

MACHINE LEARNING for Causal Inference in Econometrics

Instructor: Anna Simoni

16 hours

The goal of this course is to give an overview of Machine Learning methods for Econometric models with a large number of covariates or nonlinearities. We will analyze methods for causal inference in economic settings where one wants to learn the effects of counterfactual policies (policy evaluation).

The machine learning methods that will be presented in the course are well suited to deal with datasets with many observations and/or many covariates. We will present state-of-the-art econometrics methods based on machine learning that allow to answer the causal questions of interest in economics. We will emphasize the difference with respect to statistical machine learning methods.

All along the course, the different models will be illustrated through applications and case studies. During the course, students will also learn how to implement in R the methods seen in class.

Syllabus:

1. Review concepts:
 - a. Statistical Learning, review of linear regression and least squares.
 - b. Challenges in High-dimension.
 - c. Review of the concept of endogeneity and of instrumental variables.
 - d. Introduction or review of R (depending on the level of knowledge of R by the students).
 - e. Review of linear Regression Models with many covariates: subset selection, shrinkage methods (Ridge regression, Lasso).
2. Inference: Post-Lasso and debiased Lasso.
3. High-Dimensional Instrumental Variables for causal inference, inference for Treatment Effect models.
 - a. Lasso-based methods
 - b. Tree-Based methods.
4. Case Studies based on research articles. Some examples include: impact of internet and social media on news, search advertising, policy evaluation.

Evaluation: by project in group.

References:

G. James, D. Witten, T. Hastie and R. Tibshirani, "An Introduction to Statistical Learning with applications in R", 2013, Springer.

Other references to journal articles will be provided during the course.