

Machine Learning II

Course Description

The set of methods which are known today as Machine Learning has a long history in economic applications. With the rapid development in computing technology and the availability of large economic data sets, these methods have recently been more relevant than ever. This is an introductory level course in Machine Learning which has the objective of covering the most common and recent techniques developed in the area that are relevant for economists.

Organization

- 21 hours on Spring term

Evaluation

- Around 5 applied homeworks

Textbooks

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York: Springer.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. New York: Springer.

Plan

1. Introduction to Statistical Learning (Session 1)
2. Resampling methods (Session 1)
3. Linear model selection and regularization (Sessions 2 and 3)
4. Tree-Based Methods (Session 3)
5. Unsupervised Learning (Session 4)
6. Deep Learning (Session 5)
7. Selective Inference in Econometrics (Session 7)

Detailed Plan

1. Statistical Learning (Chapter 2 of ISL)

- What is statistical learning?
- Assessing model accuracy
- The bias-variance trade-off

Lab: Bias-variance decomposition using simulations

2. Resampling Methods (Chapter 5)

- Cross-validation
- Bootstrap

Lab: Implementing K-fold CV and bootstrap in R

3. Linear Model Selection and Regularization (Chapter 6)

- Subset selection
- Shrinkage methods
- Dimension reduction

Lab: Lasso and Ridge regression using glmnet

4. Tree-Based Methods (Chapter 8)

- Decision trees
- Bagging, Random Forests
- Boosting

Lab: Forecasting with ensemble trees

5. Unsupervised Learning (Chapter 10)

- K-means and hierarchical clustering
- Principal Components Analysis

Lab: PCA and clustering with macroeconomic indicators

6. Deep Learning (Chapter 13)

- Introduction to neural networks
- Fitting neural nets

Lab: MLP for macroeconomic forecasting

7. Selective Inference in Econometrics (Additional Chapter)

- Selective inference for Lasso and stepwise regression
- Selective inference in clustering
- Application to grouped panel models (Bonhomme and Manresa 2015)

Lab: Post-selection inference in R

Selected Readings

Literature Reviews

- [3] Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.
- [7] Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210), 1243089.
- [11] Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of economic surveys*, 37(1), 76–111.
- [13] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- [17] Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of economic perspectives*, 28(2), 3–28.

Clustering

- [1] Ando, T., & Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1), 163–191.
- [6] Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3), 1147–1184.
- [14] Sarafidis, V., & Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*, 77(2), 274–296.
- [16] Su, L., Shi, Z., & Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6), 2215–2264.

Forecasting Applications

- [4] Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2), 304–317.
- [5] Bai, J., & Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4), 607–629.
- [8] Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5), 920–964.
- [9] Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2), 339–354.
- [12] Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1), 98–119.
- [15] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460), 1167–1179.

Regularization

- [2] Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519), 1182–1198.
- [10] Lam, C., & Souza, P. C. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 38(3), 693–710.