PRINCIPLES OF
MACHINE LEARNING

CS 363M, SPRING 2024

TTH 9:30-11:00, GDC 5.302 (51110)
TTH 11:00-12:30, GDC 5.302 (51115)
TTH 2:00-3:30, UTC 3.122 (51120)

PROFESSOR

Angie Beasley
angiebeasley@utexas.edu
GDC 6.314

TAs

Anshul Moondra - amoondra@utexas.edu
Jiaru Song - jiarus@cs.utexas.edu
Nilesh Gupta - nilesh@cs.utexas.edu
Son Nguyen - sonnv77@utexas.edu

See Canvas for up to date office hours.

COURSE DESCRIPTION

Machine learning is a subset of artificial intelligence that enables software applications to improve their performance over time, as they acquire more information, data, or experience.

Most of machine learning is done by identifying patterns in the input, whether that input is a table of data with many rows, or a large set of images, or repeated trials of playing a game, etc. Over time, given more input, the machine better learns the patterns, and therefore makes better decisions.

In this class, you will learn machine learning algorithms to find patterns in large input sets. We will cover the three major machine learning techniques: supervised, unsupervised, and reinforcement learning. You will learn about regression, classification, clustering, anomaly detection, recommender systems, and reinforcement learning. You will use Python and Jupyter Notebooks, two industry-standard tools for machine learning.

In this "big data" world, where data is constantly collected on everything and everyone, machine learning is drastically changing the way in which many important decisions are made. Now more than ever, it is critical to understand how these algorithms come to their conclusions, the correct ways to interpret and apply their results, and the social impacts of their use.

Only one of the following may be counted: CS 363D, 363H, 363M, 378 (Topic: Intro to Data Mining), 378H (Topic: Intro to Data Mining: Honors). Prerequisite: The following coursework with a grade of at least C-: Computer Science 429 or 429H; Mathematics 362K or Statistics and Data Sciences 321; and Mathematics 340L, 341, or Statistics and Data Sciences 329C.

TEXTBOOK

Introduction to Data Mining
--> Second Edition <--
by Pang-Ning Tan, Michael Steinbach, Vipin Kumar

The textbook is optional. It is for your reference.

CLASS

This class will be held in person. You are welcome to attend any of the three class times, regardless of which one you are actually registered for, EXCEPT on exam days and on your assigned social impact day, when you must attend the class time for which you are officially registered.

If you feel sick, please do not come to class in person!

If you feel sick, you may stay home and watch the Lectures Online recording of the class. Class recordings will be posted to Lectures Online one day after class.

Devices

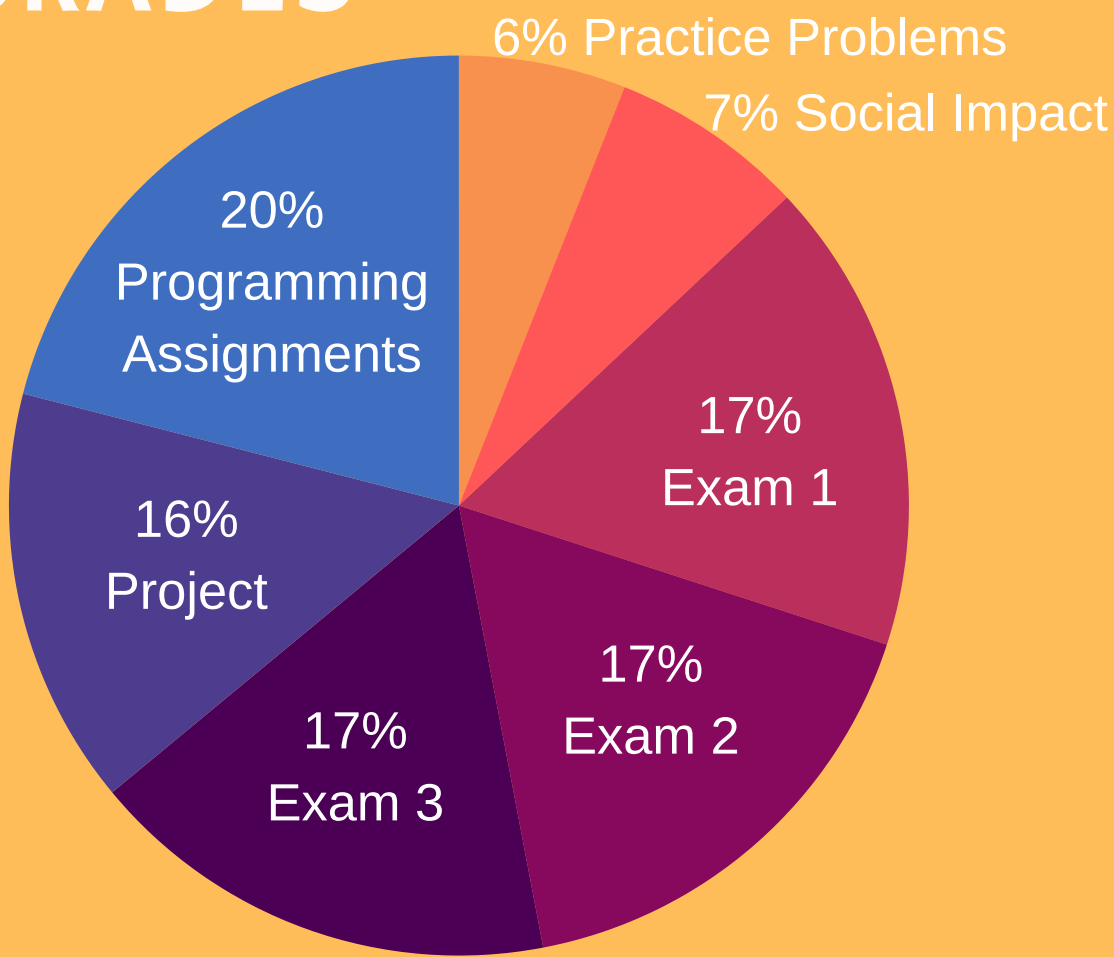We are all challenged by the many ways in which digital devices are designed to steal our attention. Additionally, there is a significant amount of research demonstrating that when students engage in off-task digital behaviors in class, it hurts the learning of peers sitting around them.

For these reasons, I do not allow the use of laptops during class, and I ask that you put your phones away for the 1 hour and 15 mins we have together to focus on learning.

I do allow flat tablet devices for note-taking, but I highly recommend that you turn off all notifications and close all other distractions during class time so you are not tempted to lose focus or do other things. If these devices become a problem, the policy will be revised.

GRADES

- 6% Practice Problems
- 7% Social Impact
- 20% Programming Assignments
- 17% Exam 1
- 16% Project
- 17% Exam 2
- 17% Exam 3

A	>=93
A-	< 93
B+	< 90
B	< 87
B-	< 83
C+	< 80
C	< 77
C-	< 73
D+	< 70
D	< 67
D-	< 63
F	< 60

All numbers are absolute and will not be rounded up or down at any stage.

PROGRAMMING ASSIGNMENTS

There will be 6 programming assignments. The first assignment must be completed individually. Assignments 2-6 may be worked individually or in pairs. If you work in pairs you are expected to use the proper pair programming method (i.e. working together in real-time). Any deviation from proper pair programming (i.e. divide-and-conquer) will be considered academic dishonesty. Programming assignments will be completed using Python 3 and either Jupyter Notebooks or Google Colab.

LATE ASSIGNMENTS

You will have 3 late days in 1-day units (that is, 1 minute to 24 hours late = 1 late day) to use throughout the semester. You may divide your late days across the programming assignments in any way you wish. Once you have used all of your late days, late assignments will no longer be accepted.

In the case of pair programming, each member of the pair must have enough late days to cover the late submission. So if the pair submits their code 2 days late, each member must have two late days remaining to use and each member will lose two late days.

To use late days, you only need to submit the assignment. You do not need to email the instructor or the TA, you do not need to indicate that you are using late days. Your late days will be deducted according to when your assignment is submitted. If you submit a late assignment without enough late days to support it, you will receive a zero for that assignment.

Contact me if there are extenuating circumstances or if you get sick.

SOCIAL IMPACT

As more and more important decisions become automated via machine learning, it is increasingly imperative that we consider the impacts of this on us, and on each other, as humans.

You will be split into groups of 4 and assigned a real-world use-case of machine learning to learn about and present to the class. This will be a short (10 min) presentation, where your group addresses the social impact of the scenario.

Additionally, each class, you will write a short response (1-3 sentences) to the social impact presentation and submit it. **You may drop 3 of these.**

PRACTICE PROBLEMS

Throughout the semester, there will be graded practice problems. Some will be graded for correctness, and some will be graded only for completion. **You may drop your 2 lowest of these**.

PROJECT

Your project will be to perform machine learning on a dataset of your choosing using the techniques taught in class. This project gives you the opportunity to practice the full machine learning process on a real-world problem. You will work in groups of 4 people. The results of your project will be presented in a Jupyter Notebook.
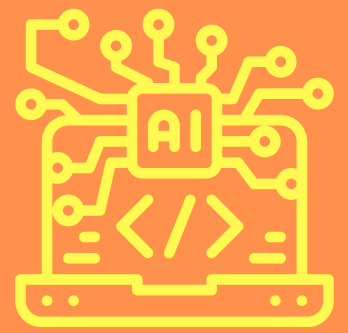
PROJECT SCHEDULE

2/27	Group selection deadline
3/8	Dataset selection deadline
4/21	Final project due

REGRADE REQUESTS

All grades will be posted on Canvas.
You have **one week** from the date the grade is posted to dispute your grade. The TAs will be grading the assignments. First contact the TAs and see if you can resolve your differences. If you can not resolve your differences, you may contact me to explain the situation. We will not entertain any grade disputes after one week.

COURSE SCHEDULE

Subject to change at instructor's discretion.

DATA PREP

1/16	Introduction
1/18	Data Prep, Data Exploration
1/23	Feature Engineering, Dimensionality

SUPERVISED LEARNING

1/25	Classification & Decision Trees [Ch 3.1-3.3]
1/30	Decision Trees (cont.)
2/1	Linear Regression
2/6	Overfitting & Cross-Validation [Ch 3.4-3.9]
2/8	Nearest Neighbor [Ch 4.3]
2/13	Naive Bayes [Ch 4.4]
2/15	Evaluating Classifiers [Ch 4.11]
2/20	Ensemble Methods [Ch 4.10]
2/22	SVMs [Ch 4.9]; Non-linear regression
2/27	EXAM 1
2/29	Neural Nets [Ch 4.7]
3/5	Neural Nets (cont.)

UNSUPERVISED LEARNING

3/7	Clustering & K-means [Ch 7.1-7.2]
3/12-14	Spring Break
3/19	Density-Based Clustering [Ch 7.4]
3/21	Hierarchical Clustering [Ch 7.3]
3/26	Cluster Validation [Ch 7.5]
3/28	Anomaly Detection [Ch 9]
4/2	EXAM 2

REINFORCEMENT LEARNING & RECOMMENDER SYSTEMS

4/4	Guest Speaker
4/9	Reinforcement Learning
4/11	Apriori [Ch 5.1-5.5]
4/16	Rule Gen & Evaluation [5.3-5.4, 5.7-5.8]
4/18	Sequence Mining
4/21	Project Due
4/23	Collaborative Filtering
4/25	Semester Wrap-Up
TBD	EXAM 3

ACADEMIC INTEGRITY

Each student in the course is expected to abide by
the University of Texas Honor Code:
"As a student of The University of Texas at Austin, I shall abide
by the core values of the University and
uphold academic integrity."

This means that **work you produce on assignments and
exams must be all your own work**, or all your own group's
work. I will make it clear for each assignment whether group
work is allowed or not.

Academic misconduct may consist of any of the following:
Copying; Using unauthorized materials or sources;
Providing or receiving unauthorized aid or assistance;
Substituting for another person or allowing another person to
substitute for you; Collusion or unauthorized collaboration;
Falsification or fabrication of information, data, citations, or
work; Plagiarism or representing another's work as your own;
Failure to to comply with course requirements or instructions;
Or any action that compromises the integrity
or fairness of your work.

The use of generative AI tools (ChatGPT, Copilot, or similar)
is prohibited in this class. Therefore, any use of AI tools for
work in this class will be considered academic misconduct,
since the submitted work would not be your own.

Penalties for academic misconduct may include grade
reduction or course failure, as well as being reported to the
Academic Integrity Office in the Dean of Students Office.

→ If you submit code or work that is not your
own, you will be guilty of academic
dishonesty and subject to academic
disciplinary action, including failure of the
course and being reported to the
Dean of Students.

UNIVERSITY RESOURCES

The Counseling and Mental Health Center (CMHC) provides
counseling, psychiatric, consultation, and prevention services:
http://cmhc.utexas.edu/

Student Emergency Services (SES) can be contacted in cases of
family emergency/death in the family, medical emergencies, fire or
natural disasters, academic difficulties due to crisis or emergency
situations, interpersonal violence (stalking, harassment, physical
and/or sexual assault):
http://deanofstudents.utexas.edu/emergency/

If you have concerns about the safety or behavior of fellow students,
TAs or professors, call BCAL (the Behavior Concerns Advice Line):
512-232-5050. Your call can be anonymous. If something doesn't
feel right – it probably isn't. Trust your instincts and share your
concerns.

STUDENT SUPPORT & ACCOMODATIONS

I am committed to creating an accessible and inclusive learning environment for everyone. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. Please contact me as soon as possible if the material being presented in class is not accessible to you, if any of the physical space is difficult for you, or to discuss any other accommodations you may need.

If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability & Access (D&A): https://diversity.utexas.edu/disability/.

SHARING COURSE MATERIALS IS STRICTLY PROHIBITED

Sharing of Course Materials is Prohibited. No materials used in this class, including, but not limited to, videos, assessments, quizzes, exams, papers, projects, homework assignments, in-class materials, lecture hand-outs, review sheets, and problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission.

Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure of the course.

Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA federal law (20 U.S.C. § 1232g; 34 CFR Part 99). Class recordings may not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Students may not record all or part of class, livestream all or part of class, or make/distribute screen captures, without advanced written consent of the instructor. Classes may be recorded by the instructor. Students may use instructor's recordings for their own studying and notetaking. Instructor's recordings are not authorized to be shared with anyone without the prior written approval of the instructor. Failure to comply with requirements regarding recordings will result in a disciplinary referral to the Dean of Students Office and may result in disciplinary action.

Notice of Copyright: Materials in this course—unless otherwise indicated—are protected by United States copyright law (Title 17, U.S. Code). No material from this course may be copied, reproduced, re-published, uploaded, posted, transmitted, or distributed in any way without the permission of the original copyright holder.