Economics 712: Non-Parametric and ML Approaches to Structural Estimation Problems

Course Description

This course explores themes at the intersection of machine learning and structural estimation problems in econometrics. Structural estimation problems arise when taking economic structures to data. The existence of certain economic primitives (demand functions, production functions, e.g., the underlying structure) are used to model the data generating process characterizing observables in the data, and the estimation problem is defined when structure is estimated from a sample of data drawn from the DGP. The usefulness of structural approaches is that they can used to perform potential policy experiments, most notably out of sample prediction.

Modern machine learning provides a set of empirical techniques that are often associated with solving a particular class of structural estimation problems – namely prediction problems. In this class we take the point of view of an emerging area of research that seeks to adapt and leverage the many recent advances in machine learning methods to design solutions to structural estimation problems more broadly. The content of the course will be based around the study the theory and practice of specific classes of ML/AI techniques and survey their application to economic questions. There are four main categories of interest with a progression that mirrors the typical phases of a structural estimation paper: "Stylized Fact – Reduced Form Description – Structural Estimation and Policy Counterfactual". These interestingly map to the "Descriptive-Predictive-Prescriptive" paradigm used to apply data to business problems in industry analytics.

0) Asymptotic experiments in Econometrics vs ML (1 week)
1) Unsupervised Learning and Feature Extraction (2 weeks)
    - Bayesian Inference, Bayesian Belief Networks, Principal Components, SVD, K Means, Applications to NLP.
2) Prediction Problems and Regularization (4 weeks)
    - Prediction decision problems, Bias-Bariance tradeoff, Penalized Least Squares Estimators (ridge/GAM/lasso), Classifiers (discriminant analysis, ANN's, SVM's).
3) Causal Inference (4 weeks)
    - Structural Equations and Instrumental Variables, the Partially Linear Model, Non-parametric Regression and Splines/Kernels, Panel Data and Group Fixed Effects, Double ML.
4) Selected Topics in Structural Estimation. (remaining time)
    - Discrete Choice Models, Demand Functions, Production Functions, Auctions and Games.

Useful Texts

In addition to selected journal articles, the following useful texts provide much of the core material for lectures in the first half of the course. Many are available freely online.

1. "Elements of Statistical Learning" (ESL) by Hastie, Tibshirani, and Friedman (free download)

   This text is clearly written and beautifully illustrated and gives both a theoretically foundational and computationally concrete basis for thinking about "ML technique" from a statistical point of view.

2. "Introduction to Machine Learning" (IML) by Ethem Alpaydın.

   This book is slightly less advance than ESL but with more of an ML/AI bent. It is a wonderful linear combination of two classic texts in ML/AL – Tom Mitchell's 1987 "Machine Learning" and Chris Bishop's "Pattern Recognition and Machine Learning", with a similar topical coverage as the former but emphasizing the probabilistic foundations of the latter. The Bayesian approach to inference is more readily embraced in the ML paradigm, which interestingly connects naturally to many economic/econometric questions that are framed as Bayesian problems. We will lean on this book to give a Bayesian point of view for the tools that are developed in ESL.

I will also draw from elements of the following texts for certain subjects. Most are freely available online but all would constitute useful additions to your library.

1. "Machine Learning" by Tom Mitchell.
2. "Analog Estimation Methods in Econometrics" (AEM) by Manski (free download)
3. "High Dimensional Statistics" (HDS) Lecture Notes by Rigollet and Hutter (free download)
4. "Economic Forecasting" (EF) by Elliot and Timmermann
5. "Econometric Analysis of Cross Section and Panel Data" by Wooldridge
6. "Elements of Causal Inference" (ECL) by Peters, Janzig, and Scholkopf (free download)
7. "Mostly Harmless Econometrics" by Angrist and Pischke
8. "Statistical Learning with Sparsity" by Hastie and Tibshirani (free download)

Tentative Topic Schedule

1. Thurs Jan 17	Intro to ML applications in Economics:
2. Tues Jan 22	Asymptotics in Econometrics vs ML
3. Thurs Jan 24	Induction and PAC Learning.
4. Tues Jan 29	Bayesian Inference, Bayesian Belief Networks, NLP
5. Thurs Jan 31	Feature Extraction: PCA and SVD analysis. Applications to Preference Forecasting: Netflix Challenge Application to NLP: LSA
6. Tues Feb 5	K Means Clustering.
7. Thurs Feb 7	EM Algorithm and LDA.
8. Tues Feb 12	Varieties of Structural Estimation problems; Regression, Best Linear Predictor, and Least Squares Estimators
9. Thurs Feb 14	Prediction Problems, Alternative Loss functions, and Quantile Regression
10. Tues Feb 19	Short vs Long Regression, Variable Selection, and Multiple Hypothesis Testing.
11. Thurs Feb 21	Ridge Regression, PC Regression, GAM, Project Pursuit, Model Selection
12. Tues Feb 26	The Lasso
13. Thurs Feb 28	Regression Trees, Boosting, Bagging, Random Forests Tree Based Decision Models
14. Tues Mar 5	Classifiers – Logistic Regression and Linear Discriminant Analysis
15. Thurs Mar 7	Perceptron, ANN's, and SVM's
16. Tues Mar 12	Regression vs Structure; Prediction vs Causality.
17. Thurs Mar 14	Single Equation Model as Causal Structure Non-Parametric Regression
18. Tues Mar 19	Splines and the "Kernel Trick"
19. Thurs Mar 21	Structural Equations and Instrumental Variables
20. Tues Mar 26	Potential Outcomes and LATE
21. Thurs Mar 28	Application of Lasso to Structural Models
22. Tues Apr 2	Non-Parametric IV, Weak IV, Deep IV, and the Cross Learner
23. Thurs Apr 4	Panel Data and Group Fixed Effects
24. Tues Apr 9	Two Step Estimators and Double ML
25. Thurs Apr 11	Non-Parametric Identification and Estimation of Demand Functions

26. Tues Apr 16	Non-Parametric Identification of Production Functions
27. Thurs Apr 18	Auctions and Games;
28. Tues Apr 23	Markov Decision Problems and Approximative Dynamic Programming
29. Thurs Apr 25	Reinforcement Learning
30. Tues April 30	TBD