PRINCIPLES OF
MACHINE LEARNING

CS 363M, FALL 2022

MW 9:30-11:00, BUR 220 (52885)
TTH 11:00-12:30, PAR 201 (52890)
TTH 12:30-2:00, PAR 201 (52895)

PROFESSOR

Angie Beasley
angie.beasley@utexas.edu
GDC 6.314

TAs

Chloe Chen - chloetchen@utexas.edu
Jiaru Song - jiarus@cs.utexas.edu
Irene Xi - christin124216@gmail.com
Haoran Zhang - hz6453@utexas.edu

Please 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.
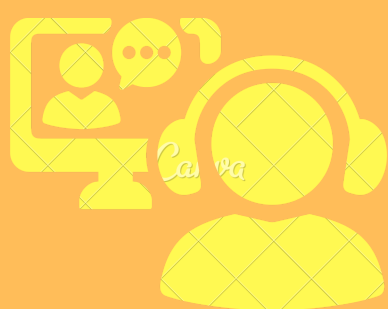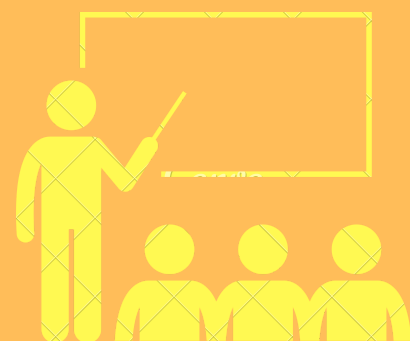
In this class, you will learn machine learning algorithms to find patterns in large input sets. We will cover supervised, unsupervised, and reinforcement learning. You will learn about regression, classification, clustering, anomaly detection, and association analysis. You will use Python and Jupyter 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.

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.

MODALITY

This class will be offered in a hybrid format. What that means is: I will be in the classroom teaching in-person, and I will also run Zoom during class, so you may choose either to come to class in person, or you may watch from home over Zoom. Additionally, you will not be restricted to the class time for which you are registered - you are welcome to attend any of the three class times (either in person or on Zoom), regardless of which one you are actually registered for... with two exceptions:

1) To take exams, you must attend in person during your officially registered class time.

2) On the day of your assigned social impact use-case, you must attend in person during your officially registered class time.

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

If you feel sick on a regular class day, you may watch from home on Zoom, or you may watch the Zoom recording at a later time, when you feel better.

If you feel sick on an exam day, or on your social-impact use-case day, contact me and we will work out a make-up exam, or a social-impact alternative.

If you do not feel comfortable (in general) attending exams or your social-impact use-case in person, please contact me to work out an alternative.
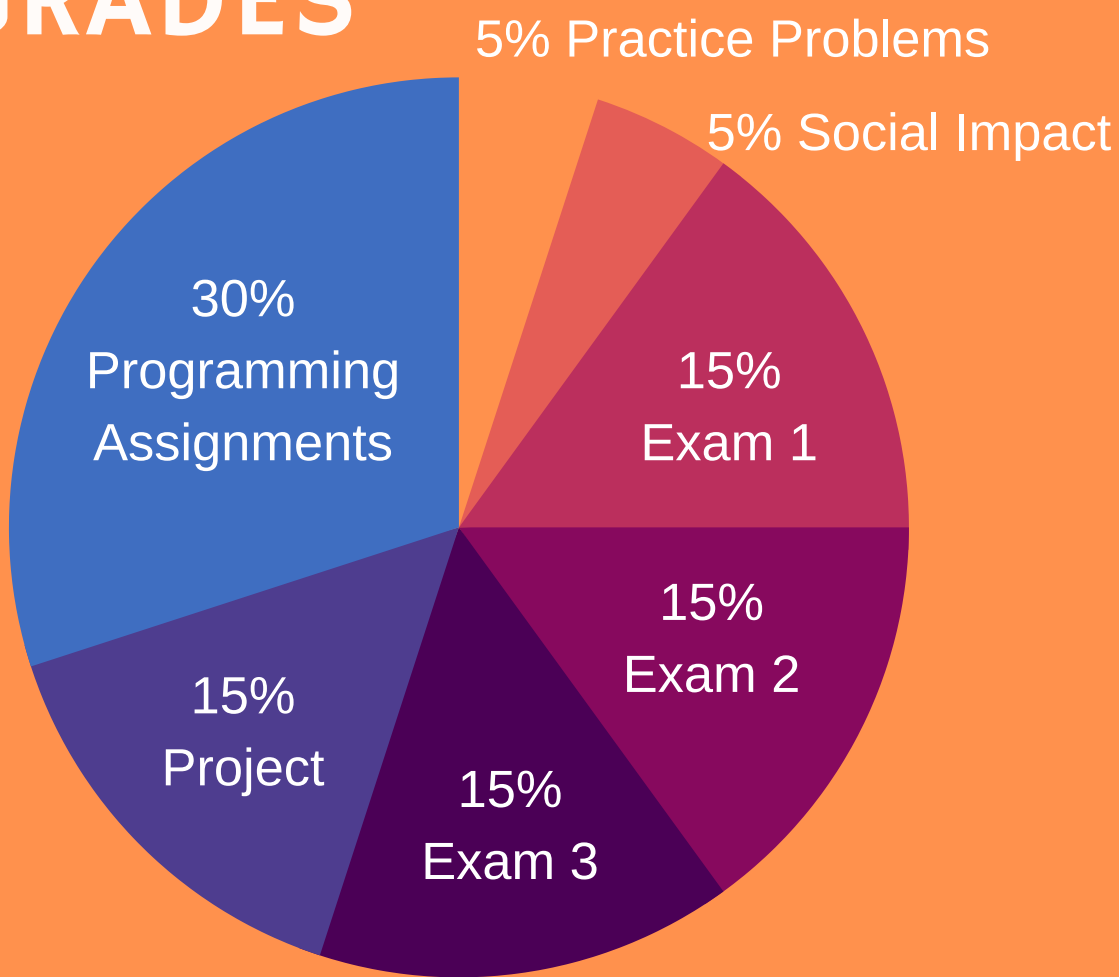
TEXTBOOK

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

The textbook is optional. It is for your reference.

GRADES

5% Practice Problems

5% Social Impact

30% Programming Assignments

15% Exam 1

15% Exam 2

15% Project

15% Exam 3

Grade	Threshold
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.

SOCIAL IMPACT

As more and more important 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 of 4 and assigned a 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.

PROJECT

Your project will be to model a dataset 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. The project will make up 15% of your final grade.

PROJECT SCHEDULE

10/14	Group selection deadline
10/28	Dataset selection deadline
11/20	Final project due

PROGRAMMING ASSIGNMENTS

There will be 6 programming assignments, each equally weighted to total 30% of your final grade.

Programming assignments must be completed using Python 3 and either Jupyter Notebooks or Google Colab. Assignments may be worked individually or in pairs. If you work in pairs, you are expected to use the proper pair programming method.

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 5% of your final grade.

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

8/22, 8/23	Introduction
8/24, 8/25	Data Prep, Exploration, Jupyter, Pandas
8/29, 8/30	Feature Engineering, Dimensionality

SUPERVISED LEARNING

8/31, 9/1	Regression
9/5	Labor Day Holiday
9/6, 9/7	Classification & Decision Trees [Ch 3.1-3.3]
9/8, 9/12	Decision Trees (cont.)
9/13, 9/14	Overfitting & Cross-Validation [Ch 3.4-3.9]
9/15, 9/19	Nearest Neighbor [Ch 4.3]
9/20, 9/21	Naive Bayes [Ch 4.4]
9/22, 9/26	Evaluating Classifiers [Ch 4.11]
9/27, 9/28	Ensemble Methods [Ch 4.10]
9/29, 10/3	SVMs [Ch 4.9]
10/4, 10/5	EXAM 1
10/6, 10/10	Neural Nets [Ch 4.7]
10/11, 10/12	Neural Nets (cont.)

UNSUPERVISED LEARNING

10/13, 10/17	Clustering & K-means [Ch 7.1-7.2]
10/18, 10/19	Density-Based Clustering [Ch 7.4]
10/20, 10/24	Hierarchical Clustering [Ch 7.3]
10/25, 10/26	Evaluating Clusters [Ch 7.5]
10/27, 10/31	Anomaly Detection [Ch 9]
11/1, 11/2	Guest speaker

REINFORCEMENT LEARNING & RECOMMENDER SYSTEMS

11/3	Apriori [Ch 5.1-5.5]
11/7, 11/8	EXAM 2
11/9	Apriori [Ch 5.1-5.5]
11/10, 11/14	Scalability & Rule Generation [Ch 5.3]
11/15, 11/16	Sequential Patterns [Ch 6.4]
11/17	Collaborative Filtering
11/21-11/25	Fall Break
11/28	Collaborative Filtering
11/29, 11/30	Reinforcement Learning
12/1, 12/5	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 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 Services for
Students with Disabilities (SSD):
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.

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.