ISyE 6740 – Computational Data Analysis
Tentative Syllabus Spring 2025

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor: Yao Xie; yao.xie@isye.gatech.edu
Professor Office Hour: Thursday 2-2:30 pm, after class
Office: Groseclose 445

Teaching Assistants:
TBA

Prerequisites

- Undergraduate level probability, linear algebra, and statistics.
- Basic programming using MATLAB or Python.

Course Description

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. The course is designed to answer the most fundamental questions about machine learning: What are the most important methods to know about, and why? How can we answer questions such as "is this method better than that one"? What can we say about the errors our method will make on future data? What's the "right" objective function? How to tune parameters? What does it mean to be statistically rigorous?

This course is designed to give graduate students a thorough grounding in the methods, theory, mathematics and algorithms needed to do research and applications in machine learning. The course covers topics from machine learning, classical statistics, and data mining. Students entering the class with a pre-existing working knowledge of probability, statistics and algorithms will be at an advantage, but the class has been designed so that anyone with a strong numerate background can catch up and fully participate. Some experience with coding is expected (in Python, or MATLAB).

For detailed course topics, please see the tentative course schedule.

Learning Objectives
After taking this course, students should be able to:

- Gain thorough understanding in the methods, theory, mathematics and algorithms needed to do research and applications in machine learning.
- Implementing and use machine learning algorithms.
- Gain experience with analyzing real data.

Textbooks/Readings

- **Textbook**: The course material will be based on lectures slides provided in the course.
- **Recommended References**:
    - (PRML) Pattern recognition and Machine Learning, Christopher M. Bishop.
    - (ESL) The elements of Statistical Learning: Data Mining, Inference, and Predictions, 2nd edition, Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
    - (FML) Foundations of Machine Learning, 2nd edition. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar.

- Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares, Stephen Boyd and Lieven Vandenberghe.
- The Matrix Cookbook, Kaare Brandt Petersen and Michael Syskind Pedersen.

COMMUNICATION

Content Questions and Help: Because an answer to an individual's question can often help others, please do not email your questions directly to the instructor. Instead, course and content questions will be addressed on **Ed Discussion**. Feel free to make a private post (**sent to all instructors**, so the professor and all the TAs will see your question) if your question involves private material such as your derivation work for a homework problem.

Ed Discussion Post Protocols: Please be courteous when posting on Ed Discussion and treat fellow students, the TAs, and professor with respect. In any public posts, please do not show any of your answers related to the homework problems, such as code snippets, mathematical derivations/proofs, etc. If you would like to show any plots (which do not disclose any explicit answer to the questions) from your implementation in the discussion, please either make them private (only share with instructors) and/or add watermarks to those images/results. Please make your questions specific, **do not ask, for example, for instructors to verify your results, or debug your code.**

When asking questions on Ed Discussion, please be specific. In principle, instructors are not responsible for debugging programs and will not comment on purely coding questions. For example, please do not send a code file to a TA or post a question asking why a section of code "doesn't work".

GRADING

Grades will be assigned on the following basis:

Homework:	70%
Project:	30%
- Project proposal	5%
- Project presentation	5%
- Project report	20%

Important: Make sure the scores in Canvas are consistent with the feedback you receive. We will not make any change in grading if asked more than 2 weeks after releasing grades. **We cannot accept any homework regarding requests after the final project report's due date. There is no regrade for the final project report.** If you have any questions about your grade/feedback, please **do not respond/discuss through the Canvas Grade pages**, instead, please send an email directly to your grading TA with your questions, as well as the regrade request.
The following grading scale will be used in the course:

- 90-100%: A
- 80-89%: B
- 70-79%: C
- 60-69%: D
- below 60%: F

HOMEWORK

Due Dates and Submissions: Homework should be submitted to Canvas **by 11:59 pm EST on the date it is due**. No submission will be accepted through email. We strongly encourage the use of LaTeX for your submission. Assignments will include both exercises and programming problems; the programming problems will ask you to carry out statistical analysis using statistical software. Keep in mind that you should NOT hand in raw program output. **Conclusions and interpretation of results are more important than printouts.** Computing output without proper explanation will not receive full grades. That said, notebook style submissions, Including Jupyter Notebooks (.ipynb), with code and explanation blocks, are acceptable. Please use MATLAB or Python for homework programming.

Please double-check your homework submission to make sure it is complete with all your answers before submission. Before the deadline, you can submit as many times as you like (your grader will only review the last submission). Additional submissions or replacing any part of the original submission past the deadline will have an appropriate penalty applied to it.

(Important) Homework code submission requirement

Please make sure you read through the below guide for homework submission and strictly follow the instructions. For each assignment, two files need to be submitted through Canvas before the deadline:

(1) A pdf with a name: 'GTUsername_HWx_report'. i.e. 'yxie77_HW1_report' In this report file, you need to summarize **ALL** your write-up answers to questions, including analytical answers, program output/images/summaries, etc. (Please do not include your code in the pdf, unless Jupyter Notebook is used, then you should print it into the required pdf). For this report, we strongly recommend typing your answers. You can use either Jupyter Notebook, Latex, or any other mark-down tool. If you choose to handwrite your report, please make sure your writing is legible enough for your grader.
(2) A zip file with name "GTUsername_HWx.zip": In this zip file, you need to have a folder with the same name: 'GTUsername_HWx.', which should include all your program files.

Failure to follow submission requirements will result in a 10-point penalty on your assignment grade. Please make sure you strictly follow the above file format requirements to avoid unnecessary loss of points. In addition, for programming problems, you can use either Matlab or Python (3.X). When you submit code, **please also include data in the folder**. Your TA will not be allowed to modify your code, so please make sure your code doesn't have directory dependence on your computer. In other words, **avoid the use of absolute paths.** If your code has a pathing issue, it will be considered as "not executable," and receive no credit.

You can work together with other students on homework, as long as your write-up and solutions represent your own work. You are also allowed (and encouraged) to ask the instruction team questions, but keep in mind that a good faith effort to work through the material beforehand will often help you form good questions. Any academic misconduct is subject to an F grade as well as referral to the Dean of students.

Homework accommodations: We have the following policies to help with emergent situations:

(1) You can have up to 10 days of homework extension without penalty. You do not need to make an extension request for individual homework assignments. Your assigned TA will keep track of your total extension usage, and you will receive a reminder email if your allowed days are used up. **Note that you may only use 7 extension days on any single assignment. No extension can be used for any final project deliverable.**

(2) If you have already used the above 10 days of homework extension, and if you submit the homework late: one day late the grade will be discounted to 75% of your total, two days late the grade will discount to 50% of your total, three days late the grade will discount to 25% of your total. Past three days, your homework will not be accepted, and you will receive no credit.

FORMAL EXCUSES

Officially, there are 2 channels via which a student can get formal excuses.

(1) A student can request a formal exemption from the Dean's office for illness or other personal issues by filling out a form at https://studentlife.gatech.edu/request-assistance
(2) A student may also make requests to the registrar for an institute absence, such as attending conference, religious observances (within the first 2 weeks of the course) https://registrar.gatech.edu/info/institute-approved-absence-form-for-students

PROJECT

Project Team: Projects will be done in groups of **1-3** students. Please plan for the project early; you need to form your own team (e.g., you can ask on Ed Discussion). You will also need to come up with the topic of your project. You can discuss with the professor or TAs if you need some suggestions to come up with a topic. If you decide to do a team project, each team member will receive the same grade. Also, your project grading will take into account the size of your team. You are responsible for the collaboration and work division inside your group. With this in mind, the instruction team will not provide judgment for any disputes among team members; such cases will be escalated to the Dean's office.

Project Submission: In the middle of the semester, each team will need to submit a project proposal (see Course Schedule for the due date). By the end of the semester (see Course Schedule for the due date), each group needs to submit one project report. No data or code submission will be required (only proposal and report pdfs). Please see the Project Guidelines document on Canvas for additional details and format requirements.

Project Grading:

- 5% Project Proposal (this is mainly a format and progress check; while you will receive no detailed feedback on the proposal, you are welcome to discuss topics and ideas with the TAs and professor).
- 20% on Project Final Report (5% on creativity and project scope; 5% on formulation; 5% on the rigor of implementation, 5% on report writing quality).
- 5% on Project Presenation, which will happen at the last lecture of the semester. More details be posted later in the semester.

Project due dates:

Please note that the final project due date is firm, and **there is no extension policy for any project deliverable.** In both cases, this is to ensure that we are able to provide feedback/grades in a timely manner, particularly for the final report. The final project report due date is close to the university registrar's due date for a letter grade, and we need the time to grade the final project and calculate final letter grades. If you foresee any situation that prevents you from submitting final project deliverables by the due date, please let us know as soon as possible. (For documented reasons, we may be able to assign your letter grade as "incomplete" without the final project report and update the letter grade in the next semester).

PLAGIARISM

Plagiarism is considered a serious offense. You are not allowed to copy and paste or submit materials created or published by others, as if you created the materials. All materials submitted and posted must be your own original work.

STUDENT HONOR CODE

You are responsible for completing your own work. All students are expected and required to abide by the *letter* and the *spirit* of the Georgia Tech Honor Code. The teaching assistants and I will also abide by the Honor Code. I am very serious about this expectation because ethical behavior is extremely important in all facets of life. To review the Georgia Tech Honor Code, please visit http://osi.gatech.edu/content/honor-code . Any student suspected of behavior in violation of the Georgia Tech Honor Code will be referred to Georgia Tech's Office of Student Integrity.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404) 894-2563 or http://disabilityservices.gatech.edu/, 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.

AI POLICY

We treat AI-based assistance, such as ChatGPT and Copilot, the same way we treat collaboration with other people: you are welcome to talk about your ideas and work with other people, both inside and outside the class, as well as with AI-based assistants. **However, all work you submit must be your own**. You should never include in your assignment anything that was not written directly by you without proper citation (including quotation marks and in-line citation for direct quotes).
Including anything you did not write in your assignment without proper citation will be treated as an academic misconduct case. If you are unsure where the line is between collaborating with AI and copying AI, we recommend the following heuristics:
Heuristic 1: Never hit "Copy" within your conversation with an AI assistant. You can copy your own work into your own conversation, but do not copy anything from the conversation back into your assignment.
Instead, use your interaction with the AI assistant as a learning experience, then let your assignment reflect your improved understanding.
Heuristic 2: Do not have your assignment and the AI agent open at the same time. Similar to the above, use your conversation with the AI as a learning experience, then close the interaction down, open your assignment, and let your assignment reflect your revised knowledge.
This heuristic includes avoiding using AI directly integrated into your composition environment: just as you should not let a classmate write content or code directly into your submission, so also you should avoid using tools that directly add content to your submission.
Deviating from these heuristics does not automatically qualify as academic misconduct; however, following these heuristics essentially guarantees your collaboration will not cross the line into misconduct.

Course Schedule

All assignments are due by 11:59 pm EST (electronic submission via Canvas) on the date listed below.

Week/Dates	Module/Topic	Weekly Overview	Deliverables
Week 1 01/06-01/12	Introduction and Overview, Clustering and k-means	Overview of the topics and scope of the class. We will introduce a building block of a fundamental problem in unsupervised learning (clustering): K-means.	Homework 1 Release 01/06 Due 01/19
Week 2 01/13-01/19	Spectral Clustering	We will discuss another type of clustering algorithm: spectral clustering, which is different from k-means since it is based on geometry (connectivity) of data	
Week 3 01/20-01/26 (01/20 MLK Day)	Dimensionality Reduction and PCA	We will present a linear dimensionality reduction technique called PCA.	Homework 2, Release 01/20 Due 02/02
Week 4 01/27-02/02	Nonlinear Dimensionality Reduction	We will introduce general non-linear dimensionality reduction techniques.	
Week 5 02/03-02/09	Density Estimation	Discuss basic density estimation method, which captures the distributional information of the data	Homework 3 Release 02/03 Due 02/16
Week 6 02/10-02/16	Gaussian Mixture Model and EM Algorithm	We will present a popular type of model for densities called Gaussian mixture models and discuss how to fit such models	
Week 7 02/17-02/23	Basics of Optimization Theory	We will introduce the essentials of optimization theory, foundational to developing machine learning algorithms.	Homework 4 Release 02/17 Due 03/02
Week 8 02/24-03/02	Classification Naïve Bayes and Logistic Regression	Introduce classification problems and basic methods for classification.	**Project Proposal Due 03/02.**
Week 9 03/03-03/09	Support Vector Machine (SVM), Neural Networks	Introduce SVM classifier and related theory, Understand basic neural networks.	Homework 5 Release 03/03 Due 03/23
Week 10 03/10-03/16 03/17-03/23 Break	Feature selection and anomaly detection	Feature selection methods and introduction to anomaly detection	
Week 11 03/24-03/30	Boosting Algorithms and AdaBoost	Introduce basic boosting algorithms and AdaBoost.	Homework 6 Release 03/24 Due 04/13
Week 12 03/31-04/06	Random Forest	Introduce tree-based methods for regression and classification, and random forest	
Week 13 04/07-04/13	Bias-Variance Tradeoff and Cross-Validation	Introduce principle of bias-variance tradeoff and how it is used to cross-validation for model selection and parameter tuning	
Week 14 04/14-04/20	(Bonus topics, no homework)	Kernel methods, Reinforcement learning	
Week 15 04/22	Project presentation in class		**Project Report Due – 04/27**