ECON7225: Machine Learning and Econometrics

Jui-Chung Yang

Spring, 2024

E-mail: jcyang1225@ntu.edu.tw   Web: sites.google.com/site/juichungyang
Class Hours: Th 9:10-12:10   Office Hours: T/Th 12:20-13:10
Class Room: College of Social Sciences 605   Office: College of Social Sciences 715

Course Description

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 e§ects (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 that 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.

Required Materials

- Fan, J., Li, R., Zhang, C.-H., and Zou, H. (2020). *Statistical Foundations of Data Science*. CRC Press.
  fan.princeton.edu/sites/g/files/toruqf5476/files/documents/chapters1-3.pdf

- Chan, F., and Matyas, L. (2022), *Econometrics with Machine Learning*, Springer.
  link.springer.com/book/10.1007/978-3-031-15149-1

Grading

One midterm presentation (April 11, 40%). One final presentation (May 30, 60%).