CSE/ISyE-6740A Fall 2026

Computational Data Analysis (CDA)

Georgia Tech, College of Computing

Monday Wednesday, 12:30 - 1:45pm

Technology Square Research Building Room 132

Instructor Information

Instructor	Email	Drop-in Hours & Location
Kai Wang	kwang692@gatech.edu	TBD

General Information

Description

Machine learning studies the question "how can we build computer programs that automatically improve their performance through experience?" This includes learning to perform many types of tasks based on many types of experience. For example, it includes robots learning to better navigate based on experience gained by roaming their environments, medical decision aids that learn to predict which therapies work best for which diseases based on data mining of historical health records, and speech recognition systems that learn to better understand your speech based on experience listening to you.

The course is designed to answer the most fundamental questions about machine learning: How can we conceptually organize the large collection of available methods? What are the most important methods to know about, and why? How can we answer the question “is this method better than that one” with some theoretical guidance or for a specific dataset of interest? What can we say about the errors our method will make on future data? What's the 'right' objective function? What does it mean to be statistically rigorous? Should I be a Bayesian? What computer science ideas can make ML methods tractable on modern large or complex datasets? What are the open questions?

Throughout this course, you will gain a thorough understanding in the concepts, methods and algorithms needed to do research and applications in machine learning and data science. The course covers topics from machine learning, classical statistics, data mining, Bayesian statistics and information theory. Students entering the class with a pre-existing working knowledge of probability, statistics, linear algebra and algorithms will be at an advantage.

In this generation, data science and machine learning have brought numerous transformative changes to many interdisciplinary fields, including traditional computer science related topics, e.g., computer vision, language models, robotics, to more broadly science and society related topics, e.g., biology, physics, urban planning, medicine, and social science, etc. I expect this course to help students gain experiences and interests, and equip with the necessary ability to explore more applications outside of the course.

Pre- &/or Co-Requisites

Algorithms, linear algebra, probability, and statistics.

We will not be providing basic support for numpy, sklearn, git, etc. You will be expected to have the software engineering skills to work with data sets of 100,000+ rows. All code will be written in Python. We will primarily use PyTorch for all the programming examples and homework. No other languages will be supported.

Course Goals and Learning Outcomes

Upon successfully completion of this course, you should be able to:

- Understand the theory, foundation, and limitations of different machine learning and data analysis algorithms
- Implement machine learning and data analysis algorithms
- Apply machine learning and data analysis algorithms to different applications
- Create new machine learning and data analysis algorithms to solve new challenges

Course Requirements & Grading

Assignment	Weight (Percentage, points, etc)	Date
Homework	45% • Hw0: 5% • Hw1-hw4: 10% each	See course schedule
Midterm (in person)	20%	12:30pm – 1:45pm (in class) Date: 10/14/2026
Final (in person)	35%	Date and time: TBD

Homework deadlines: Sunday (11:59pm ET)!!

Extra Credit Opportunities

Assignment	Weight (Percentage, points, etc)	Date
LaTeX class notes (individual)	3% (graded SAT/UNSAT) – first come first serve, notes in latex	Voluntary, starting from Week 2
Bonus homework questions	There will be additional challenging questions in the homework that you can solve and get bonus.	To be assigned in each homework

Grading Scale

Your final grade will be assigned as a letter grade according to the following scale. We will round it to the nearest integer for the final scores.

A	90-100%
B	80-89%
C	70-79%
D	60-69%
F	0-59%

All the exams will be closed book. We will curve the score distribution linearly if necessary. If you have an approval from the institute for extra exam time, we will accommodate it, and you can take the exam at the Disability Services Testing Center. If you encounter any unexpected situations and need special accommodations, please contact Dr. Wang and the head TA as soon as possible and we will try to help.

Course Materials

Course Text

- "Pattern Recognition and Machine Learning", Christopher M. Bishop
- "The Elements of Statistical Learning", Trevor Hastie, Robert Tibshirani, Jerome Friedman

Additional Materials/Resources

To be assigned on Canvas.

Course Expectations & Guidelines

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 http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Any evidence of cheating or other violations will be referred to the Dean of Students with a recommendation that the penalty be an award of zero points for the graded requirement, and a one letter grade reduction in the course. Cheating includes, but is not limited to: using unauthorized references or notes; copying directly from any source, including friends, classmates, tutors, or a solutions manual; allowing another person to copy your work; taking an exam or handing in a graded requirement in someone else's name, or having someone else take an exam or hand in a graded requirement in your name; or asking for a re-grade of a paper that has been altered from its original form.

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.

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.

Assignment, Collaboration, and Late Assignment Policies

Each homework will be announced around the beginning of the week each with at least 2 weeks for students to work on it (due by **Sunday 11:59pm ET**, unless further notice). Homework is **penalized by 20% for each day** that it is late (this applies additively, meaning that no credit is gained after 5 late days). We strongly encourage the use of LaTeX for your submission. Unreadable handwriting is subject to zero credit. We encourage you to discuss course content, homework problems, and prepare midterms together with your classmates. However, all answers and codes should be prepared independently. If you refer to any material, it should be properly cited. If you discussed homework problems with your classmates, indicate which problems you discussed with whom. Any kind of academic misconduct is subject to F grade will be reported to the Dean of Students.

Extensions & Re-Scheduled/Missed Exams

Make-up exams will be given only for documented reasons of illness, family emergency, or participation in approved Institute activities (such as field trips and athletic events, see http://catalog.gatech.edu/rules/12/ for more information). Exams will be closed book, and no calculators will be allowed. Final exams will be re-scheduled in accordance with Georgia Tech's policy, also found at http://www.catalog.gatech.edu/rules/12/. If you happen to need any accommodation that meets the circumstances mentioned above, please contact the teaching assistants and the instructor as soon as possible.

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. See http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectation 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.

Generative AI Policy

This course is about growing in your ability to write, communicate, and think critically. Generative AI agents such as ChatGPT, DALL-E 2, and others present great opportunities for learning and for communicating. However, AI cannot learn or communicate for you, and so cannot meet the course requirements for you.

In this course, using generative AI tools in the work of the course (including assignments, discussions, ungraded work, etc.) is allowed only in instances specified by your instructor.

As with any technology, generative AI tools need to be used critically and according to academic and professional expectations. Thus, in instances in which your instructor allows generative AI tool use, you are expected to adhere to these principles:

- Responsibility: You are responsible for the work you submit. In instances in which your instructor allows generative AI tool use, this means that any work you submit should be your own, with any AI assistance appropriately disclosed (see "Transparency" below) and any AI-generated content appropriately cited (see "Documentation" below). This also means you must ensure that any factual statements produced by a generative AI tool are true and that any references or citations produced by the AI tool are correct.
- Transparency: Any generative AI tools you use in the work of the course should be clearly acknowledged as indicated by the instructor. This work includes not only when you use content directly produced by a generative AI tool but also when you use a generative AI tool in the process of composition (for example, for brainstorming, outlining, or translation purposes).
- Documentation: You should cite any content generated by an AI tool as you would when quoting, paraphrasing, or summarizing ideas, text, images, or other content made by other people.

Using generative AI tools at times not allowed by the instructor will be considered an infraction of the Georgia Tech Honor Code subject to investigation by the Office of Student Integrity. Likewise, using generative AI tools in the course without adhering to these principles will be considered an infraction of the Georgia Tech Honor Code subject to investigation by the Office of Student Integrity.

Campus Resources: Knack — a New Leading Peer Tutoring Platform

Students looking for additional assistance outside of the classroom are advised to consider working with a peer tutor through Knack. Georgia Tech has partnered with Knack to provide students access to verified peer tutors who have previously aced this course. To view available tutors, visit gatech.joinknack.com and sign in with your student account.

Course Schedule

The course schedule of this class will be updated on Canvas progressively during the semester at least one week prior to the class. This following is a temporary one and may change over time depending on students' feedback. It will include class topics, assigned readings, and assignment due dates. You can refer to the "Course Requirements & Grading" for the grading policies of the assignments.

Date	Topic	Reading, notes, due dates, and more
8/24	Introduction and logistics	Homework 0: familiarizing yourself with Gradescope, autograder
8/26	Clustering, K-means algorithm	
8/31	Dimensionality reduction Principal component analysis	
9/2	Density estimation, histograms, kernel density estimator	Homework 0 due on 9/6 Homework 1 out before 9/6
9/7	**Labor Day, no class**	
9/9	Gaussian mixture models	
9/14	Expectation maximization (EM) algorithm	
9/16	Feature selection and decision trees	Homework 1 due on 9/20 Homework 2 out before 9/20
9/21	Naive bayes, KNN and logistic regression	
9/23	Support vector machines	
9/28	Support vector machines	
9/30	Kernel methods	Homework 2 due on 10/4
10/5	**Fall break**	
10/7	Gaussian process regression	
10/12	Midterm review	
10/14	**Midterm exam**	
10/19	Gradient descent	
10/21	Regularization and bias-variance tradeoff	Homework 3 out before 10/25
10/26	Neural networks	
10/28	Convolutional neural networks	10/31 withdrawal deadline

Date	Topic	Reading, notes, due dates, and more
11/2	**No class. Instructor conference traveling**	
11/4	Autoencoders and variational autoencoders	Homework 3 due on 11/8 Homework 4 before 11/8
11/9	Diffusion models	
11/11	Recurrent neural networks	
11/16	Transformer	
11/18	Introduction to reinforcement learning	Homework 4 due on 11/22
11/23	Policy iteration and value iteration	
11/25	**Thanksgiving holiday, no class**	
11/30	Model-free RL	
12/2	Deep reinforcement learning	
12/7	Final review	Last class
TBD	**Final exam**	TBD