

Advanced Machine Learning Applications for Finance

Course content

This course surveys applications of machine learning techniques to various types of financial datasets. This course starts with financial data structure and features, then introduces deep learning and advanced supervised learning techniques. We will examine several machine learning applications in pricing, hedging, and portfolio management. Advanced methods for clustering and classification such as support vector machine and unsupervised learning will be introduced. Reinforcement learning and its connection with optimal control will be discussed. Text data will be introduced and analyzed using text mining techniques. Machine learning techniques will be applied to asset allocation. Strategy back-testing and strategy risk will also be discussed.

Outline

Part I: Statistical Machine Learning

- Classification: logistic regression, discriminant analysis, naive Bayes, KNN, support vector machine (maximal margin classification, support vector classifier, SVM)
- Clustering: K-means clustering, hierarchical clustering
- Case studies: Credit card fraud data, client segmentation

Reference: An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer
<https://www.statlearning.com/>

Part II: Deep neural network

- Theoretical foundation: universal approximation theorem, neural network, training
- Applications: deep hedging, deep portfolio optimization, deep calibration

Reference: Hands-on Machine Learning with Scikit_Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, by Aurelien Geron, O'Reilly.
Machine Learning in Finance, a course at ETH Zurich by Josef Teichmann
https://people.math.ethz.ch/~jteichma/index.php?content=teach_mlf2019

Part III: Natural Language Processing

- Text pre-processing: stop words removal, tokenization, stemming
- Features creation: vectorization, bag of word, tf-idf, word2vec
- Application: Tweets sentiment analysis using logistic regression, naive Bayes, and neural network

Reference: Speech and Language Processing, by Dan Jurafsky and James H. Martin
<https://web.stanford.edu/~jurafsky/slp3/>
Deeplearning.ai <https://www.deeplearning.ai/>

Part IV: Industry lecturer (TBD)

Lectures: Wed 6:30pm – 9:15pm EST HAR 302

For students who want to attend in-class learning, please sign up every week on Questiontools course website

Instructor: Hao Xing

Email: haoxing@bu.edu

Office hours: Friday 9-10:30pm

Teaching Assistants: Jingyan Wang jingyanw@bu.edu

Office hours: Monday 7-9pm EST

Liyan Wang liyanwong@bu.edu

Office hours: Tuesday 9-11am EST

Grading: The grade for the course will be determined as follows:

Homework (written and coding): 50%

Midterm project (individual): 25%

Final project (group): 25%

QUESTROM COMMUNITY STATEMENT OF NORMS

In addition to the expectations outlined in the [COVID-19 Health Commitments & Expectations for Boston University Students](#), **we expect all members of our Questrom community to adhere to the following classroom norms:**

- Use the designated paths and doors to enter and exit the classrooms and move around the Hariri building.
- Use the wipes provided in classrooms to clean desktops and other spaces before and after use.
- Wear an appropriate personal protection equipment (PPE) face covering when in the Hariri building, including in the Hariri classrooms and offices. Students not wearing a face covering will be asked to leave and are expected to comply with the request.
- Be willing to display your “green screen” compliance app upon request (e.g., in class or for a meeting).
- Eat only in the designated locations in the Rafik B. Hariri Building. Eating is NOT permitted in the classrooms. Eating is only permitted in Starbuck/Breadwinners and in student lounges. Only drinking with a straw is permitted in classrooms.
- Update your LfA location and class status (in-person or 100% remote) on the StudentLink as changes occur throughout the semester.
- Attend class remotely if your behavior outside of the classroom might put others in the community at risk.

Adhering to these norms will allow us to continue to enjoy a residential learning experience

COURSE RECORDING POLICY STATEMENT

All class sessions will be recorded for the benefit of registered students who are unable to attend live sessions (either in person or remotely) due to time zone differences, illness or other special circumstances. Recorded sessions will be made available to registered students ONLY via their password-protected QuestromTools account. Students may not share such sessions with anyone not registered in the course and may certainly not repost them in a public platform. Students have the right to opt-out of being part of the class recording. Please contact your instructor or teaching assistant to discuss options for attending the course in such cases.