ECON4061 Machine Learning in Economics

Module Outline

Instructor: Marit Hinnosaar

Autumn 2024

Lectures: 10 2-hour sessions, at Monica Partridge C17; Wed 9-11am
Computer classes: 8 1-hour sessions in October & November, at Law and Social Sciences A25, Wed 1-2pm
Office hour sign-up:
`https://outlook.office365.com/owa/calendar/SchoolofEconomicsOfficeHours@UniofNottm.onmicrosoft.com/bookings/`

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 consists of group coursework and in-person closed book written 2-hour exam. Coursework consists of in-person in-class group presentation and group essay. The exam gives 75% of the mark in the module, group coursework presentation gives 10%, and group coursework essay gives 15%. The exam will take place during the Autumn examination period: Monday 13th January 2025 - Saturday 25th January 2025. Group courseworks are presented in the last three weeks of the module during lecture times. The deadline for the group coursework essay is Friday, December 13, 2024 at 15:00.

List of topics

1. Module overview, introduction to coding in Python, data cleaning and preparation
2. Natural language processing
   - Regular expressions, dictionaries, sentiment analysis, bag-of-words, tokenization, similarity
   - 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

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
     - Marit Hinnosaar and Toomas Hinnosaar. Influencer Cartels. SSRN Scholarly Paper 3786617, Rochester, NY, May 2024

Tentative schedule of lectures

1. Introduction and overview of the module
2. Natural language processing I: dictionaries, dictionary-based sentiment analysis, applications in economics
3. Natural language processing II: regular expressions, term frequency - inverse document frequency, cosine similarity, applications in economics
4. Neural networks, word embeddings, word2vec, applications in economics
5. Recurrent Neural Networks and Long Short-Term Memory Networks
6. Sequence to sequence (seq2seq) models, attention
7. Transformers, Large Language Models, applications in economics
8. Group coursework presentations
9. Group coursework presentations
10. Group coursework presentations

Tentative schedule of computer classes

1. Introduction to Python (assigining values, logical operators and conditionals, lists and indexing, loops, defining functions), Jupyter Notebooks, Pandas, importing and exporting data, taking a look at the data
2. Data preparation and aggregation in Pandas, summary statistics, data visualization
3. Regular expressions, dictionaries, sentiment analysis
4. Cosine similarity
5. Word2vec
6. Large Language Models
7. Transformer-based models to work with text and images
8. Review