COLUMBIA BUSINESS SCHOOL

MACHINE LEARNING — B9653 - SPRING 2024

Dr. George A. Lentzas
Email: george.lentzas@columbia.edu

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, marketing and business. We will cover topics such as regularization, tree methods, bagging/boosting, support vector machines and recommendation algorithms. and review several real-world applications. 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**), and (2) James et al, "The Elements of Statistical Learning" (referred to **ESL**). These are in increasing difficulty.

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 business topic using the machine learning tools covered in class.

Grading: Quizzes 25%, Homework 25%, Research Paper 25%, Final 25%. Homework is due promptly at 12:00pm, a week after it is released. There will be a 5% per 1hr delay penalty for late submissions (both homeworks / research paper).

Office Hours: Weekly with TA (hours tba) and by appointment with professor.

SCHEDULE

SESSION 1 (INTRODUCTION TO MACHINE LEARNING)

- Types of Learning
- Regression vs Classification
- A Brief History of Machine Learning
- Types of Error
- R Lab: A brief introduction to R - Part I
- Reading: ISLR Ch. 1-4 | R Cookbook Ch. 1-4 | ESL Ch. 2 (Optional)

SESSION 2 (RESAMPLING METHODS)

- Resampling methods: Cross Validation
- Comparison and Analysis of Regularization Methods
- R Lab: A brief introduction to R - Part II
- Reading: ISLR Ch. 5-6 | ESL Ch. 3, 7

SESSION 3 (REGURALIZED LINEAR MODELS)

- Linear Regularization: Ridge Regression
- Linear Regularization: LASSO
- R Lab: the package glmnet and a salary prediction example
- Reading: ISLR Ch. 8.1 | ESL Ch. 9.2, 15

SESSION 4 (TREES, RANDOM FORESTS & BOOSTING)

- Basic Decision Trees (CART)
- Ensemble Learning
- Random Forests
- Boosting & XGBoost
- R Lab: Sales Prediction Example
- Reading: ISLR Ch. 8.2.1, 8.3 | ESL Ch. 10, 15, 16

SESSION 5 (UNSUPERVISED LEARNING)

- Principal Components Analysis
- Clustering Methods
- R Lab: Practical Applications of PCA
- Reading: ISLR Ch. 7, 10.3-10.7 | ESL Ch. 14, 9.1

SESSION 6 (RECOMMENDATION ALGORITHMS)

- Introduction to Recommender Systems
- Neighborhood Based Collaborative Filtering
- Model Based Collaborative Filtering
- Content Based Recommender Systems
- Reading: Recommender System, Charu C. Aggarwal, Ch 1-2.3, 3-4