SpringerBriefs in Computer Science

Series Editors

Stan Zdonik, Brown University, Providence, RI, USA

Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA

Xindong Wu, University of Vermont, Burlington, VT, USA

Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

Borko Furht, Florida Atlantic University, Boca Raton, FL, USA

V. S. Subrahmanian, University of Maryland, College Park, MD, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, VA, USA

Newton Lee, Institute for Education, Research, and Scholarships, Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at http://www.springer.com/series/10028

Daniil Ryabko

Universal Time-Series Forecasting with Mixture Predictors

Springer

Daniil Ryabko
Fishlife Research S.A.
Belize City, Belize

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-030-54303-7 ISBN 978-3-030-54304-4 (eBook)
https://doi.org/10.1007/978-3-030-54304-4

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is devoted to the problem of sequential probability forecasting, that is, predicting the probabilities of the next outcome of a growing sequence of observations given the past. This problem is considered in a very general setting that unifies commonly used probabilistic and non-probabilistic settings, trying to make as few as possible assumptions on the mechanism generating the observations. A common form that arises in various formulations of this problem is that of mixture predictors, which are formed as a combination of a finite or infinite set of other predictors attempting to combine their predictive powers. The main subject of this book are such mixture predictors, and the main results demonstrate the universality of this method in a very general probabilistic setting, but also show some of its limitations. While the problems considered are motivated by practical applications, involving, for example, financial, biological or behavioural data, this motivation is left implicit and all the results exposed are theoretical.

The book targets graduate students and researchers interested in the problem of sequential prediction, and, more generally, in theoretical analysis of problems in machine learning and nonparametric statistics, as well as mathematical and philosophical foundations of these fields.

The material in this volume is presented in a way that presumes familiarity with basic concepts of probability and statistics, up to and including probability distributions over spaces of infinite sequences. Familiarity with the literature on learning or stochastic processes is not required.

Antsiranana, Madagascar
2019

Daniil Ryabko

Contents

1 Introduction 1
- 1.1 General Motivation 2
- 1.2 Mixture Predictors 4
- 1.3 Loss: KL Divergence and Total Variation 5
- 1.4 Some of the Results 6
- 1.5 Mixture Predictors in the Literature on Sequence Prediction and Related Settings 7
- 1.6 Organization 9

2 Notation and Definitions 11
- 2.1 Loss 12
  - 2.1.1 KL Divergence 13
  - 2.1.2 Total Variation 14
- 2.2 Regret 15

3 Prediction in Total Variation: Characterizations 17
- 3.1 Optimality of Mixture Predictors 17
- 3.2 Topological and Algebraic Characterization of Predictability 19
- 3.3 Examples 22
  - 3.3.1 Example: Countable Classes of Measures 22
  - 3.3.2 Example: Bernoulli i.i.d. Processes 22

4 Prediction in KL Divergence 23
- 4.1 Realizable Case: Finite-Time Optimality of Mixture Predictors 23
  - 4.1.1 Finite-Time Upper-Bound and Asymptotic Optimality 23
  - 4.1.2 Lower Bound 24
  - 4.1.3 Examples 26
  - 4.1.4 Proof of Theorem 4.1 31
- 4.2 Conditions on $\mathcal{C}$ Sufficient for Vanishing Loss ($V_{\mathcal{C}} = 0$) 36
  - 4.2.1 Separability 37
  - 4.2.2 Conditions Based on the Local Behaviour of Measures 38
  - 4.2.3 Optimal Rates of Uniform Convergence 41

4.3 Non-Realizable Case and Suboptimality of Mixture Predictors 44
4.3.1 Suboptimality of Mixture Predictors 44
4.3.2 Some Sufficient Conditions on a Set $\mathcal{C}$ for the Existence of a Predictor with Vanishing Regret 47

5 Decision-Theoretic Interpretations 53
5.1 Players and Strategies 54
5.2 Minimax 54
5.3 Complete Class 55
5.4 Total Variation 56

6 Middle-Case: Combining Predictors Whose Loss Vanishes 57
6.1 Examples and an Impossibility Result for Stationary Processes 59
6.2 Proof of Theorem 6.1 61

7 Conditions Under Which One Measure Is a Predictor for Another 67
7.1 Measuring Performance of Prediction 69
7.2 Preservation of the Predictive Ability Under Summation with an Arbitrary Measure 70
7.3 Dominance with Decreasing Coefficients 71

8 Conclusion and Outlook 77
8.1 Generalizations: Infinite Alphabets and Different Losses 78
8.2 The General Sequence Prediction Questions 79
8.3 Beyond Sequential Prediction 81

References 83