PRINCIPLES OF
MACHINE LEARNING

CS 363M, SUMMER 2023
ONLINE/ASYNCHRONOUS (85800)

PROFESSOR

Angie Beasley
angiebeasley@utexas.edu
GDC 6.314

TA

Anshul Moondra
amoondra@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 decisions.

Computer Science 363D and 378 (Topic: Introduction to Data Mining) may not both be counted. Prerequisites: The following coursework with a grade of at least C- : Computer Science 429 (or 310) or 429H (or 310H); Mathematics 362K or Statistics and Data Sciences 321 (or Statistics and Scientific Computation 321); and Mathematics 340L, 341, or Statistics and Data Sciences 329C (or Statistics and Scientific Computation 329C).

MODALITY

The summer version of this class will be a 9-week online asynchronous course (June 1 - July 27). What this means is:

- You will watch recorded videos of the lectures. The videos are approximately 1.5 hours each, so 4 videos will be posted per week to make a total of 6 hours of lecture video per week.
- There will be one programming assignment due per week (except the first week and the last week). The assignments will be of the same length and difficulty as the regular semester.
- There will be 3 exams. You will take them remotely on specific, designated days.
- There will be two weekly office hours over Zoom, one with me, and one with the TA. You can always schedule an appointment with me or the TA if the office hour times do not work for you or if you need additional help.
- You can ask questions at all times on Piazza, where me, the TA, and other students will be answering daily.
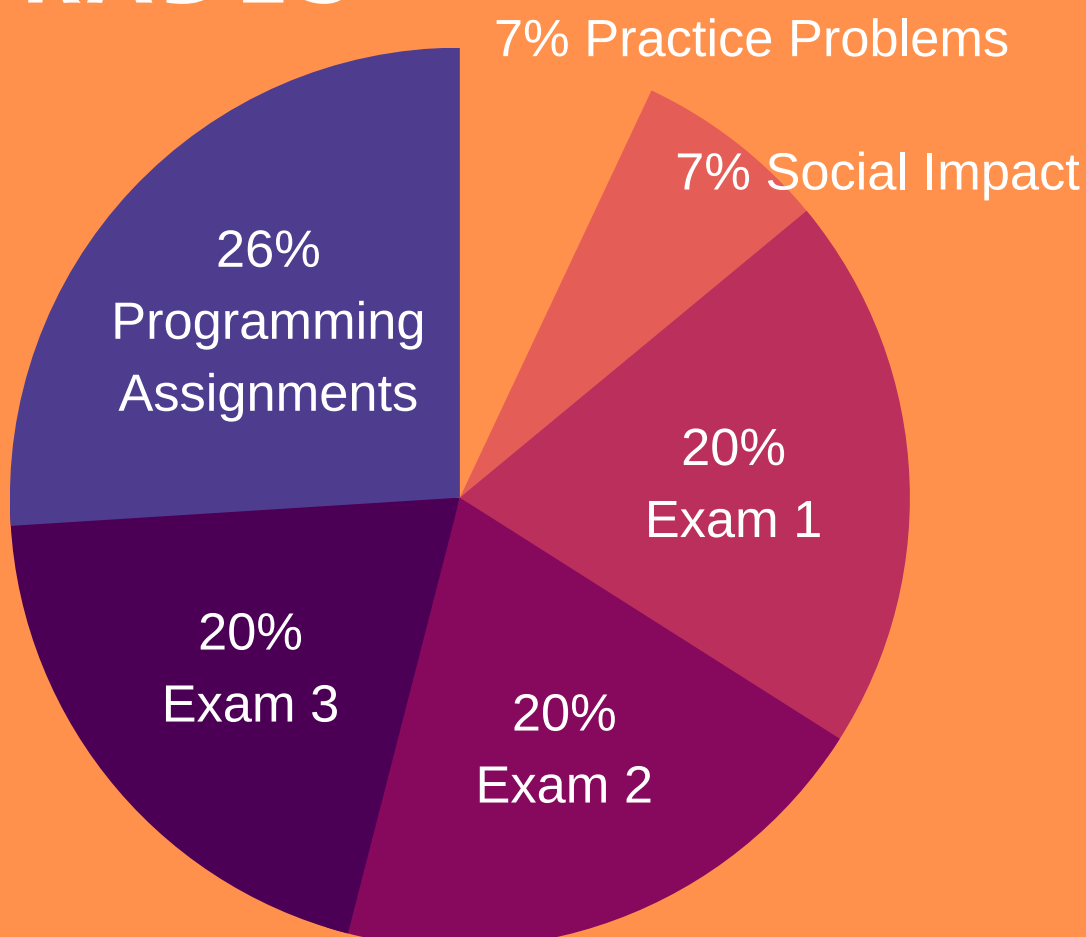
TEXTBOOK (optional)

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

GRADES

7% Practice Problems

7% Social Impact

26% Programming Assignments

20% Exam 1

20% Exam 2

20% 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 7 programming assignments, each equally weighted to total 26% of your final grade. The first assignment must be completed individually. Assignments 2-7 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.

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.

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 and the remaining will make up 7% of your final grade.

SOCIAL IMPACT

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

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

Additionally, each week, you will need to watch at least one social impact presentation video and write a response to it.

COURSE SCHEDULE

Subject to change at instructor's discretion.

SUPERVISED LEARNING ● ● ● ● ● ● ● ●

Week 1 & 2
6/1 - 6/11

Introduction
Data Prep, Exploration, Jupyter, Pandas
Feature Engineering, Dimensionality
Classification & Decision Trees [Ch 3.1-3.3]
Decision Trees (cont.)
HW 1 due 6/11, 11:59pm

Week 3
6/12 - 6/18

Linear Regression
Overfitting & Cross-Validation [Ch 3.4-3.9]
Nearest Neighbor [Ch 4.3]
Naive Bayes [Ch 4.4]
HW 2 due 6/18, 11:59pm

Week 4
6/19 - 6/25

Evaluating Classifiers [Ch 4.11]
Ensemble Methods [Ch 4.10]
SVMs [Ch 4.9]
Non-linear Regression
HW 3 due 6/25, 11:59pm

UNSUPERVISED LEARNING ● ● ● ● ● ● ●

Week 5
6/26 - 7/2

EXAM 1 due 6/27, 11:59p
Neural Nets [Ch 4.7]
Neural Nets (cont.)
Clustering & K-means [Ch 7.1-7.2]
HW 4 due 7/2, 11:59pm

Week 6
7/3 - 7/9

July 4th Holiday
Density-Based Clustering [Ch 7.4]
Hierarchical Clustering [Ch 7.3]
Evaluating Clusters [Ch 7.5]
HW 5 due 7/10, 11:59pm

RECOMMENDER SYSTEMS &
REINFORCEMENT LEARNING ● ● ● ● ●

Week 7
7/10 - 7/16

Anomaly Detection [Ch 9]
EXAM 2 due 7/12, 11:59p
Apriori [Ch 5.1-5.5]
Rule Gen & Evaluation [5.3-5.4, 5.7-5.8]
HW 6 due 7/16, 11:59pm

Week 8 & 9
7/17 - 7/25

Collaborative Filtering
Reinforcement Learning
Semester Wrap-Up
HW 7 due 7/23, 11:59pm
EXAM 3 due 7/25, 11:59pm

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 is
all your own work, unless it is assigned as group work. I will make
it clear for each exam or assignment whether collaboration is
allowed or not.

You are responsible for understanding UT's Academic Honesty
Policy which can be found here:
https://catalog.utexas.edu/general-
information/appendices/appendix-c/student-discipline-and-
conduct/#subchapter11400.prohibitedconduct

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

STUDENT SUPPORT
AND 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
and Access (D&A): http://diversity.utexas.edu/disability/.

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.

RELIGIOUS HOLY DAYS

By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an exam, a work assignment, or a project in order to observe a religious holy day, I will give you an opportunity to complete the missed work within a reasonable time after the absence.

Q DROP POLICY

If you want to drop a class after the 12th class day, you'll need to execute a Q drop before the Q-drop deadline, which typically occurs near the middle of the semester. Under Texas law, you are only allowed six Q drops while you are in college at any public Texas institution. For more information, see: http://www.utexas.edu/ugs/csacc/academic/adddrop/qdrop

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 in 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.