MACHINE LEARNING FOR ECONOMISTS

Fall 2020

Instructor: Lilia Maliar
E-mail: lmaliar@gc.cuny.edu
https://lmaliar.ws.gc.cuny.edu/

Lecture time: TBD
Location: TBD
Office Hours: TBD
Course material will be posted on Blackboard.

Course Summary:

Recent developments in artificial intelligence and constantly growing computational power provide economists with unprecedented capacities for the data analysis. This course provides a broad overview of numerical methods at the intersection of mathematics, statistics and computer science that constitute the workhorse of the modern data analytics. In particular, the course provides an introduction to machine learning, deep learning, reinforcement learning, parallel computing and big data methods, as well as data manipulation, visualization, presentation and interpretation techniques. The studied applications are not limited to conventional econometric regressions models but contain some prominent examples from computer science, such as recognition of handwritten numbers. The course also introduces students to programming in Python with the emphasis on economic applications.

Learning goals and outcomes

1) Demonstrate good understanding of the existing machine-learning techniques.
2) Develop good programming skills.
3) Demonstrate the ability to use the existing machine-learning techniques for analyzing the given economic problem.
4) Demonstrate the ability to find new research questions in light of existing theories.

Prerequisites:

This course assumes having completed the first-year PhD-level microeconomics, macroeconomics and econometrics sequences. Good knowledge of Python programming languages is recommended but not required, as long as you are ready to learn it as you go. You may find helpful to do the following online course.

Online edx Course: Introduction to Computer Science using Python.

Assessment:

The grade breakdown is as follows:
(1) Mandatory attendance (5%).
(2) Problem sets (30%). These problem sets provide opportunities to bring lecture material into practice, and they will emphasize coding in Python. This assignment relates to learning goals 1), 2), 3) and 4).
(3) Midterm exam (25%). This relates to learning goals 1), 2) and 3).
(4) Final exam (40%). This relates to learning goals 1), 2) and 3).

Policies:

1) There will be no make-up exams and late problem sets receive a grade of zero; no exceptions.
2) Laptops are highly discouraged: use them only if absolutely necessary and cell phones are strictly prohibited.

Programming Language:

We will use Python. Most of the modern data science applications are written in Python, supplemented with data-science platforms such as Google TensorFlow and Scikit. An additional advantage of Python is that it is an open source software. All the problem sets will use examples and data related to economics.

Course Outline:

1) Introduction to Python.
2) Linear regression: cost function and gradient descent
3) Logistic regression and classification
4) Overfitting and regularization
5) Neural networks
6) Testing hypothesis: bias vs. variance
7) Support vector machines
8) Unsupervised learning
9) Dimensionality reduction
10) Anomaly detection
11) Recommender systems
12) Large-scale machine learning
13) Reinforcement learning: computer games versus dynamic economics models

14) Data science with Python, TensorFlow and Scikit

Note: The content is subject to changes depending on the student's progress and feedback

Online Textbooks, Readings and Lecture Notes

There are no required textbooks for this course; all relevant information will be from course materials provided in the class (lecture slides, problem sets, etc.). Below are few references which may be helpful for the course:

Banerjee, A., V. Chandola and V et al. (2009), Anomaly detection: a survey, ACM Computing Surveys.

Boyd, S. and Vandenberghe , L, (2004). Convex optimization. Cambridge University Press.

Chakraborty, C. and A. Joseph, (2017). Machine Learning at Central Banks. Bank of England. Staff working paper 674. Python code.

Coleman, C., S. Lyon, L. Maliar and S. Maliar, (2017). Matlab, Python, Julia: What to choose in economics? CEPR working paper DP-13210.

Efron, B. and T. Hastie, (2017). Computer Age Statistical Inference. Algorithm, Evidence and Data Science. Cambridge University Press.

Goodfellow, I., Y. Bengio and A. Courville, (2016). Deep Learning. The MIT Press.

Hastie, T., R. Tibshirani and J. Friedman, (2009). The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer.

Hastie, T., R. Tibshirani and W. Wainwright, (2016). Statistical Learning with Sparsity. The Lasso and Generalizations. CRC Press.

James, G., D. Witten, T. Hastie and R. Tibshirani, (2013). An Introduction to Statistical Learning with Applications in R, Springer. Python code.

Karim, M., A. Menshawi and G. Zaccone (2017). Deep Learning with TensorFlow: Explore neural networks with Python. Packt.

Leskovec, J., A. Rajaraman and J. Ullman (2014). Mining of Massive Datasets. Cambridge University Press.

Murphy, K., (2012). Machine Learning: A Probabilistic Perspective. The MIT press.

Introduction to Machine Learning. Wikipedia Guide.

Maliar, L. and S. Maliar, (2014). Numerical Methods for Large Scale Dynamic Economic Models. In: Schmedders, K. and K. Judd (Eds.), Handbook of Computational Economics, Volume 3, Chapter 7, Amsterdam: Elsevier Science.

Maliar, L., Maliar, S. and P. Winant, (2019). Will artificial intelligence replace computational economists any time soon? CEPR working paper DP 14024.

Menshawy, A., R. Karim, and G. Zaccone, (2017). Deep Learning with TensorFlow. O'Reilly.

Nielsen, M. (2019). Neural Networks and Deep Learning. Free online book.

Nilsson, N., (2005). Introduction to Machine Learning. Notes.

Shalev-Shwartz, S. and S. Ben-David (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.

Silver, D. (2015). Introduction to Reinforcement Learning. Lecture notes.

Sutton, R. and A. Barto, (2018). Reinforcement Learning: An Introduction. Second edition, in progress. The MIT Press.

Pedregosa et al. (2011), Scikit-learn: Machine Learning in Python, JMLR.

Powell, V., (2011). Approximate Dynamic Programming: Solving the Curses of Dimensionality. 2nd Edition. Wiley Series in Probability and Statistics.

VanderPlas, J., (2016). Python Data Science Handbook, O'Relly Press. Python code.

Free Online Courses:

- Mathematics for Machine Learning, MIT course. License.
- Machine Learning with Python: from Linear Models to Deep Learning, MIT course.
- Introduction to Machine Learning, MIT course. License.
- Introduction to Computational Thinking and Data Science, MIT course. License.
- Machine Learning, Stanford course.
- Machine Learning, University of Maryland course.
- Machine Learning for Graduate, Washington University course.
- Introduction to Machine Learning for Coders, USF course.
- Deep Learning for Coders, USF course.
- Reinforcement Learning, UCL course.

Other Free Online Resources:

Such resources include data repository and online codes.

- UCI machine learning, Data repository.
- A curated list of practical financial machine learning tools and applications in Python.
- Data Science Association, Resources.
- Scikit-Learn: Machine Learning in Python, Resources.