CS4641 Syllabus

Course Information

Course Prefix and Number: CS4641
Course Name: Machine Learning
Instructor: Borela Valente, Rodrigo

Course Description

This course introduces techniques in machine learning with an emphasis on algorithms and their applications to real-world data. We will investigate the following question: how to computationally extract useful knowledge from data for decision making and task support? The course will also cover briefly Ethics in Machine Learning and Secure Computing.

Course Learning Outcomes

Upon successful completion of this course, you should be able to implement major machine learning algorithms, apply them to real-world data analysis and critically evaluate their performance given their assumptions, task objectives, and data distribution. We will focus on methods organized into three main course objectives:

1. Basic math for data science and machine learning
   - Linear algebra
   - Probability and statistics
   - Information theory
   - Optimization
2. Unsupervised machine learning for data exploration
   - Clustering analysis
   - Dimensionality reduction
   - Kernel density estimation
3. Supervised learning for predictive data analysis
   - Tree-based models
   - Support vector machines
   - Linear classification and regression
   - Neural networks

In addition to the specified technical content, this class includes the following learning objectives:

- Structuring a task into a machine learning workflow
- Collaborating effectively on team projects
- Conducting peer evaluation in a constructive format
- Communicating technical content in a concise and effective manner

Required Course Materials

Our primary text for the course is Bishop's Pattern Recognition and Machine Learning. Additionally, we will use a selection of research papers, articles, class notes, and tutorials, which will be provided to you.

Grading Policy

This course uses a fixed grading scale. Grades are not curved. Final grades are calculated to the nearest tenth. Scores are not rounded up. For example, a final score of 89.9 will be recorded as a B, not an A. Scores are calculated using the following assignment category percentages:

Assignments (4):		**40%**
HW 1	10%	
HW 2	10%	
HW 3	10%	
HW 4	10%	
Project:		**30%**
Proposal		5%
Midterm Checkpoint		10%
Final report		15%
Quizzes:	**15%**	
Syllabus Quiz:	**1%**	
Class participation:	**4%**	
In-Class Assessment:		**10%**

Letter Grade Scale:

A	90.0 and above
B	80.0 – 89.9
C	70.0 – 79.9
D	60.0 – 69.9
F	Below 60.0

Description of Graded Components

- **Assignments.** There will be four assignments, each carrying 10% of the course grade. Each one is designed to improve and test your understanding of the material. Assignments will have both programming and written analysis components. The topics are subject to change and cover the entire breadth of the technical learning outcomes.
- **Project.** The project is worth 30% of your grade and will focus on applying machine learning algorithms and methods to real-world datasets to create meaningful insights and useful predictions. You will create a website with GT GitHub Pages for your

project which will be used to publish three deliverables: a proposal (5%), a midterm checkpoint (10%), and a final report (15%).
    - CS 4641 students are required to use supervised learning. It is highly encouraged to use unsupervised learning methods as well.
    - CS 7641 students are required to use both unsupervised and supervised learning.
- **Quizzes.** The total number of quizzes for the semester is listed on the official class schedule (excluding the syllabus quiz and warm-up quiz 0). The weight of each individual quiz is calculated by dividing the total category weight (15%) by the total number of quizzes administered. All quizzes are mandatory. Quizzes are open notes and will be proctored using Honorlock. The topic of each quiz will coincide very closely with the content covered in class on that week. Quizzes will have a duration of seven-minutes for Undergrad students and six-minutes for Grad students. Each quiz will have five multiple choice questions. Quizzes measure your understanding of the topics and they will be mostly conceptual questions.
- **Syllabus quiz.** This quiz will test you on the course deadlines and rules. You can simply obtain 1% if you carefully read all the contents of the website and our class rules.
- **Class participation.** Edstem has statistics which give us many measurements regarding how much a student has been involved on Edstem's activities such as viewing posts, answering questions, asking questions and so on. We use this to account for your Class Participation score. We also will add class attendance to this score. At the end of the semester, we will define a minimum and maximum number of involvement considering all the students and your grade will be defined based on that.
- **In-class assessment.** We will hold one in-class assessment on Canvas. The exam will focus on conceptual questions, and all students are required to take it.

Attendance Policy

Our class will be offered on campus for both Undergrad (4641) and Grad (7641). Lectures might be recorded IF class has the recording system. Any class that I am able to record [which sometimes does not work even if we have the recording system in place], I will make it available to all students (both undergrad and grad) by the end of the day. The attendance is required for both undergrad and grad. Also, the class attendance will be counted toward your class participation at the end of semester.

Academic and Research Honesty/Integrity Statement

All learners are expected to know and abide by the Georgia Tech Academic Honor Code and the student Code of Conduct.

Ethical behavior is extremely important in all facets of life.

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

2. You may discuss high-level ideas with other students at the "whiteboard" level (e.g., how cross validation works, use hashmap instead of array) and review any relevant materials online. However, each student must write up and submit his or her own answers.
3. You must not put your code on public domain (e.g., public GitHub), because a (future) student could copy your code. That student obviously violates the honor code, and you may also be implicated.
4. All incidents of suspected dishonesty, plagiarism, or violations of the Georgia Tech Honor Code will be subject to the institute's Academic Integrity procedures (e.g., reported to and directly handled by the Office of Student Integrity (OSI)). Consequences can be severe, e.g., academic probation or dismissal, grade penalties, a 0 grade for assignments concerned, and prohibition from withdrawing from the class.

Artificial Intelligence Use Policy

We are using the AI assistant policy developed by David Joyner and shared by other classes at Georgia Tech (CS 7643 Deep Learning). The summary is that you should treat your AI source like a human source, with all accompanying plagiarism implications:

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.

Core IMPACTS

Not applicable.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, 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 contact the Office of Disability Services

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. This summarizes my expectations for you and what you can expect from me. 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. Expectations of Advisors and Advisees