Bus241a: Machine Learning and Data Analysis for Business and Finance: Spring 2023 (preliminary)

Contact Details

Blake LeBaron
Email: blebaron@brandeis.edu
Office location: Sachar 204
Telephone: (781) 736-2258

TA : TBA

Communication
The best way to contact me will be over email or zoom. Phone is not reliable since I will probably check campus voice mail very sporadically. I will also use Latte announcements frequently for the class.

Continuity
If there are any problems with campus, we will be able to shift all aspects of the class to full zoom remote. If I have internet or power issues at my location I will try to get a short message out to the class that I have technical issues, and I will try to reach a location with power.

Meeting Times/Locations

Classes
Mon/Wed: 4:05pm-5:25pm (in person)

Office hours:
TBA

Student Office Hours
TBA

Accommodations

Brandeis seeks to create a learning environment that is welcoming and inclusive of all students, and I want to support you in your learning. Live auto transcription is available for all meetings or classes hosted on Zoom and you can turn it on or off to support your learning. Please check for Zoom updates to take advantage of this new feature. To learn more, visit the Zoom Live Transcription webpage. For questions, contact help@brandeis.edu
If you think you may require disability accommodations, you will need to work with Student Accessibility Support (SAS) (781-736-3470, access@brandeis.edu). You can find helpful student FAQs and other resources on the SAS website, including guidance on how to know whether you might be eligible for support from SAS. If you already have an accommodation letter from SAS, please provide me with a copy as soon as you can so that I can ensure effective implementation of accommodations for this class. In order to coordinate exam accommodations, ideally you should provide the accommodation letter at least 48 hours before an exam.

Brandeis UNIVERSITY

Course Description

Course Prerequisite(s):

1. Econ213a/Econ184a (equivalent to one semester of econometrics)
2. Bus215f: Python for Business and Finance, or solid knowledge of Python
    a. Knowledge of Numpy, SciPy, and Pandas is a plus
    b. Fin285a also covers enough Python to prepare you for this class.

Learning Goals:

This course is a general topics course on machine learning tools, and their implementation through Python, and the Python packages, Scikit Learn, Keras, TensorFlow. The course is oriented heavily to applications in business and finance, giving students the tools needed to survive in the modern data analytics space. It is not intended as a deep theoretical approach to machine learning. Students will finish the class with a basic understanding of how to execute predictive analytic algorithms, as well as rigorously test their performance. The course is statistical in nature. It will draw on tools from our basic econometrics class, Bus213a. Finally, the course assumes a good working knowledge of the Python programming language at the start. Bus215 meets this requirement. Online courses in Python may be acceptable to meet this requirement.

Goal list:

1. Basic data processing and handling with Python/Pandas
2. Machine learning tools available in Scikit Learn
3. Testing and evaluating forecasts/predictions
4. Presenting/describing results
5. Neural network/deep learning tools using Keras/TensorFlow

Credit Hours:

Success in this four credit course is based on the expectation that students will spend a minimum of 9 hours of study time per week in preparation for class (readings,papers, discussion sections, preparation for exams, etc.)

Textbook(required):

Geron, *Hands-on Machine Learning with Scikit-learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly, 2019. (This book is available online for free through the Brandeis Library.)

Textbooks(optional):

1. James, Witten, Hastie, Tibshirani, *An Introduction to Machine Learning*, Springer 2013. (This book gives a slightly more detailed mathematical approach than the core textbook, and it is extremely useful for more math savvy students. It can be a useful reference to have. This book is available in pdf format for free on the web.)
2. Hastie, Tibshirani, Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer 2017. (This book presents the technical mathematical core of much of machine learning. Its math level is well beyond the scope of this class, but more math prepared students may find it useful. It is also available for free in pdf.)
3. McKinney, Python for Data Analysis: Data Wrangling with Pandas, Numpy, and IPython, O'Reilly, 2017, second edition. (This book is a useful Python overview to have on your shelf if you feel you need it. Covers Python and all the critical packages for data analytics. A third edition is coming out soon.)

Lecture notes:

Extensive lecture notes in pdf format are provided on Latte.

Software (all software is open source):

1. Python 3.7/3.8 and the entire Anaconda suite of tools.
2. The Keras package.
3. Tensorflow II

Data:

Many data sets will be provided for analysis and experiments. These will be available through the class website. Also, additional data can be retrieved through many public websites. This includes the **Machine Learning Repository**, and **Kaggle**

Course Requirements

Attendance

Class attendance (in person) is required. However, classes will be recorded, so missing a class can be made up by watching the off line video.

Assignments

The class will involve weekly problem sets which should be submitted individually. Students may use others as resources for working on problem sets. Understanding problem sets is crucial to doing well on exams. All problem sets are submitted using Latte.

Exams/Quizzes

There will be two midterm exams given using Latte at roughly halfway, and two thirds through the class, and one final exam during the final exam period. Exams will be open book/open notes/open internet. However, no human resources can be used to assist. This means others in the class, or anyone else on the planet cannot not be used to help. All exams will be given in person, and in class, or in a room determined for the final exam.

Participation

Students are expected to be active in class whether in person on online. I encourage discussion and participation through questions and active thinking. Software will be provided, and students can try experiments on their own machines during class. Questions and critiques about class code is strongly encouraged. I would like you to keep your zoom views on for remote class meeting

Course Plan

These sections all line up exactly with the sections and resources on the Latte page.

(Gn) refers to the textbook, Geron, chapter n.)

1. Landscape of machine learning problems (G1)
2. Python basics
3. Modeling preliminaries
    a. Software
    b. Terminology
    c. Notation
    d. Model performance
4. Overfitting/Generalization
    a. Overfitting
    b. Regularization
    c. Cross-validation
5. Classification (G3)
    a. Logistic regression
    b. K-nearest neighbors
    c. Performance measures
    d. Classification thresholds
    e. ROC curves
    f. Multi-class problems
    g. Error analysis
6. Regression and regularization (G4)
    a. Regression
        i. Performance measures
        ii. Linear regression
        iii. Polynomial regression
        iv. K-nearest neighbors
    b. Regularization
        i. Overfitting
        ii. Regularized linear models (Ridge and Lasso)

        iii. Bias/variance tradeoff
        iv. Logistic regression with regularization
7. Training (G4)
    a. Matrix review
    b. Gradient descent
    c. Batching
    d. Learning curves
    e. Early stopping
8. Data preprocessing (G2)
    a. Cleaning
    b. Rescaling
    c. Grid searching
    d. Cross-validation
9. Support vector machines (G5)
    a. Linear SVM classification
    b. Nonlinear SVM
    c. Math details on kernels
10. Basic nonlinear regression tools
    a. Kernel Ridge regression
    b. SVM regression
11. Decision trees (G6)
    a. Classification
    b. Training
    c. Regression
12. Ensemble learning (G7)
    a. Voting classifiers
    b. Bagging and pasting
    c. Random forests
    d. Boosting
    e. Stacking
13. Unsupervised learning (G9)
    a. K-means clustering
        i. Data preprocessing
        ii. Semi-supervised learning
    b. DBSCAN
14. Neural networks/deep learning (G10)
    a. Network structure
    b. Keras
    c. Image detector
    d. Neural network hyperparameters
    e. Advanced training for deep networks
15. Dynamic networks and time series
    a. Recurrent networks
    b. LSTM/GRU units
    c. Convolutional layers

Brandeis Calendar Dates

The Brandeis calendar can be a tricky thing to follow. For convenience I will present here all the critical calendar dates for Fall 2022.

Jan 17: First day of classes

No class: Feb 20-24, April 6-13

May 2: Thursday schedule

May 2: Last day of classes

May 3-4: Study days

May 5-16: Final exams

May 21: Commencement

Evaluation and Grading

All grading progress will be transparent and can be monitored through Latte. All exams will be during class times.

Class Element	Grade Percentage
Problem sets	10%
Midterm exam I	20%
Midterm exam II	20%
Final exam	30%
Group projects	20%

Important Policies and Resources

Academic Integrity

Every member of the University community is expected to maintain the highest standards of academic integrity. A student shall not submit work that is falsified or is not the result of the student's own effort. Infringement of academic integrity by a student subjects that student to serious penalties, which may include failure on the assignment, failure in the course, suspension from the University or other sanctions. Please consult Brandeis University Rights and Responsibilities for all policies and procedures related to academic integrity. Students may be required to submit work via TurnItIn.com or similar software to verify originality. A student who is in doubt regarding standards of academic integrity as they apply to a specific course or assignment should consult the faculty member responsible for that course or

assignment before submitting the work. Allegations of alleged academic dishonesty will be forwarded to the Department of Student Rights and Community Standards. Citation and research assistance can be found at Brandeis Library Guides - Citing Sources.

Snow Days

Brandeis can be closed for weather on some winter days. Fortunately, we have technology that will enable classes to continue remotely on these days. If Brandeis has closed, then stay home, and check your email. Instructions on how we will proceed will go out on email. We should continue as usual. Classes will be recorded over zoom. This is subject to my having electricity, so if I do not show up, then it means I'm having technical problems, and in this case we will need to cancel the class. (I should make it up with an asynchronous recorded lecture.)

Classroom Health and Safety

- Register for the Brandeis Emergency Notification System. Students who receive an emergency notification while attending class should notify their instructor immediately. In the case of a life-threatening emergency, call 911. As a precaution, review this active shooter information sheet.
- Brandeis provides this shuttle service for traveling across campus or to downtown Waltham, Cambridge and Boston.
- On the Brandeis campus, all students, faculty, staff and guests are required to observe the university's policies on physical distancing and mask-wearing to support the health and safety of all classroom participants. Face coverings must be worn by all students and instructors in classes with in-person meetings. Review up to date COVID-19 Response | Brandeis University regularly. These can change as campus status level changes.

Course Materials/Books/Apps/Equipment

If you are having difficulty purchasing course materials, please make an appointment with your Student Financial Services or Academic Services advisor to discuss possible funding options, including vouchers for purchases made at the Brandeis Bookstore.

LATTE

LATTE is the Brandeis learning management system. Login using your UNET ID and password. For LATTE help, contact Library@brandeis.edu.

Library

The Brandeis Library collections and staff offer resources and services to support Brandeis students, faculty and staff. Librarians and Specialists from Research & Instructional Services, Public Services, Archives & Special Collections, Sound & Image Media Studios, MakerLab, AutomationLab, and Digital Scholarship Lab are available to help you through consultations and workshops.

Privacy

To protect your privacy in any case where this course involves online student work outside of Brandeis password-protected spaces, you may choose to use a pseudonym/alias. You must share the pseudonym/ alias with me and any teaching assistants as needed. Alternatively, with prior consultation, you may submit such work directly to me.

Student Support

Brandeis University is committed to supporting all our students so they can thrive. If a student, faculty, or staff member wants to learn more about support resources, the Support at Brandeis webpage offers a

comprehensive list that includes these staff colleagues you can consult, along with other support resources:

- The Care Team
- Academic Services (undergraduate)
- Graduate Student Affairs
- Directors of Graduate Studies in each department, School of Arts & Sciences
- Program Administrators for the Heller School and International Business School
- University Ombuds
- Office of Equal Opportunity.