Syllabus

Bus241f: Machine Learning and Data Analysis for Business and Finance

Key information

Instructor

- Blake LeBaron
- bblebaron@brandeis.edu
- http://www.brandeis.edu/~blebaron
- Sachar 204, 736-2258
- Office hours:

TA

- TBA

Times:

Class Times: TBA

Detailed information

Course Description

This course is a general topics course on machine learning tools, and their implementation through Python, and the Python package, Scikit Learn. Students will finish the class with a basic understanding of how to execute predictive analytic algorithms in both cross sectional and time series environments. They will also have a good sense for how to evaluate and test their predictive models. The course is statistical in nature, but will use only basic statistics from a standard one semester statistics class. Given the time horizon (6-7 weeks) it can only will provide a birds eye view of the many different ML technologies that are available. Finally, the course assumes a good working knowledge of the Python programming language at the start. Online courses in Python may be acceptable to meet this requirement.

Learning Goals

1. Basic data processing and handling with Python/Pandas
2. Machine learning tools available in Scikit Learn
3. Implementation of machine learning algorithms
4. Testing and evaluating forecasts/predictions (cross validation)
5. Presenting/describing results (graphics)

Prerequisites:

1. ECON213a/ECON184a (equivalent to most undergrad 1 semester classes in econometrics)
   1. Random variables, expectations, PDF’s, CDF’s
   2. Linear regression (Ordinary least squares)
2. Good working knowledge of Python computer language
   1. Knowledge of : Numpy, SciPy, and Pandas a plus
   2. FIN285a will be sufficient for this
   3. Online options such as Datacamp may be useful
3. Basic calculus (about 1 semester, undergrad level)
4. Basic matrix algebra will be useful too
5. Rudimentary knowledge of Excel

Required Readings:


Optional books:


Required Software

1. Python 2.7 and the entire Anaconda suite of tools. (This is open source and runs on all major operating systems.)

Grading

Grades will be based on:

1. Problem sets (25%)
2. Final exam (50%)
3. Group project (25%)

Rules and responsibilities

Communications

You are responsible for all announcements and materials in class. Also, much of the information in class will be sent over Latte and the class website.

Rules specific to ECON/FINNNN

- Exams
  - Your own work.
  - Closed book (no notes).
  - No laptops, no cell phones, no calculators, no pda’s.
- Problem sets
  - Hand in your own work.
  - Can talk and assist each other.
  - Use all resources.
- Group projects
  - Own work for the group.
  - Hand in one writeup per group.
- Laptops: Please bring to class if you want to.

Academic Integrity

You are expected to be honest in all of your academic work. Please consult Brandeis University Rights and Responsibilities for all policies and procedures related to academic integrity. Students may be required to submit work to TurnItIn.com software to verify originality.
Allegations of alleged academic dishonesty will be forwarded to the Director of Academic Integrity. Sanctions for academic dishonesty can include failing grades and/or suspension from the university. Citation and research assistance can be found at LTS - Library guides.

Work Load

Success in this two-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.)

Disability Statement

If you are a student with a documented disability on record at Brandeis University and wish to have a reasonable accommodation made for you in this class, please see me immediately.

Spring calendar dates

Course Outline

1. Introduction: What is Machine Learning? What is Artificial Intelligence?

2. Landscape of problems
   1. Supervised versus unsupervised learning
   2. Classification versus forecasting
   3. Time series/cross section
   4. Classifying data sets: Tall, wide, and dense data
   5. Predictive modeling/policy intervention

3. Python basics (very short)
   1. Anaconda
   2. Spyder
   3. Numpy, Scipy
   4. Scikit Learn

4. Supervised learning
   1. K nearest neighbors
   2. Linear regression
   3. Naive Bayes
   4. Decision trees
   5. Multi-trees, forests, and bagging predictors
   6. Support vector machines
   7. Neural networks and deep learning

5. Unsupervised learning
   1. Is this harder?, Challenges
   2. Preprocessing
   3. Principle components
   4. K-means clustering

6. Model evaluation
   1. Cross validation
   2. Evaluation metrics

7. Text processing (text data)
   1. Sentiment analysis
   2. Bag of words