Instructors.
Aleksandr Aravkin, sasha.aravkin@gmail.com
Aurélie C. Lozano, aclozano.huang@gmail.com

Course Description and Objectives. The course will cover major statistical learning methods for data mining under both supervised and unsupervised settings. Topics covered include linear regression and classification, model selection and regularization, tree-based methods, support vector machines, and unsupervised learning. Students will learn about the principles underlying each method, how to determine which methods are most suited to applied settings, concepts behind model fitting and parameter tuning, and how to apply methods in practice and assess their performance. We will emphasize roles of statistical modeling and optimization in data mining.

Prerequisites. Linear Algebra. Recommended: Probability and Statistics. Scientific Programming in a language such as R, Matlab is desirable. Please speak to the instructor if you have any concerns regarding prerequisites.

Recommended Textbook and Further Reading.

- Introduction to Statistical Learning by James, Witten, Tibshirani & Hastie; available online at http://www-bcf.usc.edu/~gareth/ISL/.
- Statistics for High-Dimensional Data by Bühlmann & van de Geer.

Course Webpage. https://sites.google.com/site/saravkin/teaching/ieore4570. Lecture slides will typically be posted prior to each lecture.

Grading Policy. Homeworks 30% - Midterm 20% - Class Participation 15% - Final Exam 35%

Homeworks. There will be 6 homework assignments, covering theoretical understanding, computation, and applied data analysis.
Exams. Midterm Exam: There will be a midterm exam on March 6th, and a take-home final exam assigned around May 1st. Exams must be completed individually, according to the honor code and Columbia Academic Honesty policy [http://www.cs.columbia.edu/education/honesty](http://www.cs.columbia.edu/education/honesty).

Tentative Schedule. (Subject to change with reasonable advance notice).

- January 23, 2015: Overview
- January 30, 2015: Linear regression
- February 6, 2015: Classification
- February 13, 2015: Resampling
- February 20, 2015: Regularization
- February 27, 2015: Applications (part 1)
- March 6, 2015: Midterm
- March 13, 2015: Tree methods
- March 20, 2015: No Class, spring recess
- March 27, 2015: SVMs
- April 3, 2015: Applications (part 2)
- April 10, 2015: Unsupervised learning: PCA
- April 17, 2015: Unsupervised learning: clustering
- April 24, 2015: Applications (part 3)
- May 1, 2015: Final exam