MSF 6424
Machine Learning for Finance

Spring 2022

Prof. Murray Z. Frank

Class Time: Tuesday, Thursday 3:45 PM to 5:25 PM
Class Room: CSOM 1-132
Teaching Assistant: Ramin Hassan, hassa417@umn.edu

Machine learning methods are widely used in finance. Particularly common applications include credit evaluation, measuring market sentiment, interpreting financial news, and asset pricing. This class introduces several of the fundamental methods. The goal is that each student should develop a basic feel for these methods and their application in finance. This requires both learning important ideas, and significant time invested at the computer generating code that actually works.

Machine learning is closely related to statistics and to econometrics. In statistics the goal is often to establish the connections between variables. Traditionally it was done with smaller data sets, using methods that could be proved to work well in certain settings. In econometrics the focus is often on establishing causal connections and estimating parameters that have an economic interpretation. In machine learning the focus is often on out-of-sample prediction. This is usually done with larger data sets, using methods that experience has shown to work well on a range of applications. The differences are partly historical, and they reflect how each discipline developed. The dividing lines are not sharp, and there is now substantial overlap in methods and in motivations.

A real project has several steps: 1) data collection, 2) data management, 3) exploratory data analysis, 4) estimation, learning and predicting, 5) communicating results. All of these play an important role in real world data analysis.

Grades

- 3 assignments totaling 30% (due dates: March 27, April 3, April 17)
- Quiz 10 % (April 12)
- Class participation 10%
- Exam 50% (April 28)

Software: The homework assignments provide you with practice in the use of Python. Your code is expected to follow the best practices instructions as in:
http://www.danielmsullivan.com/pages/tutorial_workflow_3bestpractice.html

Required: Geron, Aurelien, 2019, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, 2nd Edition, O'Reilly Media. We also make heavy use of the Scikit-Learn User Guide:

https://scikit-learn.org/stable/user_guide.html and Quantecon Data Science: https://datascience.quantecon.org/.

Most lectures have two assigned readings that generally overlap and complement each other. Usually the lectures will draw material from both of them, as well as going beyond the readings in some respects.

Setting the stage

March 15. **Introduction**
Read: Geron ch 1
Read: Israel, Kelly, Moskowitz, 2020, Can machines learn finance? Journal of Investment Management 18, 2, 23-36

March 17. **An end-to-end project**
Read: Geron ch 2
Read: https://scikit-learn.org/stable/modules/preprocessing.html

Fundamental methods

March 22. **Classification**
Read: Geron ch 3
Read: https://datascience.quantecon.org/applications/classification.html

March 24. **Naive Bayes and k-nearest neighbors**
Read: https://scikit-learn.org/stable/modules/naive_bayes.html
Read: https://scikit-learn.org/stable/modules/neighbors.html

March 29. **Linear regression**
Read: Geron ch 4
Read:https://datascience.quantecon.org/applications/regression.html

March 31. **Logistic regression**
Read: Geron ch 4 (continued)
Read: https://en.wikipedia.org/wiki/Logistic_regression

April 5. **Support vector machine**
Read: Geron ch 5
Read: https://scikit-learn.org/stable/modules/svm.html

April 7. **Trees**
Read: Geron ch 6

Read: https://scikit-learn.org/stable/modules/tree.html

April 12. **Quiz** and **Ensemble methods**
Read: Geron ch 7
Read: https://scikit-learn.org/stable/modules/ensemble.html

Applied issues

April 14. **Visualizing data**
Read: https://datascience.quantecon.org/applications/visualization_rules.html
Read: https://seaborn.pydata.org/tutorial/function_overview.html

April 19. **Missing data**
Read: Little and Rubin, 2020, Introduction, Ch 1 in "Statistical Analysis with Missing Data" (3rd Edition), Wiley
Read: https://scikit-learn.org/stable/modules/impute.html

April 21. **Interpreting your model**
Read: Molnar, C. 2021, Interpretability, Ch 3 and Model-agnostic methods, Ch 6 in "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable", https://christophm.github.io/interpretable-ml-book/

Advanced topic

April 26. **Neural networks**
Read: Geron ch 10
Read: Gu, Kelly, and Xiu, 2020, Empirical Asset Pricing via Machine Learning, Review of Financial Studies, 33, 5, 2223–2273.

April 28. Final Exam
Please note that the final exam will be held during the regular class period.

Python resources: When looking for help with Python and it's libraries, the first place to look is not Google. The first place to look is under the help buttons in Spyder or in Jupyter - depending on which you are using. There are also many useful Python reference resources freely available, notably: https://chrisalbon.com/ and https://datascience.quantecon.org/.

Optional Readings The literature on machine learning and on related issues in finance is huge. Obviously we cannot do everything. These may be mention in class, but they are not assigned. I suggest browsing, to see what you find interesting.

- Antweiler, W., and Frank, M. Z. 2004, Is all that talk just noise? The information content of internet stock message boards, Journal of Finance, 59, 3, 1259-1294.

- Erel, I., Stern, L.H., Tan, C. and Weisbach, M.S., 2021. Selecting directors using machine learning. Review of Financial Studies, 34, 7, 3226-3264.
- Frank, M. Z., and A. Sanati, 2018, How Does the Stock Market Absorb Shocks? Journal of Financial Economics, 129, 1, 136-153.
- Hall, P. 2019, Toward Responsible Machine Learning, https://github.com/jphall663/hc_ml
- Healy, K., Data Visualization: A Practical Introduction, https://socviz.co/
- Lipton, Z.C., 2018, The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16, 3, 31-57.
- Lopez de Prado, M., 2018, The 10 Reasons Most Machine Learning Funds Fail, Journal of Portfolio Management, 44, 6, 120-133.
- Murphy,K., 2022, Probabilistic Machine Learning: An introduction, MIT Press. A rigorous reference and some code. It has been described as "What every ML PhD student should know." https://github.com/probml/pml-book/releases/latest/download/book1.pdf
- Another graphics library: https://plotly.com/python/
- Ribeiro, M. T., S. Singh, and C. Guestrin, 2016, Model-agnostic interpretability of machine learning. https://arxiv.org/abs/1606.05386
- Yang, X., et al. 2020, Qlib: An AI-oriented Quantitative Investment Platform. https://arxiv.org/abs/2009.11189

Academic Policies

The Carlson School defines academic misconduct as any act by a student that misrepresents the student's own academic work or that compromises the academic work of another. Scholastic misconduct includes (but is not limited to) cheating on assignments or examinations, plagiarizing, i.e., misrepresenting as one's own work any work done by another, submitting the same paper, or substantially similar papers, to meet the requirement of more than one course without the approval and consent of the instructors concerned, or sabotaging another's work. Within this general definition, however, instructors determine what constitutes academic misconduct in the courses they teach. Students found guilty of academic misconduct face penalties ranging from lowering of the course grade or awarding a grade of F or N for the entire course, to suspension from the University.

http://www1.umn.edu/regents/policies/academic/Student_Conduct_Code.html

Accommodations for Students with Disabilities

The University of Minnesota is committed to providing all students equal access to learning opportunities. Disability Services is the campus office that works with students who have disabilities to provide and/or arrange reasonable accommodations. Students registered with Disability Services, who have a letter requesting accommodations, are encouraged to contact the instructor early in the semester. Students who have, or think they may have, a disability (e.g. psychiatric, attention, learning, vision, hearing, physical, or systemic), are invited to contact Disability Services for a confidential discussion at 612-626-1333 (V/TTY) or at ds@umn.edu. Additional information is available at the DS website http://ds.umn.edu