Topics in Econometrics: Introduction to the High Dimension
ECON 31703

Class: Tues. & Thurs., 8:00a-9:20a (Saieh 247)
Instructor: Stéphane Bonhomme
Office: Saieh 325
Office Hours: Tues. 3:30p-4:00p (Saieh 325)
email: sbonhomme@uchicago.edu

Teaching Assistant: Myungkou Shin
TA session: Mon. 5:30p - 6:20p (Cobb 107)
Office hours: Thurs 2:00p - 3:00p (graduate students lounge)
email: myungkoushin@uchicago.edu

Webpage: https://canvas.uchicago.edu/courses/

Goal of the course: The past two decades have seen a quick development of techniques to deal with high-dimensional models and data sets. Some of the main developments come from outside of economics, in fields such as statistics, machine learning and computer science. This course is an introduction to these techniques, with a focus on how to use them in economic applications. We emphasize theoretical properties of machine learning methods, as well as implementation.

Prerequisites: Students should be familiar with basic probability and statistics.

Background material: I will mainly draw material from journal articles. Two references that I will draw from are:

1. Hastie, Tibshirani and Friedman: Elements of Statistical Learning.
2. Hastie, Tibshirani and Wainwright: Statistical Learning with Sparsity.

Grading: There will be a few problem sets throughout the quarter. In the final week students will present a research project involving some machine learnings techniques, either focused on an application or on methodology.

Course Topics:

1. What is the high dimension? (w1 l1)
   (a) Applications
   (b) Main ideas
   (c) Organization of the course
2. Regression with a large number of regressors (w1 l2)
   (a) OLS with large p
   (b) Series regression
3. The Lasso and relatives (w2)
   (a) Subset selection
   (b) The Lasso
   (c) An example: estimating network links from panel data
   (d) Other penalization schemes: Ridge, grouped Lasso...
   (e) Matrix Lasso: nuclear norm penalization
4. The Lasso: basic theory and implementation (w3)
   (a) Computation of the Lasso
   (b) Convergence rates
   (c) Choice of λ
5. Inference on the Lasso (w4)
   (a) Post-model selection inference
   (b) Double Lasso
6. Partitioning methods (w5)
   (a) Methods based on covariates: kernel, nearest neighbor and kmeans

(b) Decision trees: CART

(c) Bagging and random forests

7. Some theory on trees (w6)

   (a) What is adaptivity?

   (b) Rate of convergence for kmeans-based regression

   (c) Rate of convergence for regression trees

   (d) Honest inference

8. Neural networks (w7)

   (a) Neural network models

   (b) Computation: backpropagation and stochastic gradient descent

   (c) Approximation property and issues in regularization and inference

9. Unsupervised learning (w8)

   (a) PCA/ factor methods/ matrix completion

   (b) Kmeans clustering and grouped fixed-effects

10. Latent variable modeling (w9)

    (a) Bayesian computation and inference in high dimensions

    (b) Variational Bayes

    (c) Topic models for text analysis, and model of link formation in networks