ISyE 7406: Data Mining and Statistical Learning

Semester: Spring 2027
Section: A (CRN 28782), in person; 3 credit hours
Instructor: Xiaoming Huo
Email huo@gatech.edu
Phone: 404-385-0354
Time 8:00 am - 9:15 am TR
Venue George Tower Scheller Tower Room: 0350
Office hours*:* Tuesday and Thursday, right after the lecture. We can walk to my office.
Or email for appointments.
My office George Tower 812
Prerequisite: graduate-level standing, plus ISYE 6413 and ISYE 6414, each with a minimum grade of D. You should also be comfortable with basic programming in R or Python.
Topics: An introduction to some commonly used data mining and statistical learning algorithms. This course emphasizes understanding the theoretical and statistical aspects of these algorithms, as well as implementing them with data examples using statistical software (e.g., R, Python). We will cover the following topics.

1. introduction to data mining and overview of supervised learning (cross-validation, K-nearest neighbor)
2. linear regression and model selection
3. shrinkage methods (ridge regression, LASSO, principal components, partial least squares)
4. linear classification (discriminant analysis, naive Bayes, logistic regression)
5. local smoothers and kernel smoothing
6. spline smoothing
7. additive models
8. tree-based methods
9. ensemble methods (bagging, random forest, boosting)
10. support vector machines and kernel methods
11. neural networks and deep neural networks
12. unsupervised learning (K-means, EM algorithm)

Catalog Description: Topics include neural networks, support vector machines, classification trees, boosting and discriminant analyses. Intended for Ph.D. students and those seeking the M.S. in Statistics.

Course Learning Outcomes

Understand the theoretical and statistical aspects of several widely used data mining and statistical learning methods.
Apply appropriate data mining and statistical learning methods to analyze real-world datasets.
Texts: optional; there is no required textbook, and all course materials will be available on Canvas. The following book is highly recommended and is available for free at https://www.statlearning.com/.

1. An Introduction to Statistical Learning, 2nd Edition (2021), Springer; Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; ISBN-13: 978-1071614174

Exam Dates: See the *Schedule* page.
Software: R and Python. You will use software for homework assignments.

Canvas: We will use Canvas (http://canvas.gatech.edu) to distribute course-related materials, including announcements, course syllabus, homework assignments and solutions, slides and other lecture material, practice exams, course grades on exams and homework, and any important announcements.

Piazza. This term, we will be using Piazza for class discussions. The system is highly catered to getting you help fast and efficiently from classmates and the TA. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Homework

- There are 10 *homework* assignments, roughly weekly/biweekly. You are allowed (and encouraged) to work together with other students on homework as long as you **write up** and turn in your own solutions. You are also allowed (and encouraged) to ask me questions, although you should try to think about the problems before asking.
- The homework should be submitted online at Canvas on the due date/time (11:59 pm ET).
- Code is mandatory as an Appendix in the report; otherwise, a 20% penalty will be applied.
- The TA will grade the homework. If you disagree with the TA's grade or comments, send a private Piazza message for a regrade. If you still disagree, the instructor will make the final decision.
- **The lowest homework score will be dropped.**
- No homework or other graded work is due on the Final Instructional Class Days (April 26-27) or during the Reading Period.

Exams

- Exams include two tests during the semester and a final. They are closed-book exams. For tests I, II, and final, **you can bring two, four, and six pages of double-sided notes**, respectively. In the notes, you can write down anything you like. But they have to be self-made. Photocopy is not allowed.

Grading Policy: Distribution of points:

- Homework 25 %
- Midterm 25 % + 25 % (25 for each exam)
- Final exam 25 %

Criteria for Final Grades

- A grade of "A" corresponds to a percentage grade of 90% or above on all graded work; a grade of "B" corresponds to a percentage grade of 80%-89.999%; a grade of "C" corresponds to a percentage grade of 70%-79.999%; a grade of "D" corresponds to a percentage grade of 60%-69.999%; and a grade of "F" corresponds to a percentage grade of 59.999% or below.

Teaching Assistant: This will be added when the information is available; details will be posted on the Canvas course web page under Syllabus.

Attendance Policy

Students are expected to attend all scheduled class sessions. Regular attendance is strongly correlated with academic success in this course. If you must miss a class, you are responsible for obtaining notes and any materials covered. Excessive unexcused absences may affect your final

grade at the instructor's discretion. Please notify the instructor in advance if you anticipate missing class due to illness, religious observances, or Institute-approved activities.

Other information and rules:

- Make sure the scores in Canvas (the GT online system) are consistent with what you got. I gave the final grades based on them.
- No extra credit unless it is arranged before any tests.
- We will not change grading for work older than two weeks.
- Zero-tolerance: any violation of honor codes (e.g., cheating on homework, tests, etc.) will result in an automatic F in this class.
- Avoid late arrival and early departure as they are distracting.
- Instructors have the discretion to cancel a class or pivot to digital instruction per the Institute Updates Weather Operations Policies. The announcement will be made accordingly.

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. For information on Georgia Tech's Academic Honor Code, please visit https://osi.gatech.edu/honor-code. Any student suspected of cheating or plagiarizing 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. www.honor.gatech.edu

Student-Faculty Expectations Agreement

- At Georgia Tech, we believe that it is important to strive for an atmosphere of mutual respect, acknowledgment, and responsibility between faculty members and the student body. See https://catalog.gatech.edu/rules/21/ for an articulation of 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. Student-Faculty Expectations Agreement

CIOS Incentives

If the CIOS response rate of this course is above 90%, the cutoffs for each level of the letter grade will be decreased by 1.

Accommodation 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 email me as soon as possible to set up a time to discuss your learning needs.

Extensions, Late Assignments, & Re-Scheduled/Missed Exams

We usually don't offer extensions, rescheduling, or make-up exams because it requires significant resources to accommodate them. I do not grant exemptions (including make-up tests) unless I receive an official letter from a legitimate Georgia Tech officer, including the registrar, dean of students, athletics department, or advisers of student organizations. Letters and reports from parents, church leaders, medical offices, and so on will not be sufficient - I do not have the resources to verify their validity. The letter must explain why an excuse is acceptable, and I will determine whether it is compelling enough.

Late submission of assignments will incur a penalty (10-75%, depending on the lateness). The penalty will be 10% if the submission is made within 24 hours, 30% if it is made within 48 hours, 50% if it is made within 72 hours, and 75% if it is made beyond 72 hours.

Incomplete grades:

An incomplete (I) grade should be considered only when a student can demonstrate a significant non-academic circumstance that impacts his or her work in the class, and **ONLY when he or she has already completed a substantial part of the course and has satisfactory academic performance.** If those conditions do not apply, the student should be counseled on how to succeed or on dropping the class while there is still time. **There is more information on the Registrar's Office website concerning Incomplete grades and grade changes – see: https://registrar.gatech.edu/faculty-and-staff/grading-and-grade-entry**.

Use of AI

When allowed for usage on an assignment, responsible use is expected. All submitted work must include a brief **AI Usage Statement** outline:

- **Which tools were used**
- **When they were used**
- **What prompts or questions were given**
- **How the AI output informed or shaped your final submission**

For specific guidelines on how to cite AI-generated content, please refer to the article "**How to cite ChatGPT**"

Failure to follow assignment guidelines for Generative AI —including using Generative AI when it is not permitted or failing to disclose its use when it is—may be considered a violation of Georgia Tech's academic integrity policies. When in doubt, always consult your instructor before using Generative AI.