

**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  
1<sup>st</sup> semester, 2023-2024**

**Time:** 3:30pm -6:20pm, every Tuesday (T7T8T9)

**Venue:** TSMC 205

**Examinations:**

**Mid-term presentation:** TBA

**Final presentation:** TBA

**Office hours:** By appointment

**Instructor:** CY (Chor-yiu) SIN

(Office: CTM-804, Email: [cysin@mx.nthu.edu.tw](mailto: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 et al. 2017). 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 **Python, R or C++**.

**This course discusses the basic concepts with examples in economics and management. Hands-on exercises with the software “Python”.**

**Course topics:**

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

4. Tree-based methods and boosting: with an introduction to gradient boosting
5. Basic topic in deep learning: Feedforward neural network
6. Basic topic in deep learning: Recurrent neural network (RNN) and Long short-term memory network (LSTM)
7. Intermediate topic in deep learning: Convolutional neural network (CNN).
8. Intermediate topic in deep learning: “You only look once” (YOLO)
9. Other topics in deep learning: (a) Generative adversarial network (GAN); (b) Reinforcement learning (RL)
10. Generalized linear model: with an introduction to support vector machine (SVM)
11. Re-sampling approach to (a) Interval estimation; (b) P-value

**Textbooks and papers:**

- (1) Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., 2017. Double/debiased/Neyman machine learning of treatment effects. American Economic Review: Papers and Proceedings 107, 261-265.
- (2) Efron, B., Hastie, T., 2016. Computer age statistical inference: algorithms, evidence and data science. Cambridge, U.S.: Cambridge University Press.
- (3) Hansen, B.E., 2022. Econometrics. New Jersey, U.S.: Princeton University Press
- (4) James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., 2023. An introduction to statistical learning with applications in Python. New York, U.S.: Springer Science+Business Media (<https://www.statlearning.com/>)
- (5) Varian, H.R., 2014. Big data: new tricks for econometrics. The Journal of Economic Perspectives 28, 3-27.

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

**Attitude taken in this course:**

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