MATH 279: Data Science and Machine Learning for Finance (4)

Winter 2026
Lecture: MS 5117
Wednesday 3:00 PM-3:50 PM; Friday 2:00 PM-3:50 PM

Instructor: Mihai Cucuringu
Email: mihai@math.ucla.edu
Office: MS 7620B
https://www.math.ucla.edu/~mihai/fin.htm

• Course Description

Lecture, three hours. This is a project-based course covering topics pertaining to applications of data science and machine learning to the field of statistical finance, that will combine both theoretical and practical approaches. The goal of the course is, on one hand, to understand at a high level the mathematical foundations behind some of the state-of-the-art algorithms for a wide range of unsupervised and supervised learning tasks arising when dealing with financial data, including linear and nonlinear dimensionality reduction, network analysis, clustering, and ranking. The selected literature will expose students to a spectrum of regression techniques - spanning both linear and nonlinear models - used in financial modeling, with a focus on handling high-dimensional, noisy data, while assessing their robustness, and balancing predictive power with interpretability.

A broad range of financial topics will be considered, including portfolio construction, statistical arbitrage, lead-lag detection, market microstructure and limit order book analysis, price impact and optimal execution, realized volatility and covariance matrix forecasting, random matrix theory and financial correlations, fundamental and statistical factor models, change-point detection, and backtesting of systematic strategies. Students will be exposed to numerous practical examples drawn from finance, involving data sets that range from mid/low frequency (eg minutely and daily prices) to high-frequency limit order book data.

• Learning Outcome/Course Objectives/Core Competencies

After successful completion of Math 279, students should be able to

- Understand the mathematical foundations of core machine learning techniques relevant to finance, including linear and nonlinear dimensionality reduction, clustering, network analysis, and ranking.
- Apply statistical learning methods such as regression models (linear, nonlinear, penalized, and high-dimensional) to financial data, and assess their robustness, predictive accuracy, and interpretability.
- Analyze complex dependence structures in high-dimensional time series using tools from random matrix theory, factor models, and covariance/volatility forecasting.
- Model financial systems with advanced data-driven techniques, including limit order book modeling, order flow imbalance, lead–lag detection, price impact, and optimal execution.
- Implement practical machine learning pipelines in finance, handling diverse datasets that span daily, intraday, and high-frequency trading data.

• Evaluate empirical asset pricing models, statistical arbitrage strategies, and portfolio construction methods with rigorous backtesting.
• Integrate theoretical insights with computational experiments to produce interpretable, reproducible results in statistical finance applications.
• Communicate complex technical findings clearly, both in written reports and in oral presentations, through project-based work that mimics real-world financial research.

Textbook

-**Machine Learning in Finance: From Theory to Practice** by Matthew F. Dixon, Igor Halperin, and Paul Bilokon, Springer (2020). ISBN 978-3-030-41067-4
eBook freely available at
https://link.springer.com/book/10.1007/978-3-030-41068-1

-**The Elements of Quantitative Investing** by Giuseppe A. Paleologo, Wiley (2025) ISBN: 978-1-394-26546-6

-**Financial Machine Learning** by Bryan T. Kelly and Dacheng Xiu, SSRN 2023, https://ssrn.com/abstract=4501707

Prerequisites

Linear algebra (at the level of MATH 115A), elementary probability and statistics, programming experience (at the level of PIC 16A). Optional but helpful: previous exposure to graphs and networks (at the level of MATH 168: Introduction to Networks), data science (MATH 118), machine learning (MATH 156), and time series analysis. While prior familiarity with basic finance concepts would be helpful, no prior knowledge of finance will be assumed.

• General Course Outline/Schedule of Lectures

Topics
• Introduction & Roadmap
• Statistical Machine Learning
• Introduction to Quantitative Finance • Modern Portfolio Theory and Asset Pricing • Evaluation metrics: risk-adjusted performance measures, statistical significance of alphas, robustness checks
• Measures of correlation and dependence: Pearson, Spearman, Kendall, Maximal Correlation, Distance Correlation, Mutual Information, Maximal Information Coefficient (MIC)
• Singular Value Decomposition and Principal Component Analysis

• PCA in high dimensions, Random Matrix Theory and financial applications
• Statistical arbitrage; development of equity market-neutral strategies
• Nonlinear Dimensionality Reduction • cMDS, ISOMAP, LLE, Laplacian Eigenmaps, Diffusion Maps • Autoencoders, Autoencoder Asset Pricing Models
• Graph partitioning • Clustering point clouds and graphs: k-means, spectral clustering • Cheeger's Inequality, spectral bipartition • Analysis and modeling of order flow in limit order markets
• Clustering signed graphs, regularization techniques • Applications to statistical arbitrage • Signed clustering for portfolio construction and hierarchical Markowitz
• Clustering Directed Graphs • Ranking from Pairwise Comparisons • Group Synchronization and Applications
• Lead-Lag detection in financial networks
• Stochastic Block Models: Spectral & SDP Relaxations
• Linear models for prediction in Quantitative Finance; Fama-MacBeth regressions; Applications to factor models • Regularization and sparse regression for financial prediction; Applications in factor selection and portfolio construction
• Introduction to Limit Order Books • Market Microstructure
• Price-Impact and Cross-Impact of Order Flow Imbalance, Conditional order flow imbalance
• HAR models for forecasting Realized Volatility • Forecasting Realized Volatility with machine learning and graph-based methods
• Econometric and machine learning approaches to forecasting realized covariance matrices: factor-based methods, shrinkage estimators, graph-structured regularization techniques for high-dimensional settings.
• Optimization methods in finance; Markowitz mean-variance framework • Hierarchical Risk Parity (HRP) and clustering-based optimization

• Empirical Asset Pricing via Machine Learning
• Topics in Quantitative Finance: Virtue of Complexity, Probabilistic Forecasting, Generative Modeling in Finance

- **Grades:**
  Project: 95%
  Scribe Notes/Participation: 5%

  Students will contribute to the course by preparing scribe notes summarizing selected lectures and by actively participating in class discussions and fostering a collaborative learning environment.

  Scribe notes will be prepared by small groups of one to two students, who will sign up to cover a specific lecture. The scribe notes should concisely summarize the main ideas from the lecture slides, as well as relevant points arising from in-class discussions and questions. In addition, each set of scribe notes should conclude with a short section outlining several potential research directions or open questions related to the lecture material, encouraging further exploration beyond the core content.

Course Project:
The main component of the course is a group or individual project applying machine learning and data science methods to a financial dataset. Students are expected to identify a topic relevant to the themes covered in class (e.g., portfolio construction, volatility forecasting, statistical arbitrage, lead-lag analysis, market microstructure or limit order book modeling) and to develop a reproducible research report combining theoretical and/or empirical analysis.

Project milestones:

Milestone	Description	Due Date	Weight
Proposal	1-2 page summary outlining the chosen problem, dataset, brief literature review, and methods	Week 4	10%
Midterm Progress Report	Literature review, preliminary analysis, initial results, challenges encountered	Week 8	15%
Presentation	15-minute presentation on the project	Week 10	20%
Final Report	10-15 page written report with full analysis, figures, and code appendix	Finals Week	50%

Deliverables must include: a written report, supporting code/scripts, figures or tables summarizing key findings, and a short presentation.

The final written report is due by the end of Week 11 (or at the latest, before the grade submission deadline).