Constrained Classification and Policy Learning*

Toru Kitagawa†, Shosei Sakaguchi‡, and Aleksey Tetenov§

September 15, 2020

Abstract

Modern machine learning approaches to classification, including AdaBoost, support vector machines, and deep neural networks, utilize the surrogate-loss techniques to circumvent computational complexity in minimizing the empirical classification risk. These techniques are useful also for causal policy learning problems as estimation of individualized treatment rules can be cast as weighted classification. Consistency of these surrogate-loss approaches studied in Zhang (2004) and Bartlett et al. (2006) crucially relies on the assumption of *correct specification*, meaning that the specified class of policies contains a first-best. This assumption is, however, less credible when the class of policies is constrained by interpretability or fairness, leaving applicability of the surrogate-loss based algorithms unknown in such second-best scenarios. This paper analyzes consistency of the surrogate-loss procedures under a constrained set of policies without assuming correct specification. We show that the hinge losses (i.e., ℓ_1-support vector machines) are the only surrogate losses that preserve consistency in the second-best scenarios. We illustrate implications and uses of our theoretical results in monotone classification by proposing computational attractive procedures that are robust to misspecification.

Keywords: Surrogate loss, support vector machine, monotone classification, statistical treatment choice, personalized medicine

*The authors gratefully acknowledge financial support from ERC grants (number 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001).

†Department of Economics, University College London. Email: t.kitagawa@ucl.ac.uk.

‡Department of Economics, University College London. Email: s.sakaguchi@ucl.ac.uk.

§Geneva School of Economics and Management, University of Geneva. Email: aleksey.tetenov@unige.ch