the use of computational methods to avoid relying on (potentially unrealistic) assumptions; (3) the ability to "learn" complex specifications, including nonlinear, hierarchical, and noncontinuous interaction effects in a high-dimensional space; and (4) the ability to disentangle the variable search from the specification search, in a manner that is robust to multicollinearity and other substitution effects.

Keywords: machine learning, unsupervised learning, supervised learning, clustering, classification, labeling, portfolio construction

JEL classifications: G0, G1, G2, G15, G24, E44

AMS classifications: 91G10, 91G60, 91G70, 62C, 60E

© True Positive Technologies, LP 2020
ISBNs: 9781108792899 (PB), 9781108883658 (OC)
ISSNs: 2631-8571 (online), 2631-8563 (print)

Contents

1	Introduction	1
2	Denoising and Detoning	24
3	Distance Metrics	38
4	Optimal Clustering	52
5	Financial Labels	65
6	Feature Importance Analysis	74
7	Portfolio Construction	92
8	Testing Set Overfitting	105
	Appendix A: Testing on Synthetic Data	125
	Appendix B: Proof of the “False Strategy” Theorem	128
	Bibliography	130
	References	136