

**ECON 504200 Machine learning and econometrics:
Concepts and procedures**

(機器學習與計量經濟：概念與程序)

**National Tsing Hua University
1st semester, 2022-2023**

Time: 3:30pm -6:20pm, every Tuesday (T7T8T9), **except November 29th**
Venue: TSMC 205

Make-up class(es): TBA

Examinations: TBA

Office hours: By appointment

Instructor: CY (Chor-yiu) SIN

(Office: CTM-804, Email: cysin@mx.nthu.edu.tw, Phone: 03-516-2134)

Course description: Since Varian (2014), people have been believing that machine learning uses data to predict some variable as a function of other variables, while econometrics use statistical methods for prediction, inference, and causal modeling of economic relationships. However, nowadays some tend to believe that machine learning is the new generation nonparametric statistical and econometric methods, branded as machine learning (Chernozhukov, 2016, Yang, 2021). While **academically rigorous theorems are much appreciated**, they are not the focus of this course. Instead, this course starts with **basic concepts** of machine learning and econometrics; and ends with **procedural computer codes**. Students are required to be familiar with at least one software such as **R, Python or C++**.

This course discusses the basic concepts with examples in economics and management. Hands-on exercises with the software “R” will also be provided.

Course topics:

1. Revision on (a) prediction with linear regression (b) classification with logit/probit model
2. Linear model selection and regularization: (a) Ridge regression; (b) Elastic net; (c) Lasso; (d) Best subset regression plus CV/information criterion (IC)
3. The essential tools in computer age statistical inference: (a) Monte-Carlo simulation; (b) Bootstrapping; (c) Cross-validation (CV)
4. Generalized linear model: with an introduction to support vector machine (SVM)

5. Tree-based methods and boosting: with an introduction to gradient boosting
6. Basic topics in deep learning: (a) Single-/Multiple- layer neural network; (b) Convolutional neural network (CNN); (c) Recurrent neural network (RNN) and Long short-term memory network (LSTM).
7. Other topics in deep learning: (a) Unsupervised and supervised learning; (b) Generative adversarial network (GAN); (c) Reinforcement learning.
8. Re-sampling approach to (a) Interval estimation; (b) P-value

Required textbooks:

- (1) Efron, B., Hastie, T. (2016). Computer age statistical inference: algorithms, evidence and data science. Cambridge, U.S.: Cambridge University Press.
- (2) Hansen, B.E. (2021). Econometrics.
(<https://www.ssc.wisc.edu/bhansen/econometrics/>)
- (3) James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). An introduction to statistical learning with applications in R, 2nd edition. New York, U.S.: Springer Science+Business Media (<https://www.statlearning.com/>)

Assessment: Attendance and participation in discussions 40%, mid-term examination 30%, final examination: 30%.

Question: What is wrong with the following computer package:

<https://www.bilibili.com/video/BV13o4y1m7Ka/>

Attitude taking this course:

Be humble, careful and simple. Needless to say, be logical.

- End -