

# CS534 - Machine Learning

This repository contains course materials for [CS 534 - Machine Learning](#) provided in Fall 2020 at [Emory University](#).

Course syllabus, slides, and homework are largely borrowed/copied from Professor Joyce Ho's [CS534 class](#).

## Overview

[“Software is eating the world,”](#) said Marc Andreessen in 2011. By then, he was referring to companies like Amazon, Netflix, Apple, and Google who are massively transforming the traditional industries with the power of the internet and software. It was true in some sense, as apparent in [their enormous growth in market capitalization](#).

How did these companies become so successful? There may be many reasons, one may say. Unfortunately, the answer to this question is beyond the scope of this course. However, I find that one commonality of these companies remarkably stands out: “their insatiable appetite for data”. Six years later, Marc Andreessen’s famous quote was wittily supplemented with [“..., but AI is going to eat software”](#) by Jensen Huang.

We are living in an exciting time. Artificial intelligence and its key engine, machine learning, are impacting [every aspect of our lives](#). From shopping groceries to treating diseases, people are experimenting and adopting machine learning algorithms, achieving the level of productivity that was never possible before.

In this course, students will learn some of the fundamental concepts, theories, and algorithms of machine learning. Students will have a chance to implement and test some of the time-tested machine learning algorithms. At the end of the class, students will be able to 1) identify and formulate machine learning problems, 2) implement and test appropriate machine learning

algorithms, and 3) understand and articulate the limitations of the approaches.

*Prerequisites:*

- Undergraduate-level linear algebra
- Undergraduate-level probability
- Undergraduate-level algorithms
- Programming ability in Python

## Logistics

- Programming Language
  - Python is the primary language for the course
- Textbook
  1. [The Elements of Statistical Learning](#) by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
  2. [A Course in Machine Learning](#) by Hal Daumé III
- Office Hours
  - 2:30 pm ET - 3:30 pm ET on Monday at Zoom (the link will be available on Slack)
  - 3:00 pm ET - 4:00 pm ET on Wednesday at Zoom (the link will be available on Slack)
- TA Office Hours
  - Han Xie (email: {firstname}.{lastname}@emory.edu)
    - Thursday 2:30 pm to 3:30 pm at Zoom (the link will be available on Slack)
- 100% Online Class
  - This course is offered 100% online using these tools:
    - [Canvas](#): Canvas will be used to track assignments, grades, and important announcements.
    - [Zoom](#): Zoom will be used for the synchronous online lectures. Please check the Zoom link on the Canvas course page.
    - [Slack](#): Slack will be used as the main communication tool for the class e.g. questions regarding homework, lectures, etc. The instructor and TAs will try to respond to any Slack messages within 24 hours during weekdays, 48 hours during weekends.
  - Please show up in the scheduled Zoom sessions
    - This course intends to teach you how to think with machine learning, not the knowledge of machine learning.
    - Attendance is up to the individual student. Students should be responsible for keeping their course schedules and deadlines.

- Lastly, please be mindful of your personal safety, and avoid including personal information, such as phone numbers or addresses, in discussion forums. All online communications should be transmitted with the intent to inform, inspire, etc. - not to offend or breach personal privacy. Never use private information about other individuals and be sensitive to the information you share about yourself.

## Schedule

**WEEKLY OBJECTIVE:** Most of the time, each lecture focuses on a different machine learning algorithm, starting from very foundational yet simple ones to more complex and sophisticated ones. Students will be able to understand 1) how such machine learning algorithms are motivated/developed, 2) when such models will be useful and avoided, and 3) how to implement and modify such models to deal with more custom settings.

**GUEST LECTURES:** We are excited to have a stellar array of guest lecturers:

- [Tian Su, PhD](#), Director of Machine Learning at Walmart
- [Hyunjoon Jung, PhD](#), Director of Personalization Data Science at Nike
- [Jenny Wang](#), VP of Episode Development at Signify Health
- [Yen Sia Low, PhD](#), ML Architect at Rally Health (Optum Digital)

Date	Topic	Reference	Assignment
08/19	<a href="#">Introduction</a>	.	<a href="#">Homework #0</a>
08/24	<a href="#">Prerequisite - Statistics</a>	.	.
08/26	<a href="#">Linear Model - Regression (1)</a>	ESL Chapter 3	<a href="#">Homework #1</a>
08/31	<a href="#">Linear Model - Regression (2)</a>	ESL Chapter 3	.
09/02	<a href="#">Linear Model - Classification</a>	ESL Chapter 4	.
09/07	<a href="#">Linear Model - Other Distributions</a>	ESL Chapter 4 and <a href="#">GLMNET</a>	.

09/09	Evaluation Metrics	ESL Chapter 7	Homework #2
09/14	Evaluation Strategies	ESL Chapter 7	.
09/16	Bias-Variance Tradeoff (1)	Chapter 20-27 of <a href="#">Machine Learning Yearning</a> (Optional)	.
09/21	Bias-Variance Tradeoff (2)	Chapter 20-27 of <a href="#">Machine Learning Yearning</a> (Optional)	.
09/23	Decision Tree	.	Homework #3
09/28	Boosting	.	.
09/30	Boosting	.	.
10/05	Bagging	.	.
10/07	Random Forests	.	.
10/12	Nearest Neighbors	Chapter 3 of <a href="#">Mining of Massive Dataset</a> (Optional)	.
10/14	Dimensionality Reduction (1)	.	Homework #4
10/19	Dimensionality Reduction (2)	.	.
10/21	Convex Optimization	.	.
10/26	<i>Project Elevator Pitch</i>	.	.
10/28	Support Vector Machine	.	.
11/02	Neural Networks	.	Homework #5
11/04	Guest Lecture (Yen Sia	.	.

	Low)		
11/09	Guest Lecture (TBD)	.	.
11/11	Guest Lecture (TBD)	.	.
11/16	<i>Presentations</i>	.	.
11/18	<i>Presentations</i>	.	.
11/23	<i>Presentations</i>	.	.

NOTE: Course materials with the `md` extension are [Markdown](#) files. To read the Markdown files, you can use various “free” Markdown viewer applications out there e.g:

- [Markdown Viewer Chrome Add-on](#)
- [Slide-style View using Remark](#)
- [Slide-style View using Remark \(web service\)](#)

## Grading

- Homeworks: 35%
- Midterm: 15%
- Project: 40%
- Participation: 10%

## Project

The goal of the final project is to:

1. identify and formulate a machine learning problem,
2. come up with appropriate solutions,
3. implement the solutions,
4. analyze the results and limitations, and
5. suggest any algorithmic/engineering improvements.

You are free to find any datasets and problems of your research interest.

You are encouraged to work in groups of 2-3 people for the project. Teams are required to hand

in a project proposal, a final report, and presentation slides of their work.

Note that the final project contributes to 40% of your final grade. The final report should include a section that details each member's contribution to the project.

### *Grading for the Final Project:*

- Proposal (due 10/30): 15%
- Elevator Pitch (due 11/04): 10%
  - Elevator pitch needs to address 1) why the problem is important or interesting, 2) what are existing solutions, and 3) how you plan to solve the problem differently
  - Teams are required to present this in **2 minutes** - that's why it's called the elevator pitch.
- Presentation (due 11/27): 25%
- Report (due 12/13): 50%

## Policy

- Assignments
  - Homeworks are due electronically on Canvas at 11:59 pm ET.
  - Late submissions are accepted with the following scoring rule:
    - $\text{Score} = 0.7^{\{\text{number\_of\_late\_days}\}} \times \text{original\_score}$
    - The number of late days counts from 1 right after the deadline.
    - Note that  $0.7^{14} = 0.00678$ . This means homework submitted after 2 weeks will get zero points regardless of their original scores.
  - No partial submission is allowed. You either submit the full homework on-time or not.
  - The latest submission will count as the submission.
    - Even if your prior submissions have higher scores, the scores from the later submissions will overwrite.
- Midterm Exam
  - The midterm exam is open-book, open-notes, but no electronic devices.
  - The exam must be taken at the required time.
  - If you need to reschedule the midterm exam, you must request at least a week before the exam date to be considered.

## Honor Code

All classwork is governed by the College Honor Code, Emory Laney Graduate School, and

Computer Science Departmental Policy. It is acceptable and encouraged to discuss homework with other students. However, this should be noted on your submitted homework and all code and writeup must be written by yourself. Any code and writeup that is found to be similar are grounds for an honor code investigation by the Director of Graduate Studies, Laney Graduate School, and the honor council.

---

**cs534** is maintained by [yubin-park](#).

This page was generated by [GitHub Pages](#).