ISYE 7406 Data Mining And Statistical Learning
(Fall 2026)

Instructor Information

Dr. Yu Ding, Email: yu.ding@isye.gatech.edu
Office hours: 9:15-10:00 am Tuesday and Thursday
Office: To be decided. Office will be relocated in summer 2026

Teaching Assistant

If this course has a TA, the TA's contact information and his/her office hours will be announced through Canvas.

Classroom and Schedule

Time: 8:00 - 9:15 AM, Tuesday and Thursday
Location: George Tower Scheller Tower, Room 0350

Course Description

This is a fundamental course teaching the theories, ideas, and algorithms behind machine learning methods, designed for graduate students who have basic machine learning knowledge and some experience with off-the-shelf machine learning software packages but want to know how and why those machine learning methods work. Although computer coding is expected in data modeling exercises, this is NOT a software-teaching course. The expectation is that students should have basic machine learning coding capability before enrolling in this course. The course is structured in four parts and covers the following topics

Part 1: Regression, subset selection, and extensions

- **(Week 1-2)** Unit 1: Regression and subset selection (lasso and ridge regressions)
- **(Week 3)** Unit 3: Logistic regression for classification
- **(Week 4)** Unit 4: Basis expansions (splines)

Part 2: Prediction error assessment and model selection

- **(Week 4-5)** Unit 5: Training and prediction errors and bias-variance decomposition
- **(Week 6)** Unit 6: Assessing in-sample prediction errors (C_p, AIC, BIC)
- **(Week 7)** Unit 7: Assessing extra-sample prediction errors (cross validation and bootstrap)

Part 3: Popular methods for statistical learning

- **(Week 7-8)** Unit 8: Gaussian process regression
- **(Week 8-10)** Unit 9: Support vector machine (for classification) and kernel trick
- **(Week 10-11)** Unit 10: Classification and regression tree (CART)
- **(Week 11)** Unit 11: Neural nets

Part 4: Prediction enhancement methods

(Week 11-12) Unit 12: Ensemble methods (boosting, MART, bagging, RF)
(Week 12-13) Unit 13: Data reduction (PCA and Autoencoder)

Academic Background Requirement

ISYE 6413 and ISYE 6414. Knowledge on linear regression and statistical testing is necessary. Knowledge on convex optimization, and on real and functional analysis, will be helpful. Proficiency in computer programming skill is assumed.

Course Leaning Outcomes

- Understand the basic principles of data science and machine learning predictive modeling;
- Comprehend the criteria and methods to evaluate data science predictive models;
- Comprehend a number of typical modeling tools in machine learning and be able to use them to solve the actual data analytic problems.

Textbook and Course Materials

GT Canvas: We will use Canvas extensively for posting course materials including datasets.

Textbooks (All books can be freely downloaded from the respective author's websites)

[**HTF09**] Hastie, Tibshirani and Friedman (2009, 2nd edition), *The Elements of Statistical Learning: Data-mining, Inference, and Prediction*, Springer-Verlag.

[**R&W06**] Rasmussen and Williams, 2006, *Gaussian Processes for Machine Learning*, MIT Press.

[**JWHT13**] James, Witten, Hastie, and Tibshirani (2013), *An Introduction to Statistical Learning*, Springer-Verlag.

Useful software: R or Python

The majority of the course materials come from HTF09 (drawing materials from a total of twelve chapters), with Unit 8 as one of the exceptions, coming from R&W06, and Unit 11 from mostly online sources.

Grade Distribution

Exam 1	25%
Exam 2	25%
Exam 3	25%
Project	25%

Homework will be assigned for the purpose of self-exercising. They will not be collected nor graded. Solutions will be posted before an exam. You are encouraged to discuss with

classmates on homework. But you will be better off to work out the final solution by yourself or write the computer code independently. Doing so will help your performance in the exams.

Letter Grade Assignment

- A-B cut off: 80 points or above, but no fewer than 50% of the letter-grade enrolled students to receive A;
- B-C cut off: 60 points or above, but no fewer than 90% of the letter-grade enrolled students to receive A or B;
- C-D cut off: 40 points or above, but no fewer than 97% of the letter-grade enrolled students to receive A or B or C.

Values are to be rounded to the nearest integer.

Examination Schedule

To be announced.

Course Project Description

Please see a separate document for details.

Attendance and Make-up Policy

In this class, students are expected to attend class and to complete all assignments. Please refer to the Student Rule IV in its entirety regarding attendance, approved Institute activities, and permissible makeups. The general rule is that only is the absence due to approved Institute activities, religious observance, and emergency allowed to make up (proper documentation or notification needed). Absences due to some special occasions might be allowed for make-up, ONLY IF you and I agree upon the make-up *prior to* that event.

Re-grading Policy

Any disagreements regarding a grade received on any graded material must be discussed within one week from the time when the corresponding score is posted in Canvas. *No grade will be changed beyond the one week limit.*

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. Review Georgia Tech's Honor Code and the student Code of Conduct.

GEORGIA TECH HONOR CHALLENGE STATEMENT
I commit to uphold the ideals of honor and integrity by refusing to betray the trust bestowed upon me as a member of the Georgia Tech community.

Any student suspected of cheating or plagiarism on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services (404-894-2563) as soon as possible to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement

At Georgia Tech, we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. The Student-Faculty Expectations articulate some basic expectations that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.