Data Analysis for Economics

ä Modern Introduction

Jiaming Mao

Course Homepage: jiamingmao.github.io/data-analysis
Github Repository: github.com/jiamingmao/data-analysis

Course Description

This two-semester course offers a unified introduction to the principles and methods of statistical learning and causal inference – two areas essential to data analysis in economics. The first semester introduces learning theory and modern statistical and machine learning methods used for pattern recognition and predictive modeling. We focus in particular on how these methods can be applied to economic and business applications.

The second semester introduces the theory of causal inference and surveys frequently used econometric techniques for causal effect learning and program evaluation. Both the Rubin causal model and causal graphs will be introduced. Parciular emphasis will be placed on the recent literature on combining machine learning and high-dimensional statistical methods with traditional econometric techniques for causal inference. We will also discuss structural estimation and attempt offer a unified perspective on the use of reduced-form and structural econometric methods.

The goal of this course is to equip students with both a solid theoretical foundation, and the tools they need to conduct hands-on empirical research using state-of-the-art technology. The lecture materials are written to be both deep conceptually and easy to follow technically. Throughout the course, methods are demonstrated with applications to actual and simulated problems in various fields of applied economics, such as labor economics, industrial organization, finance, and marketing. Students will learn how to explore and analyze large high-dimensional datasets, choose appropriate methods for answering different types of questions, as well as gaining valuable computational skills.

The course spans the fields of econometrics, statistics, and computer science. Although the focus is on the analysis of economic data, the theories and the tools presented should be useful for a wide range of research areas in business and the social sciences.

Course Plan

Semester I

We begin with the theory of learning: why we are able to learn any patterns from observed data, what a model is, how to choose the best model, and what methods we can use to estimate them. We then survey a range of statistical and machine learning techniques data scientists use today to solve supervised and unsupervised learning problems and illustrate these methods with economic applications.

Our topics include:

- Foundations of Statistical Learning
- Regression
- Classification and Discrete Choice Models
- Model Selection and Regularization
- Decision Trees and Ensemble Methods
- Support Vector Machines
- Neural Networks

Semester II

We introduce the theory of causal inference: what is causality, when associations do *not* imply causation and when they *do*, and what strategies we can use to identify and estimate causal effects from observational data. We discuss both the Rubin causal model and the causal graphical model – two dominant paradigms of causal inference frameworks today. We then survey a number of methods used by econometricians today to estimate causal effects and evaluate the effects of policy programs. Finally, we introduce scientific models. These are models that describe the data-generating causal mechanisms. In the econometrics literature, scientific models based on economic theory are referred to as structural models. We discuss their use for prediction, causal inference, welfare analysis, and counterfactual simulation.

Our topics include:

- Foundations of Causal Inference
- Causal Effect Estimation Under Unconfoundedness
  - Regression and Matching Methods
- Causal Effect Estimation Under Unmeasured Confounding
  - Instrumental Variables
  - Panel Data Models
  - Quasi-Experimental Methods
- Structural Estimation: Dynamic Discrete Choice Models and Dynamic Games

Requirements

You are expected to have some familiarity with at least one programming/statistical computing language (R, Python, Matlab, Stata, etc.). We will provide ample data analysis problems for you to work through in this course. For your homework and final project, you can choose any language that you are familiar with.

Textbook

There are no required textbooks for this course. The following are recommended readings for the topics that we cover in this course.

Undergraduate

- Angrist, J. D. and J. Pischke. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Cameron, A. C. and P. K. Trivedi. (2010). *Microeconometrics using Stata* (Revised ed.). Stata Press.
- Hernán, M. A. and J. M. Robins (2019). *Causal Inference*. CRC Press.
- Morgan, S. L. and C. Winship. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (2^{nd} ed.). Cambridge University Press.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach* (7^{th} ed.). Cengage Learning.

Graduate

- Abu-Mostafa, Y. S., M. Magdon-Ismail, and H. Lin. (2012). *Learning from Data*. AMLBook.
- Cameron, A. C. and P. K. Trivedi. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Hastie, T., R. Tibshirani, and J. Friedmand. (2008). *The Elements of Statistical Learning* (2^{nd} ed.). Springer.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2^{nd} ed.). Cambridge University Press.
- Wooldridge, J. M. (2011). *Econometric Analysis of Cross Section and Panel Data* (2^{nd} ed.). The MIT Press.