Welcome

Instructor:

- Theodore J. LaGrow, **tlagrow@gatech.edu (mailto:tlagrow@gatech.edu)**

Head TAs:

- Dan Boros, **boros@gatech.edu (mailto:boros@gatech.edu)**
- John Mansfield, **jmansfield6@gatech.edu (mailto:jmansfield6@gatech.edu)**
- Danyang Cai, **dcai38@gatech.edu (mailto:dcai38@gatech.edu)**
- Jake Knigge, **jwk@gatech.edu (mailto:jwk@gatech.edu)**

Creators of Recorded Material:

- Charles Isbell
- Michael Littman

Office Hours:

- See Ed Discussion for details.

Required Text:
Machine Learning
by Tom Mitchell, McGraw Hill, 1997

Tom now owns the copyright and has made the textbook completely *free* and online.
http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html
(http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html)

Video Lectures:

"Modules" in Canvas
Beta: "Ed Lessons" in Canvas

General Information

Machine Learning is a three-credit course on the study and application of the field of Machine Learning. Machine Learning is that area of Artificial Intelligence that is concerned with computational artifacts that modify and improve their performance through experience. The area is concerned with issues both theoretical and practical. This particular class is a part of a series of classes in Machine Learning and takes care to present algorithms and approaches in such a way that grounds them in larger systems. We will cover a variety of topics, including statistical supervised and unsupervised learning methods, randomized search algorithms, and Bayesian learning methods. The course also covers theoretical concepts such as inductive bias, the PAC and Mistake-bound learning frameworks, minimum description length principle, and Ockham's Razor. Additionally, this term will have a focus on adding advanced topics to help bring the material closer to the state-of-the-art in ML practices. In order to ground these methods, the course includes programming and practical application with a number of projects.

Objectives

There are four primary objectives for the course:

- To provide foundational knowledge with a broad survey of approaches and techniques in ML
- To develop a deeper understanding of several major topics in ML
- To develop the design and programming skills that will help you to build intelligent, adaptive artifacts
- To develop necessary skills to communicate research and practicum in ML

The last bullet point is the core objective of this course and separates this ML offering from many others available online or even at GT. Without proper and efficient communication, no amount of coding or results will matter. You should develop enough background that you can pursue any desire you have to learn more about specific techniques in ML, either to pursue ML as a research career or to apply ML techniques in other research areas.

Prerequisites

The official prerequisite for this course is an introductory course in artificial intelligence. In particular, those of you with experience in general representational issues in AI, some AI programming, and at least some background in statistics, vector calculus, information theory, and linear algebra should be adequate. Any student who did well in an introductory AI course should be fine. You will note that most semi-modern AI courses suggest at least some tentative background in some machine learning techniques as well. Of course, having said all that, the most important prerequisite for enjoying and doing well in this class is your interest in the material. We say this in every semester and in every course, but it's true. In the end, it will be your own motivation to understand the material that gets you

through it more than anything else. If you are not sure whether this class is for you, please contact the instructor or Head TAs.

Resources

- **Readings.** The textbook for the course is *Machine Learning* by Tom Mitchell. This is an older text, however the material is foundational towards understand the basics of ML and freely available online. With the rise of LLMs and GPTs, there is a LOT of misinformation afoot online. Make sure you understand your information sources. Be warned. We will follow the textbook closely for most of the semester as many of the pre-recorded lectures align week-to-week, so it is imperative that you have a copy of the book. We will also use supplemental readings, but those will be provided for you. Weekly supplemental readings will be updated to support advanced topics posted throughout the semester.
- **Computing.** Even though you will not need high-throughput dedicated resources, you will have access to CoC clusters for your programming assignments. You can test your code on the Shuttles cluster, using your GT username and password to log in - you will not need a CoC account for this course. More info can be found here: **https://support.cc.gatech.edu/facilities/general-access-servers (https://support.cc.gatech.edu/facilities/general-access-servers)** . If needed, every assignment can be queued up and run on a free instance of Colab (barring some computational lag).
- **Web.** We will use the class canvas page and Ed Discussions to post announcements. Please bookmark these pages and them early and often.

Statement of Academic Honesty

At this point in your academic careers, we feel that it would be impolite to harp on cheating, so we won't. You are all adults and are expected to follow the university's code of academic conduct (**honor code (https://policylibrary.gatech.edu/student-life/academic-honor-code)**). Furthermore, at least some of you are researchers-in-training, and we expect that you understand proper attribution and the importance of intellectual honesty.

We should also point out that "proper attribution" does not absolve the writer of the "intellectual honesty" that comes from original writing. While it is definitely the case that copying text without attribution is considered plagiarism, it is also the case that copying too much text even with attribution betrays a lack of intellectual honesty. Too many quotes of more than, say, two sentences will be considered plagiarism and a terminal lack of academic originality. Do not overthink this issue, but do not under think it either.

Please note that unauthorized use of any previous semester course materials, such as tests, quizzes, homework, projects, videos, and any other coursework, is prohibited in this course.

In particular, you are not allowed to use old exams. Using these materials will be considered a direct violation of academic policy and will be dealt with according to the GT Academic Honor Code. Furthermore, we do not allow copies of my exams outside of this course. Just as you are not to use the previous material you are not to share current material with others either now or in the future. Our policy on that is strict. If you violate the policy in any shape, form, or fashion you will be dealt with according to the GT Academic Honor Code.

Due to the size and online nature of this course, there can sometimes be students who promote behavior and language that falls outside the Student Code of Conduct, especially after assignment grades are posted. Any personal attacks or unacceptable use of language towards other students or staff on any online platform will be reported with a zero-tolerance policy. When in doubt, follow the golden rule.

Readings and Lectures

The online lectures are meant to summarize the readings and stress the important points. You are expected to critically read any assigned material. Your active participation in the material, the lectures, and office hours are crucial in making the course successful. The office hours are mandatory and will be recorded for those who cannot attend live. We want to stress that the more you put into the material, the more you will get out. The full teaching staff is to assist you in learning and growing in the area of machine learning. Reach out and communicate often!

To help you to pace yourself, we have provided a nominal schedule (see the Syllabus link) that tells you when we would be covering material if we were meeting twice a week during the term. We recommend you try to keep that pace. Since the Summer term is reduced from 17 weeks to 13 weeks, we have modified the required material for the term with more optional meetings.

Further, the Summer term will offer additional advanced topics recorded by the instructor each week. The original material was recorded around 9 years ago and the additional supplemental material will help stay integrated into the current state-of-the-art for an introduction to machine learning. The instructor will post additional recordings of advanced topics pertaining to each module. These will help and correspond the the course assignments.

Scoring and Grading

Your final grade is determined by how you do on several components: a reading/writing quiz, hypothesis report practice, 3 unit quizzes, 3 comprehensive reports, and a cumulative final exam.

- **Reading/Writing Quiz.** In an effort to kickstart the course, I have included a quiz to help with reading and writing formal reports. At a graduate level in a terminal program, it is not enough to be able to apply advanced algorithms but understand why you make specific design choice and

comment on nuanced caveats on your results. This is a key aspect of the course you will not find in other online Machine Learning classes. The teaching staff prides ourselves in providing detailed feedback and we consistently hear that this skill can be immediately applied to many positions students currently hold. Therefore, I have included a lecture and quiz to help establish many of these skills you will practice in your reports over the course of the semester. You will have unlimited attempts to earn the points on the quiz until the due date. The lecture material and quiz will be available at the beginning of the first week and due at the end of the second week.

- **Hypothesis Report Practice.** To help supplement the Reading/Writing Quiz, I have provided an exercise on hypothesis interpretation and development. There is a large disparity in understanding term-to-term, so this should be great practice with feedback before the first assignment is due. The report practice will be available at the end of the first week and due at the end of the second week. This will help with direct applications to the reports and mimic the report submission process.

- **Unit Quizzes.** To help with understanding the course material in a more nuanced manner, there will be a unit quiz associated with each of the three units (SL, UL, and RL). Since each unit will be three weeks total, you will have a shorter quiz due by the end of the unit to help in your understanding of the material. I will open each quiz during the second week of the unit and it will be open until the end of the third week. There will be a variety of questions with both selection questions and calculations randomly picked from a test bank of questions. You will have a total of 60-minutes to complete the quiz with 3-attempts total over the window of the quiz. After each attempt, you will be able to see which questions you got correct and incorrect. You will be able to have one sheet of notes and a calculator, however this will be a closed-book, closed-internet quiz. There will be some practice questions posted on Ed Discussions before each quiz opens.

- **Reports.** There will be three scored assignment reports, Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL). They will be about programming and analysis. You will be required to use Overleaf and write your reports in LaTeX. You will use Private Georgia Tech Github repos to store your code. *Use of a personal Github is prohibited and will not be scored if submitted*. Generally, these assignments are designed to give you deeper insight into the material and to prepare you for the exam. The programming will be in the service of allowing you to run and discuss experiments, do analysis, and write your reports.

We do not provide a rubric for any of the assignments. These will not be shared to avoid gamification, which will happen no matter how much you argue (we have tried both ways in many classes in OMSCS). This is by pedological design to challenge you as there is no rubric in academia or the workforce. At a graduate level in a terminal program, you will be asked to develop skills you may not

have a lot of practice in. Many times, communication is a skill lacking by many engineers and practitioners - especially seasoned personnel. Instead of a rubric, you will be given an extensive FAQ for each assignment, open office hours for each assignment twice-a-week, a forum to ask questions monitored by the staff (Ed), example exemplary conference papers, low-stakes quizzes for understanding and practice, and extensive feedback provided by a grading TA when receiving your scores. That being said, we have an extensive rubric behind the scenes while we grade. There is a vetted system to verify and calibrate grading between staff members which is iterated and monitored by the instructor. I understand this will be the first time for many of you with such a unique challenge to potential weak spots. There are many people who have succeeded and this aspect of the course continues to be a highlight for alumni, especially those seeking new jobs or attempt new projects at work. My aim to to make this a manageable but challenging experience to help you grow as ML practitioner at the highest level.

When your reports are scored, you will receive feedback explaining your errors (and your successes!) in a fair level of detail. This feedback is for your benefit, both on this assignment and for future assignments. It is considered a part of your learning goals to internalize this feedback. This is one of many learning goals for this course, such as understanding how to analyze data or the differences between each algorithm or bias/variances in each of them.

Change for Summer 2025. *Reviewer Response*. In an effort to learn and grow assignment-to-assignment, we will provide a mechanism to edit and respond to your feedback. We will call this the Reviewer Response. You will have one week from the assignment grade being posted to edit and provide a two-page maximum response with both edits made and reviewer feedback. You will need to reasonably respond and edit your initial paper submission to improve your paper in good faith. Both the initial submission, revised submission, and two-page response will be needed for a proper Reviewer Response. If satisfied, you will receive half of the missed points back for the assignment. For example, if the initial grade was a 70/100, if everything is satisfied for the Reviewer Response, there will be 15 points added resulting in an 85/100. Further examples will be provided when the assignment grades are posted.

- **Final Exam.** There will be a digital, closed-book final exam at the end of the term. The final exam will also be administered via Canvas and Honorlock this term.

- **Final Letter Grade**. At the end of the term, I will calculate grades on a curve which will be up to my discursion as each term changes significantly. No other assignments will be curved. A majority of students will typically fall into the A and B ranges, however the exact cutoff changes term-to-term.

** Please note, in previous semesters we have administered a midterm. I have made a decision to remove this from the course as the exam was not conducive to a healthy student workload balance. This term, I am establishing 3 new quizzes and the Reviewer Response, so I do expect the averages of the assignments and overall grade to change slightly.*

Due Dates

All scored assignments are due by the time and date indicated. Here "time and date" means **Eastern Time (ET)**. Canvas does not currently support Anywhere On Earth, so this is the best alternative we can offer being at Georgia Tech. Please double check your settings and assignments for the exact due dates to mark your calendars appropriately. As a good check, you should go to settings on Canvas and set your time zone.

All assignments will be due at **11:59:00 PM ET** on the date due. However, since we will not be looking at the assignments until morning, you will have officially until 7:59:00 AM ET until the assignment is marked late. I understand that there are many circumstances that you may need an additional hour or two to complete the assignment. I will be asleep through the night and see no issue in giving the extra time. However, I need to heed a stern warning. You should use the 11:59PM timestamp as your internal deadline rather than the 7:59AM official cutoff. Staying up all night is a detriment to your mental health and may not be as conducive to constructive writing and testing. I know there is a colloquialism where nothing would get done unless for the last minute, however I do hope you all manage your time wisely. Please note the exact time for the submission as many situations may incur Murphy's Law. Allow a couple of minutes for the submission upload and check as it does take a few seconds on average to upload an assignment in Canvas.

For the reports only, we accept late assignments for a 20% off the top per-day penalty, a max of 5 days, or a 0 grade. The only exceptions to late report assignment penalties will require (1) a **note** from the appropriate authority and (2) **immediate notification** of the problem when it arises. With each notification, we need a proper explanation. We are here to work with you all; please do not try to abuse the system as it will not work. Also, we only accept approved late submissions one full week after the due date, including any exceptional cases. After that week, you will automatically get a 0 for that assignment. For cases that require longer than a week, we suggest dropping the course or asking for an incomplete semester. Please reach out to the instructor if you have further questions.

For the quizzes, since there are multiple attempts, each quiz will be due at the exact due date.

Further, I try to maintain our preset due dates that are posted on Canvas. The only exception when due dates will be changed is a widespread, outside force preventing use of required technology (e.g. a hurricane taking out power and Wi-Fi for a week). Otherwise, I will not change due dates. If due date needs to be changed, you will be given at minimum a 24-hour notice. Of course, anything can happen during the term and communication will be king. This will be the course policy.

As always, start early and do your best!

Numbers

Component	
Reading/Writing Quiz	5%
Hypothesis Quiz	5%
Unit Quizzes	15%
- SL Quiz	(5%)
- UL Quiz	(5%)
- RL Quiz	(5%)
Reports	45%
- SL Report	(15%)
- UL Report	(15%)
- RL Report	(15%)
Final	30%

Extra Credit

There are two opportunities to receive extra credit in this course. We intend to provide an additional comprehensive problem set that you will be able to turn in before the Final Exam. We will provide answers but will not score the set. The problem set will help prepare you for the exam. If all the problems are attempted and turned in, we will award 1% to your overall course percentage. Remember, you need to attempt everything, including adding explanations for your answers.

Everything turned in will be double-checked. These will be calculated on Canvas later in the term but before final grades are released.

Additionally, if there are significant contributions to the Ed Discussion board throughout the term, we will award 1% to your overall course percentage. "Significant contributions" will up to our discretion, however interacting with Ed Discussions will only benefit you over the course of the term. These additional points will be calculated on Canvas later in the term but before final grades are released.

There may additional extra credit opportunities throughout the term which will be communicated on Ed Discussions.

Office Hours and Other Channels

We love the assignments in this course. As you will discover they are wonderfully open-ended, much more so than many of you will be used to. It is therefore important that in addition to watching the lectures and comprehending the required readings that you attend office hours and regularly check Ed Discussions. We will record Office Hours but strongly suggest interacting both on Ed and in the live Office Hours. The Office Hours are mandatory whether you join live or watch the recording. These are only for your benefit. You should consider your participation in both required.

Disclaimer

I reserve the right to modify any of these plans as need be during the course of the class; however, we won't do anything capriciously, anything we do change won't be too drastic, and you'll be informed as far in advance as possible. There are many, many outside factors you will not be privy to, so please do not try and make assumptions. Looking forward to an amazing learning journey together!

Reading List

Required Text:

- **Tom Mitchell, Machine Learning. McGraw-Hill, 1997. (http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html)**

Optional Text:

- **Larry Wasserman, All of Statistics. Springer, 2010. (http://www.stat.cmu.edu/~larry/all-of-statistics/)** (Read Part 1 for an intro to Probability Theory)
- Richard Sutton and Andrew Barto, Reinforcement Learning: An introduction. (for Reinforcement Learning) (**Nov 5, 2017 version (http://incompleteideas.net/book/bookdraft2017nov5.pdf)**)

A List:

- Linear Algebra
  - **Linear Algebra and Eigenproblems** (https://github.com/pushkar/4641/raw/master/downloads/Eigenproblems.fm.pdf)
- Lessons (Sidebar: Ed Lessons)
  - ML is the ROX
    - Mitchell Ch 1
  - Decision Trees
    - Mitchell Ch 3
  - Regression and Classification
  - Neural Networks
    - Mitchell Ch 4
  - Instance-Based Learning
    - Mitchell Ch 8
  - Ensemble Learning
    - **Schapire's Introduction** (https://github.com/pushkar/4641/raw/master/downloads/boosting.ps)
    - **Jiri Matas and Jan Sochman's Slides** (https://github.com/pushkar/4641/raw/master/downloads/adaboost_matas.pdf)
  - Kernel Methods and SVMs
    - **An introduction to SVMs for data mining** (https://www.cc.gatech.edu/classes/AY2008/cs7641_spring/handouts/yor12-introsvm.pdf)
    - **Christopher Burges tutorial on SVMs for pattern recognition** (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/svmtutorial.pdf)
    - **Scholkopf's NIPS tutorial slides on SVMs and kernel methods** (https://github.com/pushkar/4641/raw/master/downloads/svm-scholkopf.ps)
  - Computational Learning Theory
    - Mitchell Ch 7
  - VC Dimensions
    - Mitchell Ch 7
  - Bayesian Learning
    - Mitchell Ch 6
  - Bayesian Inference
  - Randomized Optimization
    - Mitchell Ch 9
    - **No Free Lunch Theorem** (https://ml-cs7641.s3.us-east-1.amazonaws.com/nfl-optimization-explanation.pdf)
  - Clustering
    - Mitchell Ch 6
    - **Intuitive Explanation of EM** (https://ml-cs7641.s3.us-east-1.amazonaws.com/em-intuitive-explanation.pdf)

        - Statical View of EM (https://github.com/pushkar/4641/raw/master/downloads/em.pdf)
        - Jon Kleinberg's Impossibility Theorem for Clustering (https://www.cs.cornell.edu/home/kleinber/nips15.pdf)
    - Feature Selection
        - ICA: Algorithms and Applications (https://ml-cs7641.s3.us-east-1.amazonaws.com/ica-algorithms-and-applications.pdf)
        - Restructuring High Dimensional Data by Charles and Paul Viola (https://www.cc.gatech.edu/~isbell/papers/isbell-ica-nips-1999.pdf)
    - Feature Transformation
    - Information Theory
        - Charles Isbell's Note on Information Theory (https://www.cc.gatech.edu/~isbell/tutorials/InfoTheory.fm.pdf)
        - An Introduction to Information Theory and Entropy (https://github.com/pushkar/4641/raw/master/downloads/gentle_intro_to_information_theory.pdf)
    - Markov Decision Processes
    - Reinforcement Learning
        - Mitchell Ch 13
        - Richard Sutton and Andrew Barto, Reinforcement Learning: An introduction. MIT Press, 1998. (http://incompleteideas.net/book/bookdraft2017nov5.pdf)
        - Reinforcement Learning: A Survey (https://github.com/pushkar/4641/raw/master/downloads/kaelbling96reinforcement.pdf)
    - Game Theory
        - Andrew Moore's slides (http://www.cs.cmu.edu/~awm/tutorials.html) (http://www.cs.cmu.edu/~awm/tutorials.html)
    - Outro

Software

- WEKA (https://ml.cms.waikato.ac.nz/weka) Machine learning software in JAVA that you can use for your projects
- Data Mining with Weka (https://weka.waikato.ac.nz/) A MOOC Course
- ABAGAIL (https://github.com/pushkar/ABAGAIL) Machine learning software in JAVA. This is hosted on my github, so you can contribute too
- scikit-learn (http://scikit-learn.org/stable/) A popular python library for supervised and unsupervised learning algorithms
- pybrain (http://pybrain.org/) A popular python library for artifical neural networks
- Murphy's MDP Toolbox for Matlab (http://www.cs.ubc.ca/~murphyk/Software/MDP/mdp.html)
- MATLAB Clustering Package (http://www.cc.gatech.edu/~dellaert/FrankDellaert/Software.html) By Frank Dellaert (http://www.cc.gatech.edu/~dellaert/FrankDellaert/Frank_Dellaert/Frank_Dellaert.html)

Version Control

- 05/12/2025: TJL updated course policies for changes in Summer semester. Course was published for Summer semester.