MASTER of SCIENCE PROGRAM
Mathematics in Finance

NEW YORK UNIVERSITY
Courant Institute of Mathematical Sciences

Trends in Financial Data Science
MATH-GA.2047-001
Tuesdays, 7:10-9 p.m.
IN-PERSON
CIWW room 1302
Updated August 13, 2021

Instructors:
Ivailo Dimov
Email: iid201@nyu.edu
Office hours: TBA

Petter N. Kolm
Office: CIWW room 520
Email: petter.kolm@nyu.edu
Office hours: TBA

Grader:
Helen Chien
yq204@nyu.edu
Office hours: TBA

Special dates:
September 7, first lecture
December 14, last lecture
December 21, final exam

Prerequisites: The following four courses, or equivalent: (1) Data Science and Data-Driven Modeling, (2) Financial Securities and Markets, (3) Machine Learning & Computational Statistics, and (4) Risk and Portfolio Management. It is important you have experience with the Python stack.

Course requirements: Students are expected to attend lectures and actively participate. The main work for students outside of class includes homework and one group project. In the group project, students will read, understand and present a recent machine learning paper relevant to quantitative finance. Groups will be randomly assigned by the instructors to encourage teamwork. Throughout the semester there will be several short in-class quizzes (one per month) assessing the understanding of the lectures and assigned readings. There is a final exam at the end of the semester. Each student will be graded based on their performance in (1) homework assignments and quizzes (30%); (2) group project (30%); and final exam (40%).

Late homework policy: Unless a student obtains approval from the instructors to submit a deliverable late, there is a penalty according to this schedule:

- Baseline penalty for late homework: 20%
- More than a week: 40%
- More than two weeks: 60%
- More than a month, or if other students' corrected homework has been returned: 100%

Communication & Zoom: Announcements, homework and other course related material will be posted on NYU Brightspace. The NYU Brightspace site has a discussion forum for the course. Please post any questions related to lectures, homework and group projects there in the appropriate thread.

This is an in-person course. Classroom captures of all lectures will be available in the Zoom section of the course site.

Academic integrity: Students must make a serious commitment to academic integrity. The consequences of cheating are serious for the cheater and for NYU as a whole. Students are required to immediately report cheating incidents they observe to the instructors.

A student caught cheating on an assignment may have their grade for the class reduced by one letter at the first offense and their grade reduced to an F, or dismissal from NYU, for repeated offenses. During exams and quizzes, students may not communicate in any way, nor use any materials or technology not explicitly permitted. No cellphone or other electronic devices may be used during the exam; they should be stored away. Students may not look at each other's test and/or screen during the exam. Cheating on an exam or quiz will immediately result in an F for the class.

Students are not permitted to share course materials outside of class without written permission of the instructors. Students are not allowed to record (photography, audio and/or video) any lectures or sessions without written permission of the instructors.

All disciplinary actions will be taken in accordance with the policies of the NYU Graduate School of Arts and Science: GSAS Statement on Academic Integrity (nyu.edu)

Course description: This is a full semester course covering recent and relevant topics in alternative data, machine learning and data science relevant to financial modeling and quantitative finance. This is an advanced course that is suitable for students who have taken the more basic graduate machine learning and finance courses Data Science and Data-Driven Modeling, and Machine Learning & Computational Statistics, Financial Securities and Markets, and Risk and Portfolio Management.

Tentative course outline:

Bayesian statistics and decision theory

- Likelihood principle, prior and posterior distributions
- Utility and loss
- Conjugate, improper and noninformative priors
- Bayesian decision theory, sampling and MCMC
- Kalman, particle filters, and the Black-Litterman-Bayes model

- Online forecasting

SVD, PCA & ICA methods
- SSA and non-parametric de-seasonalization
- LSA and bag-of-words text-similarity
- The multilinear PCA (MPCA) algorithm
- SVD with missing data: EM PCA algorithm (BPCA), gradient descent algorithm
- ICA and information

Google PageRank and Perron-Frobenius
- Discrete Markov processes and chains
- Perron and Perron-Frobenius theorems
- The PageRank model

Functional Bayesian methods and Gaussian processes
- Gaussian processes regression (GPR), hyperparameter learning
- Kernel trick and its Bayesian interpretation
- Reproducing-kernel Hilbert spaces and the representer theorem
- Mercer theorem and functional PCA
- Unsupervised learning with GPR, missing data problems
- GPR Kalman filters: The Morokoff-Kalman model and its generalizations
- Building risk models from asynchronous data

Hidden Markov models
- The EM algorithm
- Dynamic portfolio optimization as a sequence estimation problem
- Jump models as generalizations of HMMs
- Regime switching models (risk-on / risk-off) using jump models

Neural networks in finance
- Recursive feature maps and deep neural nets (DNNs)
- Universality and the Radon transform
- Working with DNNs: Backpropagation, vanishing gradients, regularization, adversarial noise
- Classes of DNNs and their usage in finance: CNN, RNN, LSTM
- Autoencoders (PCA and MPCA generalization)
- DNNs for mid-price return prediction in limit order books

Natural language processing & sentiment modeling for financial applications
- Topic modeling, LDA and extensions
- Transformer models (BERT, finBERT)
- Sentiment modeling & statistical arbitrage

Random projection methods
- Randomized linear algebra: Johnson-Lindenstrauss lemma, random projection PCA/SVD and linear regression (random FFTs, random single-hidden-layer neural nets)
- Compressed sensing and L1-L2 random projection methods
- Randomized neural nets and Gaussian processes
- Neural tangents

Articles, textbooks and other readings will be announced in class and on NYU Brightspace.