

# CS 4641: Machine Learning

Aaron Hillegass

Summer 2024

E-mail: hillegass@gatech.edu

Office Hours: TBD

Office: TBD

Class Hours: TBD

Class Room: TBD

---

## Course Description

*Machine Learning* 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!

## Required Materials

*The Elements of Statistical Learning, 2nd Edition* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The textbook is a good resource for the things we will cover in class. While I believe I do a good job of introducing the material through lecture, the book provides a secondary description of topics that can reinforce understanding.

## Prerequisites

- Basic understanding of probability and statistics
- Working knowledge of linear algebra
- Basic programming experience in Python

## Topics Covered

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. We will focus on machine learning methods, which are organized into three parts:

**Basic math for data science and machine learning**

- Linear algebra
- Information theory
- Optimization

**Bayesian Data Analysis**

- Bayes' Law
- Prediction and confidence
- Naive Bayesian Classifiers

**Unsupervised machine learning for data exploration**

- Clustering analysis
- Dimensionality reduction
- Kernel density estimation

**Supervised learning for predictive data analysis**

- Tree-based models
- Support vector machines
- Linear classification and regression
- Neural networks

**Grading****Assignments (60%)**

There will be six assignments. Each one is designed to improve and test your understanding of the materials. Assignments will have both programming and written analysis components.

The reason we do not call them projects is because our class has a project as well. Consider each assignment as one individual big project. Assignments take time to finish them. **YOU NEED TO START WORKING ON ASSIGNMENTS AS SOON AS THEY ARE OUT.** Visit this course's Canvas and GradeScope for the assignment documents. See the schedule table above for deliverable due dates. (Topics are subject to change):

- HW1: Linear Algebra, Probability and Statistics, Maximum Likelihood Estimation, Optimization, Information Theory
- HW2: Bayesian Data Analysis and Bayesian Classifiers
- HW3: K-Means, Expectation Maximization, Gaussian Mixture Model, Clustering Evaluation

- HW4: Singular Value Decomposition, Principal Component Analysis, Linear Regression, Regularization
- HW5: Decision Trees, Random Forest, Support Vector Machine
- HW6: Neural Networks, CNN

You are required to use LaTeX for reports. (Watch the tutorial created by our own team and OverLeaf Latex Example in the Video) Handwritten solutions WILL NOT BE ACCEPTED. You can easily export your Jupyter Notebook to a Python file and import that to your desired python IDE to debug your code for assignments.

All assignments will have a 48-hour Penalized Acceptance Period after the assignment due date where we will accept the assignment with a penalty. Submissions turned in during this period will have its grade reduced by a linear percentage deduction commensurate to how much of the 48-hour penalized acceptance period is used. The penalty is capped to a 20% deduction of the submission grade. Assignments received after the 48-hour Penalized Acceptance Period will receive zero credit.

This deduction applies separately to each component of the assignment that is submitted separately on Gradescope. Consider an example where a student submits the coding portion of their assignment on time but submits the written portion 12 hours after the deadline. In this case, the coding portion will not be penalized, and the written portion will be subject to a 5% deduction.

### **Final (20%)**

We will have a final to test your knowledge. If you have understood all the homework assignments, the final should not be challenging.

### **Quizzes (15%)**

There will be 14 quizzes throughout the semester. We will consider your top 10 quizzes' scores. Each quiz will have 1.50% of your final score.

All quizzes are mandatory to be taken even if they do not count toward your final grade. If you miss a quiz, we will deduct your score from your Class Participation score. Let's say you miss taking one quiz; we will reduce 1.5% from your class participation score. If you miss 3 quizzes, you will lose all your class participation score, which is 4%, and we will NOT go beyond that. (if you miss 4 quizzes, we only deduct 4% from your class participation score, not 6%). Your class participation score can be zero at its lowest, and it won't go to a negative number.

The topic of each quiz will coincide roughly 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. All quizzes will be released on Thursdays weekly at 6:00 pm EST and the deadlines will be on Fridays 23:59 EST. To check deadlines for Quizzes, ensure to check the class schedule table. Any possible changes on quizzes dates will be reflected on our course schedule page. Please make sure to check our class website before taking the quiz.

Quizzes have 48 hours "grace period" without any penalty:

If a student decides to make a submission during the grace period, they are responsible for all issues associated with that submission and Course staff support is not guaranteed during the grace period; We provide help only when available. You do not need to ask before using the grace period. Quizzes measure your understanding of the topics and they will be mostly conceptual questions.

Quizzes' answers will be released as soon as all our students took them including our ODS students. Please do not ask any questions about a quiz that you just take on Edstem before we release the answers.

Quizzes questions are selected randomly from our question bank, which means that students will not receive the same questions for their quiz.

### **Syllabus quiz (1%)**

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. We will ask questions like how many quizzes we have in the class; which days of the week we have most of our deadlines? is Participation required in the course? ...

### **Class Participation (4%)**

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.

We will RELEASE the class participation score on the last day of the class when we have all the score for projects, quizzes and assignments. If you ask us what is my participation score before the last day of the class; we will say we do not know. So please be patient.

### **Bonus Points (up to 8%)**

About bonus points: Bonus points will be counted to always be beneficial for your final grade. More information on bonus points for assignments will be provided as the semester progresses. If it becomes necessary to curve grades, bonus points will be applied after curving, not before.

Undergrad and Grad: You can obtain up to 5% bonus points by answering the challenging questions we may have in some of the HWs.

Undergrad: You will notice that we have bonus points for all the hws, where grad students are required to answer those questions, but it will be optional for undergrad students. You will receive up to 3%, if you answer those questions. Note that these are different than the challenging questions. Challenging questions are bonus for both grad and undergrad.

How does it work? For example, hw 1 may have 30 bonus points, hw 2 may have 20 bonus points and so on. If you receive all the bonus points for all your hws, we will add 5% to your final grade.

If you are an undergrad and you answer all the challenging and Grad students questions, you will receive 8%.

Note and Example: There's a cap to how much extra credit you can get, so it is (bonus points earned)/(total bonus points available throughout the entire semester). Let's say by the end of the semester there was a total of 100 bonus for all points (100 is just a number we are randomly choosing here) between hw1, hw2, hw3, hw4, and you earned 20 bonus for all points for the whole semester, then at the end of the semester your grade will be bumped up by  $5\% * 20/100 = 1\%$  from the bonus for all points. The calculation is similar for the 3% bonus for undergrad points.

## Honor Code

All students are expected to follow the Georgia Tech Academic Honor Code. Because of the large size of our class, if we observe any (even small) similarity/plagiarisms detected by GradeScope or our TAs, WE WILL DIRECTLY REPORT ALL CASES TO OSI, which may unfortunately lead to a very harsh outcome.

You are NOT allowed to share or discuss ANY assignment codes, information or answers with other students. Edstem is the best place to have discussion regarding assignments and course topics. Discussions can be on a whiteboard level with other students such as high level conceptual questions (i.e. what is independency in Naive Bayes model)