

COLUMBIA BUSINESS SCHOOL
MACHINE LEARNING – B9653 – SPRING 2018

Dr. George A. Lentzas
Email: gl2543@columbia.edu

Time & Location: Wednesdays 6pm – 9pm at Uris 307.

Teaching Assistant: TBD

Course Description: This course is the first of two courses that will introduce students to the exciting and growing literature in machine learning / AI with a focus on applications in finance and marketing. We will cover topics such as regularization, tree methods, bagging/boosting, support vector machines and recommendation algorithms. In the process, we will review several real-world applications drawn from the areas of finance and marketing. Students are expected to be familiar with basic probability theory, linear algebra, and multiple linear regression. Some familiarity with (and willingness to learn) programming is a prerequisite as we will make extensive use of the programming language R.

Required Text: The main readings for this course are (1) James et al, "Introduction to Statistical Learning" (referred to as **ISLR**), (2) James et al, "The Elements of Statistical Learning" (referred to **ESL**) and (3) (optional) Kevin Murphy, "Machine Learning" (referred to as **MLPP**). These are in increasing difficulty, as are the readings in each week's reading list.

Class Preparation: Students are expected to have completed the Session Reading ahead of class to facilitate class participation and discussion. You need to take this seriously to be able to make the most out of this class as we will cover a lot of advanced material.

Research Paper: Students will write a research paper that addresses a well-defined finance or marketing question using the machine learning toolkit. Potential research projects will be circulated in class [more to follow].

Grading: Class Participation 25%, Homework 25%, Research Paper 25%, Final 25%. Homework is due promptly at 11:59pm, a week after it is released. Late homework will be penalized by 10% per 1hr delay.

Office Hours: TBD

SCHEDULE

SESSION 1 (INTRODUCTION TO MACHINE LEARNING) – JANUARY 24

- Types of Learning
- Regression vs Classification
- Supervised vs Unsupervised Learning (review of PCA)
- R Lab: A brief introduction to R
- Reading: ISLR Ch. 1-4, 10.1, 10.2 | R Cookbook Ch. 1-4 | ESL Ch. 2 (Optional)

SESSION 2 (RESAMPLING & REGULARIZATION) – JANUARY 31

- Cross Validation and the Bootstrap
- Linear Regularization: Lasso & Ridge Regression
- Comparison and Analysis of Regularization Methods
- R Lab: the package glmnet and a salary prediction example
- Reading: ISLR Ch. 5-6 | ESL Ch. 3, 7
- Homework 1: Automatic Factor Selection (Finance Application)

SESSION 3 (TREE BASED METHODS & SVMs) – FEBRUARY 7

- Support Vector Machines
- Basic Decision Trees (CART)
- Reading: ISLR Ch. 8.1, 9 | ESL Ch. 9.2, 15

SESSION 4 (BOOSTING, BAGGING & ENSEMBLES) – FEBRUARY 14

- Ensemble Learning
- R Lab: Time Series Forecasting with Ensemble Learning
- Reading: ESL Ch. 16
- Boosting and Bagging
- Random Forests
- R Lab: Sales Prediction Example
- Reading: ISLR Ch. 8.2.1, 8.3 | ESL Ch. 10, 15
- Homework 2: Bank Telemarketing Success Prediction (Marketing Application)

SESSION 5 (NON-LINEAR MODELS & UNSUPERVISED LEARNING) – FEBRUARY 21

- Splines and Local Regressions
- Generalized Additive Models
- Clustering Methods
- Reading: ISLR Ch. 7, 10.3–10.7 | ESL Ch. 14, 9.1

SESSION 6 (RECOMMENDATION ALGORITHMS) – FEBRUARY 28

- Introduction to Recommender Systems
- Neighborhood Based Collaborative Filtering
- Reading: Recommender System, Charu C. Aggarwal, Ch 1–2.3
- Model Based Collaborative Filtering
- Content Based Recommender Systems
- Reading: Recommender System, Charu C. Aggarwal, Ch 3–4
- Homework 3: Product Recommendations (Marketing Application)