Anderson School of Management
UCLA

MGMT MFE 431 Prof. Lochstoer
Data Analytics and Machine Learning Spring 2025

Syllabus (***tentative***)

Course Information

This is a course in data science oriented toward decision-making and predictive analytics. Topics include predictive and prescriptive models, panel regressions, text analysis, model validation and selection, models for discrete outcomes, and common machine learning methods. We will use Python for coding and analysis. Examples and homework will focus on finance applications, including return and earnings prediction, default prediction and lending markets, portfolio choice, and trading models.

Prerequisites

Statistical Concepts: All topics covered in Introduction to Econometrics and Empirical Methods of Finance.

Computing

All computing will be done with Python. Python is a free programming environment and language for statistics and data manipulation. Python is used world-wide in both academia and the private sector. We will use only a small subset of the capabilities of Python in this course and this will not require extensive training. While all material needed for the use of Python will be developed in class, some prior knowledge of Python is useful.

Mathematical Preparation

All students are assumed familiar with exponential and logarithmic functions and summation notation.

I will also assume a familiarity with vectors, matrices as well as matrix multiplication.

Homework and Tests

The final (44%), six homework (24%), a group project and presentation (24%), and class participation (8%). The homework should be done in the pre-assigned groups. The homework group does not have to be the same as the project group, but there can be no more than four students per project group. You have to form the project groups yourself. There cannot be collaboration across groups.

There will be six homework assignments. Homework will typically involve a combination of some derivation and Python computing. Each homework will count for 4% of the grade. You may ask the teaching assistant clarifying questions regarding the homework.

Lectures, HW Due, Group Project and Final Exam Dates

There are ten lectures on 4/2, 4/7, 4/14, 4/21, 4/28, 5/5, 5/12, 5/19, including project presentations and discussions on 6/5 and 6/6.

Homework should be submitted electronically via the BruinLearn web site.

Student groups will come up with their own project, using data analytics and machine learning techniques to develop either a trading strategy, an earnings forecasting model, or a risk management tool. Group presentations will be in class Thursday June 5th and Friday June 6th. The slides and code will have to be submitted by end of June 6.

The final exam will be TBD. The exam is in-person with a cheat-sheet allowed.

Late Homework Policy: Since homework submission is available on the BruinLearn web site, no late homeworks will be accepted. If you do not submit the homework before the submission deadline, the drop-box on the BriunLearn website will close and not allow for any further submission.

Please do not submit unedited output. CUT and PASTE the relevant portions of the output into your homework document.

Exam attendance:

> You must take the exam with your section unless you secure *prior* written permission from me. It is expected that students will complete all exams for the course. **I will only entertain requests for rescheduling the exam in cases of extreme medical or personal emergency**. This ***does not*** include job interviews, weddings, vacations or class trips. Any request must be made *in advance*.
>
> Having one or more other exams on the same day is NOT a legitimate reason for exam rescheduling. Similarly, the fact that a case assignment or course project is due on the exam date is NOT legitimate either. Plan ahead to manage your time.

Exam Format:

> The exam will be in-person. You are allowed a letter-sized cheat-sheet where you can write on both pages, as well as a calculator.

Requests for Regrade:

> All requests for regrading of exams must be made in writing and state clearly the basis of the request. All requests for regrades must be made within 14 days of the date of the exam.
>
> Clerical errors will be corrected at no risk to the student. All other regrading requests will result in regrading of the entire exam. Downward as well as upward revisions of a grade are possible.

<u>Students with Learning Disabilities</u>

If you have a learning disability and require special provisions for taking exams, it is your responsibility to

get in touch with the UCLA OSD

http://www.osd.ucla.edu/

and have them contact me directly to make exam arrangements for you. I will email the exam to them and they will administer it.

Class Attendance

Class attendance is vital to obtain a thorough understanding of the material and has a high weight in the class participation grade.

Office Hours

In person:

I will stay after each class for anyone that has questions.

By E-mail:

lars.lochstoer@anderson.ucla.edu

NOTE: read the syllabus and assignments before asking procedural or logistical questions. Questions about the course *content* are always welcome. Questions because you are not willing to read the syllabus and assignments are not.

Teaching assistant:

Kibeom Lee, email: kibeom.lee.phd@anderson.ucla.edu
TA session: TBD

Class Materials

All course materials will be available on the BruinLearn web page the day before the class. Students are responsible for printing their own copies of course materials.

All course materials including data, slides, and Python code should not be distributed outside this class.

The required text for the class, "Introduction to Statistical Learning," can be found at:
http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf

My class notes are relatively self-contained, but the textbook gives more details for the underlying statistics. For a more in-depth treatment of the statistical topics, I recommend "Elements of Statistical Learning," which can be found at:
https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf

A useful reference for deep learning is:
http://www.deeplearningbook.org/

Course Outline

Topic 1: Introductory Concepts, model building and visualization (Week 1)

- Examples of data analytics in finance
- Data analytics and Python, introduction to useful tools for data management
- Visualization as a model specification and communication tool
- Graphics in Python
- Forecasting signals based on data visualization

Topic 2: Big data and panel regressions (Week 1 and week 2)

- Introduction to panel regressions.
- MR vs SR: the value added of a signal
- Clustering of standard errors.
- Time and firm fixed effects.
- Fama-MacBeth panel regressions revisited: from signal to trading strategy

Topic 3: Logistic Regressions, Credit Data, and Sample Selection (Week 2 and week 3)

- Models for discrete outcomes: Logistic regressions
- Default prediction and interest rates using credit card database

Topic 4: Machine Learning I: Model selection (Week 3 and Week 4)

- Bayesian shrinkage
- Lasso and ridge regressions
- Sparsity principle
- Cross-validation

Topic 5: Machine learning II: Nonlinear methods Part 1 (Week 5 and week 6)

- High-dimensional data; model selection and regularization revisited
- Decision trees: bagging (Random Forest) and boosting (XGBoost)

Topics 6 and 7: Unstructured data and textual analysis (Week 6 and week 7)

- Text analysis in Python
- Mapping unstructured data to numeric signals
- Similarity metrics, investor inattention
- Text as big data: 21st century reading of financial reports (web scraping, the EDGAR database)

Topic 8 and 9: Machine learning III: Nonlinear methods Part 2 (Week 8 and week 9)

- Classification: k-means clustering, support vector machines
- Deep learning and neural networks
- Student presentations

Week 9 and 10: Student presentations, final topics.