Intro to Machine Learning (PSY 341K)
Syllabus
Spring 2021

PLACE AND TIME: Zoom only, synchronous, 10-11 MWF

INSTRUCTOR:

Ian Nauhaus, Ph.D.
Office: Remote
Office hours: Friday 8:30-10:00 am
E-mail: nauhaus@utexas.edu

OVERVIEW:

This course is about how to analyze data to make autonomous predictions. The end goal is proficiency with introductory machine learning topics to address problems in psychology and neuroscience. The main focus will be on supervised learning algorithms for regression and classification, from linear models to neural networks. You will also be introduced to unsupervised learning. Foundational topics will include performance metrics, linear algebra, and statistics. Every topic will be accompanied by lectures and exercises with Python programming. The goal of programming implementation is to solidify theoretical concepts and to provide more practical knowledge. Assignments will require students to alter and run Python code that implements a machine learning algorithm, plot results, and interpret results in a written format.

Prerequisites include an upper division statistics course, such as PSY 420L. Although there is not a programming prerequisite, it is highly recommended that students arrive with prior experience using a language such as C, Matlab, or Python. Early in the course, there will be a brief introduction to use of the Python programming language for statistical analyses and plotting. You will need a computer to perform the programming assignments.

COURSE STRUCTURE

Lectures and slides will cover all of the required material for the course. Each topic will be covered in two sets of lectures:

- ***Background lectures**:* This will be a standard slide presentation where I describe a ML topic. I'll explain the rationale of its application, the math, and a few examples.

- ***Coding lectures**:* You are only as good at machine learning as your ability to implement it with code. Python is perhaps the most common language used for machine learning, and data science in general. Background lectures are typically followed by a Python implementation lecture. I will share my Python screen with you and execute pieces of code and discuss the generated figures. I encourage you to play around in Python while following along. This Python code will be given to you ahead of time, and will be required for your assignments.

COURSE EVALUATION

Credit for the course is based on 7 homework assignments (50% total), two midterms (15% each), and a final exam (20%). An 8^{th} homework assignment will be extra credit. Attendance will not be used for grading. Final grades will be assigned as follows: A, 92 and above; A-, 90-91.99; B+, 87-89.99; B, 83-86.99; B-, 80-82.99; C+, 77-79.99; C, 73-76.99; C-, 70-72.99; D+, 67-69.99; D, 63-66.99, D-, 60-62.99, F, below 60

- **Midterms and final**
  - All exams will be open book. However, if you don't understand the material, you will not have enough time to finish.
  - Test material will be mostly limited to machine learning theory, not programming. However, do expect to interpret some short snippets of code. You won't be expected to write your own Python syntax, only to interpret something I've written.
  - All test material will be covered in the lectures and slides.
  - The midterms will not be comprehensive, per se, as I will give you specific cutoff lecture dates. However, keep in mind that each lecture will often build on lectures from much earlier in the semester.
  - The final exam will pool from the beginning to end of the semester, but I will specify a few topics that will not be fair game, along with a few topics that will be more important.

- **Homework assignments**
  - Assignments will largely consist of a series of tasks that require you to alter the Python code that is the basis of the "coding lectures" (see above). Your assignment will be submitted as a consolidated pdf containing the plots, written explanations, and the Python code that you executed. I will give you some examples of how I want this formatted.
  - Most of the knowledge required to complete the homework will be in the lecture. However, some of you will have less programming experience than others. *In general, expect to do some independent inquiry of how to format your code for a given homework task. I won't be able to give you all 100% of the necessary Python background.* I've provided links to some resources below.

REQUIRED AND RECOMMENDED MATERIALS

I will attempt to minimize equations while maximizing geometrical interpretations (e.g. graphs) as much as possible. That said, some math is necessary to understand the concepts and to ultimately implement them in Python. A lot of textbooks have more math than you need for this course (or to be really good at machine learning). My slides and the required text, I believe, strike a good balance.

- **Machine learning theory**

  - Lectures and slides (required). You will be given recordings and slides. Together, they will cover the necessary material for the tests.

  - Python Machine Learning by S. Raschka and V. Mirjalili (highly recommended). This text has a combination of valuable features that other free books (see below) do not. For one, the math has relatively simple matrix *notation which will be consistent with my lectures*.

Second, it generally lacks mathematical derivation that is outside the scope of my lectures. Third, there are Python programming examples in every case, which you can download from the author's Github site. Lastly, it has a much more modern and ML-specific perspective. You can rent it and/or get it for a free trial on Packt.

- Free text (recommended). For some lectures, I will point you to specific sections from free text. However, keep in mind that these will have slightly different math notation than my lectures. If you have a stronger math background, notation differences shouldn't trip you up. I simply mention these books since they are free and may help concepts "click" for you better than my lecture - everyone is different.
  - Trevor Hastie, Robert Tishirani, Jerome Friedman
    Elements of Statistical Learning
  - Christopher M. Bishop
    Pattern Recognition and Machine Learning

- Curated YouTube videos (recommended). Machine learning is a hot topic, which means there is ton of great content on each of the topics we will cover. I encourage you to look around on your own. But in some cases, I will put a link in the lecture slides to something that I think is especially good.

- **Python implementation**

  - Computer (required)

  - My Python code and coding lectures (required). For every topic, you will be given the Python code discussed in the "coding lectures". The code has lots of comments. For the homework, you will be expected to understand the code so that you can alter it, execute it, and interpret the figures it generates.

  - You are required to use the Spyder IDE to execute Python script. To make installation easy, you can just download Spyder via the Anaconda package, which will automatically install all the necessary Python toolboxes for this course. FYI, Anaconda contains other things you don't need for this course, such as RStudio and Jupyter notebook.

  - Vanderplas, *Python Data Science Handbook* (recommended) - There will not be any required reading for programming, but you may find this useful while doing your homework assignments.

  - Google (recommended). Specifically, expect to run up against problems surrounding Python syntax when doing your homework. When I have a Python question (still quite often), I start with Google, which often brings me to one of the developer sites: *https://numpy.org/*; *https://scikit-learn.org/stable/*; *https://matplotlib.org/*, *https://www.python.org/*.

COURSE OUTLINE

Please note the topics or their order are subject to change at my discretion. It is your responsibility to note these changes when announced. Expect each section to take between 1 and 3 weeks.

I. Introduction

- Syllabus
- What is machine learning?
- Python programming introduction

II. Making functions (efficiently) in a computer program

- Linear algebra basics: Types of multiplication and addition
- Converting functions into linear algebraic notation and Python code

III. Probability Theory

- Review of probability theory needed for this course: Joint distributions, moments, conditional probabilities, likelihoods, Bayes' theorem.
- Probabilistic observation (**x**) and state (y) models: p(**x**,y).

Assignment #1 on linear algebra and probability theory

IV. Linear regression

- Fitting lines and hyperplanes
- Fitting polynomials

V. Identifying the best ML model

- Which observations (for a given model) are most useful for making predictions?
- Which model makes the best predictions?
- Cross validation: We will discuss K-folds cross validation.
- Cross-validation will be introduced with linear regression, but they are fundamental to ML. We will come back to cross validation with other algorithms.

Assignment #2 on linear regression and cross validation

VI. Fitting model parameters with a cost function

- Cost functions will be introduced with linear models, but they are fundamental to ML. We will come back to this for the rest of the semester for other ML algorithms.
- Mean-squared error (MSE)
- Maximize the likelihood (MLE)
- Maximize the posterior (MAP)

Assignment #3 on cost functions
Midterm #1 on II-VI (March 5th)

VII. Supervised classifiers part 1: Basics

- Unlike regression, the output of a classifier is a discrete category (e.g. 'A' vs. 'B').
- Signal detection theory basics: Confusion matrix and ROC curves

VIII. Supervised classifiers part 2: Naïve Bayes

- A simple use of Bayes theorem to make predictions. Naïve assumption of independent observations.

Assignment #4 on signal detection theory and naïve Bayes

IX. Supervised classifiers part 3: Logistic regression

- Appends the output of your linear model with a sigmoid of Bernouli likelihoods. Set a threshold to make it a classifier.

Assignment #5 on logistic regression

X. Unsupervised classifier: K-means clustering

- Training data does not always tell us the "category" of each observation.
- We will look at K-means clustering, which is the most common unsupervised classifier

Assignment #6 on K-means clustering
Midterm #2 on VII-X (April 5th)

XI. Supervised classifiers part 4: Decision trees

- Classification trees
- Ensemble methods: Random forest

Assignment #7 on decision trees

XII. Working with Scikit-learn toolboxes

XIII. Supervised classifiers part 5: Neural networks

- This will be a brief introduction with no math. A more in-depth discussion is beyond scope. We will go over terminology of their components, useful applications, and how to implement them in scikit-learn.

Assignment #8 on neural networks

Cumulative final exam: Saturday, May 15th, 7pm-10pm

University Resources for Students and Safety information

COVID-19 Update: While we will post information related to the contemporary situation on campus, you are encouraged to stay up-to-date on the latest news as related to the student experience. https://coronavirus.utexas.edu/students

Students with Disabilities: This class respects and welcomes students of all backgrounds, identities, and abilities. If there are circumstances that make our learning environment and activities difficult, if you have medical information that you need to share with me, or if you need specific arrangements in case the building needs to be evacuated, please let me know. I am committed to creating an effective learning environment for all students, but I can only do so if you discuss your needs with me as early as possible. I promise to maintain the confidentiality of these discussions. Any student with a documented disability who requires academic accommodations should contact Services for Students with Disabilities at 471-6259 (voice) or 512-410-6644 (Video Phone) as soon as possible to request an official letter outlining authorized accommodations. For more information, visit http://ddce.utexas.edu/disability/about/.

Counseling and Mental Health Center: All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful. If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. http://www.cmhc.utexas.edu/individualcounseling.html

Title IX Reporting: Title IX is a federal law that protects against sex and gender-based discrimination, sexual harassment, sexual assault, sexual misconduct, dating/domestic violence and stalking at federally funded educational institutions. UT Austin is committed to fostering a learning and working environment free from discrimination in all its forms. When sexual misconduct occurs in our community, the university can:

1. Intervene to prevent harmful behavior from continuing or escalating.
2. Provide support and remedies to students and employees who have experienced harm or have become involved in a Title IX investigation.
3. Investigate and discipline violations of the university's relevant policies (https://titleix.utexas.edu/relevant-polices/).

Beginning January 1, 2020, Texas Senate Bill 212 requires all employees of Texas universities, including faculty, report any information to the Title IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Texas law requires that all employees who witness or receive any information of this type (including, but not limited to, writing assignments, class discussions, or one-on-one conversations) must be reported. *I am a Responsible Employee and must report any Title IX related incidents* that are disclosed in writing, discussion, or one-on-one. Before talking with me, or with any faculty or staff member about a Title IX related incident, be sure to ask whether they are a responsible employee. If you would like to speak with someone who can provide support or remedies without making an official report to the university, please e-mail advocate@austin.utexas.edu. For more information about reporting options and resources, visit http://www.titleix.utexas.edu/, contact the Title IX Office via e-mail at titleix@austin.utexas.edu, or call 512-471-0419.

Although graduate teaching and research assistants are not subject to Texas Senate Bill 212, they are still mandatory reporters under Federal Title IX laws and are required to report a wide range of behaviors we refer to as sexual misconduct, including the types of sexual misconduct covered under Texas Senate Bill 212. The Title IX office has developed supportive ways to respond to a survivor and compiled campus resources to support survivors.

Emergency Evacuation Procedures*:* The following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, http://www.utexas.edu/safety/

Sharing of Course Materials is Prohibited: No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional 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.

Class Recordings: Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.