ISTA 421: Introductory Machine Learning

Fall 2024 Syllabus

Instructor:
Salena Torres Ashton | salena@arizona.edu
454 Harvill College of Information, University of Arizona

Class Hours: 11:00 am - 12:15 pm, Tuesdays & Thursdays
Class Location: 332B Harvill, in-person only
Office Hours: TBD

Description

In this course, we will investigate how algorithms and parameters can be used to recognize patterns in data, how to make decisions based off of these patterns, and how to decide which algorithms are the best to use for certain problems. Automating these pattern predictions and classifications is known as machine learning. While this may sound like a theoretical form of data science, it is not. We will investigate key similarities between machine learning and data science, key differences, and why these differences matter.

Course Logistics

Objectives

As a multidisciplinary field, the course introduces concepts and work from many areas critical to information studies including statistics, machine learning, pattern recognition, database technology, and data visualization.

By the end of this course, students will:

- Understand a large set of concepts of machine learning.
- Evaluate and use software packages to perform machine learning analyses.
- Explain and interpret results from different machine learning algorithms.

Competencies addressed by the College of Information:

- F1.2: Students will demonstrate facility using basic research methods, for example: research design; statistics and analysis; organization, identification, and location of data and information including open- and closed-access sources; and/or presentation of findings in oral, written and multi-media form, including proper use of and citation of sources.
- DAISBS2.2: Students will establish the ability to exercise the four key techniques of computational thinking (decomposition, pattern recognition, abstraction, and algorithms)in solving information and data challenges.

Prerequisites

We will use statistical learning packages, texts, and resources for machine learning in Python.

- Introductory statistics and probability theory (ISTA 311 recommended; ISTA 116 or comparable is sufficient).
- 2 or 3 semesters of programming (does not have to have been in Python)
- Calculus I is highly recommended, not required. If you have not taken calculus, please talk with me.
- Please dust off your algebra skills!

Workloads

For each lecture hour, students are expected to spend 3-5 hours, completing the required reading and course work. If you are spending more than 5 hours of time outside of class, please come to office hours. Within reason, I also welcome your feedback about course load and workload adjustments.

Materials

- **An Introduction to Statistical Learning with Applications in Python:** by Gareth James, Daniela Witten, Trevor Hastie, et al. July 2023) https://www.statlearning.com

Additional resources recommended for this course:

- **StatsQuest with Josh Starmer:** https://www.youtube.com/@statquest/playlists Great playlists for statistics and machine learning. He also has nice playlists for specific models like linear regression, support vector machines, and random forests. His visualizations make concepts easier to understand and he uses words in his visuals, which make the concepts more accessible. (You may notice that he's also a musician.)

- **Software Carpentries: Unix Shell:** https://swcarpentry.github.io/shell-novice/. A beginner-friendly tutorial about how to use the terminal. Data science and machine learning often use large datasets that are best run by automated scripts. This means that using Quarto or Jupyter interactive notebooks may not be the best IDE. These scripts can automatically run your data through the code you write using the terminal. 3 extra credit points if you can tell me why interactive notebooks would be better than using scripts designed for automation.

- **Math Reviews from other sheet**

Course Structure

University Schedule

Event	Date
Classes begin	August 26, 2024
Labor Day - no classes	September 2, 2024
Veterans Day - no classes	November 11, 2024
Thanksgiving recess	November 28 - December 1, 2024
Last day of classes	December 11, 2024
Final examinations	December 13-19, 2024
Degree award date for students completing by close of Fall Session	December 20, 2024

Grades

Your final grade is composed from:
- **Participation:** 10%
- **Coding and Written Homework:** 40%
- **Papers:** 20%
- **Exams:** 30%

Cite all of your sources: documentation from code and tutorials, online repositories, datasets, sources for online help, discussions with fellow classmates, etc. When in doubt, cite it. You will be marked down for uncited sources and sources that cannot be verified. Read more about academic integrity in this syllabus.

Coding and Written Homework (40% Final Grade)

All homework is to be written in Python. Homework submissions are required to be pushed onto GitHub, be fully reproducible as either a ipynb, qmd, or py file. We will have approximately 2 weeks per homework assignment. *This schedule is subject to change*– I may assign fewer assignments than those listed below but I will not assign more than what is listed below.

Week	Assignment	Reading Assignment
1	Introduction and Review	Chapter 1-2 (James et al)
2-3	Linear Regression and Multiple Linear Regression	Chapter 3 (James et al)
4-5	Classification	Chapter 4 (James et al)
6	Resampling Methods	Chapter 5 (James et al)
7	Model Selection	Chapter 6 (James et al)
8	Midterms and Catch-up Week	Review previous materials
9-10	Moving Beyond Linearity	Chapter 7 (James et al)
11	Tree-based Methods	Chapter 8 (James et al)
12	Large Language Models	TBD
13-14	Unsupervised Learning	Chapter 12 (James et al)
15-16	Final Project, Exam Review, Catch-up	TBD

Exams (30% Final Grade)

1. Pretest of Concepts
2. Midterm Exam
3. Final Exam or a Final Project of Your Choice (TBD)
4. In-class mini-quizzes. **Be sure to bring paper and pencil to class!**

Papers (20% Final Grade)

Topics are subject to change.

1. Machine Learning, Data Science, Information Science and Computer Science
2. Machine learning applications in your discipline of interest
3. Final Exam or a Final Project of Your Choice (TBD)

Participation (10% Final Grade)

- **Participation** in class discussions will be a part of your grade (10%). Participating does not require that we understand everything. Asking questions, requesting examples to clarify concepts, answering each others' questions are all forms of active participation.
- **Correcting, clarifying, debugging, and communicating about code:** these are all forms of participation and collaboration. A large part of information science and machine learning in the real world will require that you can clearly explain your code and applied solutions to employers who do not program.

Incomplete Work, Etc

The grade of "I" may be awarded only at the end of a term, when all but a minor portion of the course work has been satisfactorily completed. The grade of I is not to be awarded in place of a failing grade or when the student is expected to repeat the course; in such a case, a grade other than I must be assigned. Students should make arrangements with the instructor to receive an incomplete grade before the end of the term.

Scale

A = 90 and above
B = 80 - 89
C = 70 - 79
D = 60 - 69
F = 59 and below

I am happy to round grades, no more than one full percentage point, if and only if your final participation grade is 80% or higher.

Due Dates

• All work must be turned in on the date due by midnight (11:59pm) Tucson time.
• I will accept two late assignments, and only two, as follows:

1. First late assignment receives 0% reduction if turned in within 3 days and I have given you written approval through email. If the assignment is turned in day 4 or later, I will reduce by 30%.
2. The second late assignment receives a 10% reduction if turned in within 3 days and I have given you written approval through email. If the assignment is turned in 4 or more days after the due date, I will reduce the grade by 50%.
3. Life does happen. Talk with me about any barriers you are experiencing with assignments. Prior notice and communication are key!

Accessibility

- Because I wear hearing aids, I ask that you speak loud and clear enough. It's perfectly okay to ask questions about this hearing loss— I welcome the questions!
- If you have accommodations you need addressed, please consult your [DRC advocate at the University of Arizona](#).
- I welcome discussions about disability and accessibility; I will respect your privacy and I will work with you to find reasonable accommodations. Please keep in mind that:
- You are not required to disclose your disability to me or any other instructor, to any student, or to any other person. If you have been pressured to do so, please report this to [Amanda Kraus](#), Executive Director of the Disability Resources and ADA/504 Compliance officer.your DRC advocate.
- I prefer to hold all office hours in person. If this is a challenge for you, please contact me.

Safety on Campus and in the Classroom

For a list of emergency procedures for all types of incidents, please visit the website of the [Critical Incident Response Team (CIRT)](#). A [video](#) is also available.

Academic Integrity

All submitted work must be your original work. Do not turn in any work from other classes or by other people. Plagiarism, undisclosed generative AI (see below), or any other form of academic dishonesty will not be tolerated and I will take aggressive steps against this dishonesty. If you don't know how to do something or you feel that you don't have enough time, please don't resort to dishonesty. Speak up in class, contact me directly and we'll work together.

University Policies

University syllabus policies can be found at https://catalog.arizona.edu/syllabuspolicies.

iSchool Academic Integrity Policy Syllabus Statement

This policy, agreed upon by faculty in the UArizona iSchool, applies in addition to the [Dean of Students' Code of Academic Integrity](#).

Students in courses at the UArizona iSchool are expected to maintain rigor in their academic performance with intent to learn, practice, and overcome challenges toward personal growth and enrichment. As future professionals in digital environments, iSchool students are also expected to exercise transparency and integrity in collaborations and in the use of tools and resources that may aid completion in assignments for our courses.

Consider the following PROHIBITED practices in this course, unless I have *specific and written* instructions to do otherwise:

- Posting a question on an online site such as Chegg.com, and copying and pasting some or all of the response into an assessment.
- Posting an assessment from the course on online sharing sites such as Course Hero. Aiding other students in violation of academic integrity is also a violation, and is potential copyright infringement.
- Using, in whole or in part, computer code not written by the student (for example, from another student, a book, or the internet) in an assignment or project. This includes using such code in modified or unmodified form.
- Searching for solutions to projects or assignments on the internet or through other tools, when I intended for you to learn the solution through exercises (e.g. Googling for the solution to a question on an assignment).
- Simultaneously submitting the same assignment as another student enrolled into the course without prior permission from the instructor.

Exceptions: Clear Instructions will be Provided.

In any cases in which this course requires or permits students to use practices in the list above, clear written instructions will specify the tools allowed or required, so students can be certain they are working as instructed. See the [UArizona iSchool Academic Integrity Policy](#), the [UArizona Code of Academic Integrity](#) and [Syllabus Policy](#) for more information.

LLMs and ChatGPT

Large language models (LLMs) like ChatGPT are a type of artificial intelligence (AI) engine that can look like it generates the code you need for labs and short answer questions. You are encouraged to use ChatGPT to debug code and experiment. You are also held accountable for anything generated that is incorrect, which will result in marked off points for incorrect answers or implementations. This means that if you use generative AI and cite its use, but the answer or implementation it generated was inappropriate or incorrect, it is *your responsibility* to verify its accuracy.

I am 100% okay with students using generative AI. Just like the Internet, the slide rule, and the cotton gin, this technology is here to stay. You may use generative AI under these conditions:

- Explicitly state your use: which AI model, when, and save the entire conversations and email me links and give me screen shots. We will demonstrate in class what this means.
- Simply telling me "I used Chat GPT to ask how to use descriptive statistics for…" will not work. You must show me the entire conversation used.
- You must demonstrate that you understand the code. If you cannot talk your way through the code or answer questions about the code, the program structure, or the inputs/outputs, I may assume that you were not its author.
- Remember that these models are not fully-reasoning, functioning sources of information. They will calculate mistakes and sound confident in their mistaken answers. If you use generative AI, you will be held responsible for using these answers, just like you would be held accountable for the sources you cite for your written work.
- A word of warning: I enjoy debugging code and I enjoy researching the differences in human-written code versus generative AI-written code. My first career was in historical research– looking for nearly-impossible needles in forgotten haystacks, often in the forms of falsified documents from one century to the next. It prepared me well for deciphering plagiarism in today's world. Another fair warning: I love history and if you ask me historical questions, we run the risk of talking about history all hour long. Three extra credit points for telling me about your favorite period of history in your first assignment.

Information contained in this course syllabus, may be subject to change, as deemed appropriate by the instructor.