Manhattan College
Mathematics Department

Statistical Learning

Class Time: TF 12:30-1:45 P.M. *Class Room*: RLC 104

Instructor: Angel R. Pineda, Ph.D. *Office*: RLC 201.J
Email: angel.pineda@manhattan.edu *Phone*: 718-862-7730
Web Page: http://turing.manhattan.edu/~apineda01/

Office Hours: Tuesday 10-10:50 A.M., Wednesday 11-11:50 A.M., Friday 10-10:50 A.M. and or by appointment.

Textbooks:

Required:

1) The Elements of Statistical Learning: Data Mining, Inference and Prediction, Second Edition, (Springer Series in Statistics) 2009.
by Trevor Hastie, Robert Tibshirani and Jerome Friedman
Publisher: Springer
This text is available for free online:
https://web.stanford.edu/~hastie/ElemStatLearn/

2) An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) 2013.
by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
Publisher: Springer
This text is available for free online:
http://www-bcf.usc.edu/~gareth/ISL/

Recommended:
1) Introductory Statistics with R (Statistics and Computing) 2nd Edition
by Peter Dalgaard
Publisher: Springer
This text is available in digital format from Pollack Library.

Supplemental Video Lectures by Book Authors:
https://www.r-bloggers.com/in-depth-introduction-to-machine-learning-in-15-hours-of-expert-videos/

Catalog Course Description:
This course develops the fundamental ideas of statistical learning for drawing conclusions from multivariate data sets using statistical theory and applied linear algebra. The course combines a theoretical presentation with computation of the resulting algorithms on real data sets to develop intuition of both how the methods work and how they perform in practice. It will cover the major techniques and concepts for both supervised and unsupervised learning. Topics will include regression, classification, resampling methods, model selection, regularization, principal components and clustering. Optional selected topics include tree-based methods, support vector machines and neural networks.

Prerequisites: MATG 633 and MATG 671 or permission of the Graduate Director.

Course Objectives:
After completing this course, the students should be able to:

- Understand machine learning using statistical modeling and inference
- Understand and apply the trade-off between bias and variance in model building
- Understand and apply the trade-off between prediction accuracy and model interpretation
- Apply and understand multivariate linear regression for prediction and inference
- Apply and understand linear classification methods
- Apply and understand cross-validation and bootstrap for model evaluation
- Apply and understand linear model selection
- Apply and understand linear model regularization using LASSO and Ridge Regression
- Apply and understand principal component analysis for unsupervised learning
- Apply and understand K-means clustering
- Apply and understand statistical learning methods to real world data science problems

Course Homepage (Moodle):
Here you will find four features that will be used in this course:

- *Email:* make sure that your email on Moodle is one that you check regularly. Homework assignments, announcements and other class related information will be sent via email.
- *Course Information and Documents:* material covered each week, assignments and solution keys.
- *Student Discussion Board:* this online forum allows for students and faculty to communicate about the course.
- *Grades:* students will be able to keep track of their grades online.

Grading:
Homework (50 %)
There will be weekly assignments covering the material in lecture. The major amount of learning for the class will be in these weekly assignments.

Take Home Midterm Exam (20 %) Friday November 3, 2017

Final Project (30 %), Tuesday December 12, 1:30-3:30 p.m.
The final project for this class will emphasize practical application of statistical learning to a data science problem generating code, a paper and presentation.

Tentative Grading Scale

Percent	90-100	85-89	80-84	75-79	70-74	65-69	60-64	50-59	0-49
Grade	A	A-	B+	B	B-	C+	C	D	F

The exact grading scale will be determined after the final exam. The numerical scores in the tentative grading scale guarantee the associated letter grade but the instructor may change the scale to the student's benefit.

Dates to Remember

September 1: Late Registration & Add/Drop Ends
October 10: Monday Schedule (No Classes)
October 16: Midterm Grades Due
November 17: Last day to withdraw from Courses
November 22-24: Thanksgiving (No Classes)
December 8: Last Day of Classes

Class Policies

- Late homework will not be accepted after the solutions are distributed. In case the homework is handed in before the solutions are posted it will be marked 20% off for every day (or part thereof) it is late.
- The lowest HW grade will be dropped.
- No make-up exams will be given, unless you have a medical or family emergency. These emergencies require valid documentation. The grade for a missed exam is zero.
- Cell phones (or other technology not related to the class) in the classroom is only allowed with express permission of the instructor for special circumstances. In general cell phone or other potentially disruptive technology use is not allowed in class.

Suggestions

- The course requires a time commitment of about 9 hours outside of class time. Make sure to make enough space in your schedule to spend the time needed.
- I suggest you work in groups on your homework but hand in individual solutions, not copied from each other. Doing the homework is when most of the learning occurs.
- I encourage you to come to office hours regularly. I will do my best to help you.

Academic Integrity:
Recall that as students of Manhattan College, you have each signed The Manhattan College Honor Pledge as a part of the Honor Code:

As a Manhattan College student, I will not lie, cheat, or steal in my academic endeavors, nor will I accept the actions of those who do. I will conduct myself responsibly and honorably in all my activities as a Manhattan College student. I am accountable to the Manhattan College community and dedicate myself to a life of honor.

Whenever you put your name on work to be handed in for grading in this class, you are reaffirming the above pledge. Violations of the Honor Code include, but are not limited to, cheating, plagiarism, fabrication, and other forms of academic misconduct. Please see the Manhattan College Community Standards, pp 45-47, for specific examples of the above.

Special Accommodations:

- Students with special needs should bring appropriate documentation to the Specialized Resource Center, Miguel 300, http://manhattan.edu/academics/specialized-resource-center, to obtain an Academic Adjustment/Auxiliary Aid form. Bring the completed form to me as soon as possible, and together we will decide on how best to fulfill the adjustments and/or aids listed on the form.
- Student athletes should bring their event schedules to me as soon as possible.

The material in this syllabus may be changed at the instructor's discretion. Any changes will be communicated to the students.