Machine Learning and Econometrics, Spring 2023

Jui-Chung Yang, jcyang1225@ntu.edu.tw Location: Social Science 605

Time: Thursday 234

> *Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.* (Clive Humby, 2006)
>
> *Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.* (Dan Ariely, 2013)

This is a course about machine learning and econometrics for graduate students. Since Varian (2014), people have believed that **machine learning** uses data to predict some variables as a function of other variables, while **econometrics** uses statistical methods for prediction, inference, and causal modeling of economic relationships. However, nowadays, some believe machine learning is the *new generation* of **nonparametric** statistical and econometric methods branded as machine learning (Chernozhukov, 2016). Moreover, there are some exciting developments in applying machine learning algorithms in econometric methods, such as the estimation of conditional average treatment effects (e.g., Athey and Imbens, 2016), and as the first stage in two-stage estimation (e.g., Chernozhukov et al., 2018).

In this course, we, as a group of well-trained econometricians, will learn machine learning together. On the estimation/prediction side, we will talk about techniques such as Lasso, neural networks, support vector machines, and random forests. On the inference/causal modeling side, we will discuss some ideas about inference after using these machine learning techniques and inference methods based on / inspired by computer-based statistical algorithms, such as inference after model selection and large-scale hypothesis testing. More importantly, we will talk about the inference questions which still need to be answered, and you are more than welcome to join us and work on these unanswered questions as your research projects.

By definition, computer-based statistical algorithms will be nothing without computers. Students must be familiar with at least one statistical programming language, such as `Python` or `R`. You are also allowed to use any other programming languages you like as long as you are sure you can use them to do the tricks covered in this course.

Grading

One midterm presentation (April 20, **40%**). One final presentation (June 1, **50%**). Assignments (**10%**).

The final project can be an individual work or a teamwork with ≤ 2 team members in total. It can be a theoretical study, a Monde Carlo experiment or an empirical work.

Required Reading

1. Chan, Felix and Laszlo Matyas (2022), *Econometrics with Machine Learning*, Springer.
   `https://link.springer.com/book/10.1007/978-3-031-15149-1`

Supplemental Reading

1. Efron, Bradley and Trevor Hastie (2016), *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*, Cambridge.
   `https://web.stanford.edu/~hastie/CASI/index.html`

2. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2021), *An Introduction to Statistical Learning: With Applications in* `R`, Second Edition, Springer.
   `https://www.statlearning.com/`

3. Hastie, Trevor, Robert Tibshirani and Jerome Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer.
   `https://web.stanford.edu/~hastie/Papers/ESLII.pdf`

Schedule

- **Week 1**. Linear Econometric Models with Machine Learning (Part 1)
- **Week 2**. Linear Econometric Models with Machine Learning (Part 2)
- **Week 3**. Nonlinear Econometric Models with Machine Learning (Part 1)
- **Week 4**. Nonlinear Econometric Models with Machine Learning (Part 2)
- **Week 5**. The Use of Machine Learning in Treatment Effect Estimation (Part 1)
- **Week 6**. The Use of Machine Learning in Treatment Effect Estimation (Part 2)
- **Week 7**. Forecasting with Machine Learning Methods (Part 1)
- **Week 8**. **Midterm Presentation**
- **Week 9**. Forecasting with Machine Learning Methods (Part 2)
- **Week 10**. Causal Estimation of Treatment Effects From Observational Health Care Data Using Machine Learning Methods (Part 1)
- **Week 11**. Causal Estimation of Treatment Effects From Observational Health Care Data Using Machine Learning Methods (Part 2)
- **Week 12**. Econometrics of Networks with Machine Learning (Part 1)
- **Week 13**. Econometrics of Networks with Machine Learning (Part 2)
- **Week 14**. Machine Learning for Asset Pricing (Part 1)
- **Week 15**. Machine Learning for Asset Pricing (Part 2)
- **Week 16**. **Final Presentation**