ECON4061 Machine Learning in Economics

Module Outline

Instructor: Marit Hinnosaar

Autumn 2023

Lectures: Mondays 9:00–11:00, SCGB A42
Computer classes: Tuesdays 10:00–11:00 in October & November, George Green Library A13
Office hours sign-up:
https://calendly.com/marit-hinnosaar/marit-hinnosaar-office-hours-2023-fall

Module aims and objectives

The module will introduce students to Machine Learning and the analysis of large datasets ("big data") in economics. Large datasets have become common in economic analysis, especially because many production and consumption activities leave digital footprints. Machine learning algorithms provide useful tools for analyzing such data. The module teaches coding in Python and provides basic knowledge and practice in implementing Machine Learning algorithms, including Large Language Models that are used in the recent literature. The module gives an introduction to those techniques, applies these to real-world datasets, and presents examples from the current economic research of the use of the techniques.

Assessment

The assessment is via an in-person closed book 3-hour exam. The exam gives 100% of the mark in the module. The exam will take place during the Autumn examination period: Monday 15 January 2024 - Saturday 27 January 2024.

List of topics

1. Module overview, introduction to coding in Python, data cleaning and preparation, visualizing data

2. Natural language processing

   - Regular expression, dictionaries, sentiment analysis, bag-of-words, tokenization, similarity
   - Unsupervised learning, topic models, Latent Dirichlet Allocation
   - References of ML methods:
     - David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, March 2003. https://dl.acm.org/doi/10.5555/944919.944937
     - David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 262–272. EMNLP '11, 2011. https://dl.acm.org/doi/10.5555/2145432.2145462
   - Examples of applications of dictionary based methods, including dictionary based sentiment analysis:
     - Matthew Gentzkow and Jesse M. Shapiro. What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1):35–71, 2010. https://doi.org/10.3982/ECTA7195
     - Shane Greenstein and Feng Zhu. Is Wikipedia Biased? *American Economic Review*, 102(3):343–348, May 2012. http://dx.doi.org/10.1257/aer.102.3.343

- Scott R. Baker, Nicholas Bloom, and Steven J. Davis. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, November 2016. https://doi.org/10.1093/qje/qjw024
- Paul C. Tetlock. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3):1139–1168, 2007. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2007.01232.x

- Examples of applications of similarity:

- Gerard Hoberg and Gordon Phillips. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016. https://doi.org/10.1086/688176
- Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, 3(3):303–320, 2021. https://doi.org/10.1257/aeri.20190499

- Examples of applications of LDA:

- Oriana Bandiera, Andrea Prat, Stephen Hansen, and Raffaella Sadun. CEO Behavior and Firm Performance. *Journal of Political Economy*, 128(4):1325–1369, 2020. https://doi.org/10.1086/705331

3. Large Language Models

- Preliminaries: neural networks, word embeddings, word2vec, recurrent neural networks, Long Short-Term Memory (LSTM) networks, sequence to sequence (seq2seq) models, encoder-decoder neural networks

- Attention, transformers, BERT, RoBERTa

- References of ML methods:

- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to Sequence Learning with Neural Networks, December 2014. arXiv:1409.3215 [cs]
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, 2017. arXiv:1706.03762 [cs]

- Examples of applications of word embeddings:

- Elliott Ash, Daniel L. Chen, and Arianna Ornaghi. Gender Attitudes in the Judiciary: Evidence from U.S. Circuit Courts. *American Economic Journal: Applied Economics*, forthcoming, 2023

- Examples of applications of LLMs:

- Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. The Voice of Monetary Policy. *American Economic Review*, 113(2):548–584, 2023. https://www.aeaweb.org/articles?id=10.1257/aer.20220129