Accepted Manuscript

Best subset binary prediction

Le-Yu Chen, Sokbae Lee

PII: S0304-4076(18)30077-0
DOI: https://doi.org/10.1016/j.jeconom.2018.05.001
Reference: ECONOM 4505

To appear in: *Journal of Econometrics*

Received date : 19 March 2017
Revised date : 30 April 2018
Accepted date : 6 May 2018

Please cite this article as: Chen L., Lee S., Best subset binary prediction. *Journal of Econometrics* (2018), https://doi.org/10.1016/j.jeconom.2018.05.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Best Subset Binary Prediction*

Le-Yu Chen†

Institute of Economics, Academia Sinica

Sokbae Lee‡

Department of Economics, Columbia University

Centre for Microdata Methods and Practice, Institute for Fiscal Studies

May 2018

Abstract

We consider a variable selection problem for the prediction of binary outcomes. We study the best subset selection procedure by which the covariates are chosen by maximizing Manski (1975, 1985)'s maximum score objective function subject to a constraint on the maximal number of selected variables. We show that this procedure can be equivalently reformulated as solving a mixed integer optimization problem, which enables computation of the exact or an approximate solution with a definite approximation error bound. In terms of theoretical results, we obtain non-asymptotic upper and lower risk bounds when the dimension of potential covariates is possibly much larger than the sample size. Our upper and lower risk bounds are minimax rate-optimal when the maximal number of selected variables is fixed and does not increase with the sample size. We illustrate usefulness of the best subset binary prediction approach via Monte Carlo simulations and an empirical application of the work-trip transportation mode choice.

*We are grateful to the co-editor, Jianqing Fan, an associate editor and three anonymous referees for constructive comments and suggestions. We also thank Stefan Hoderlein, Joel Horowitz, Shakeeb Khan, Toru Kitagawa, Arthur Lewbel, and participants in 2017 Asian Meeting of the Econometric Society and 2017 annual conference of the International Association for Applied Econometrics for helpful comments. This work was supported by the Ministry of Science and Technology, Taiwan (MOST106-2410-H-001-015-), Academia Sinica (Career Development Award research grant), the European Research Council (ERC-2014-CoG-646917-ROMIA), the UK Economic and Social Research Council (ES/P008909/1 via CeMMAP), and the British Academy (International Partnership and Mobility Scheme Grant, reference number PM140162).

†E-mail: lychen@econ.sinica.edu.tw

‡E-mail: sl3841@columbia.edu

Download English Version:

https://daneshyari.com/en/article/7357798

Download Persian Version:

https://daneshyari.com/article/7357798

Daneshyari.com