



THE UNIVERSITY OF
CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

MAHTIYAR BONAKDARPOUR

Department of Statistics
The University of Chicago

Statistical Machine Learning Methods for Complex,
Heterogeneous Data

FRIDAY, February 1, 2019, at 2:30 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

We develop statistical machine learning methodology for three distinct tasks. Each method blends classical statistical approaches with machine learning methods to provide principled solutions to problems with complex, heterogeneous datasets. The first framework proposes two methods for high-dimensional shape-constrained regression and classification. These methods reshape pre-trained prediction rules to satisfy shape constraints like monotonicity and convexity. The second method provides a nonparametric approach to the econometric analysis of discrete choice. This method provides a scalable algorithm for estimating utility functions with random forests and combines this with random effects to properly model preference heterogeneity. The final method draws inspiration from early work in statistical machine translation to construct embeddings for variable-length objects like mathematical equations.