CSC 380: Principles of Data Science

Tue/Thu, 3:30 – 4:45pm, Meinel Optical Sci, Rm 410

Description of Course

The course introduces students to principles of data science that are necessary for computer scientists to make effective decisions in their professional careers. A number of computer science sub-disciplines now rely on data collection and analysis. For example, computer systems are now complicated enough that comparing the execution performance of two different programs becomes a statistical estimation problem rather than a deterministic computation. This course teaches students the basic principles of how to properly collect and process data sources in order to derive appropriate conclusions from them. The course has three main components: data analysis, machine learning, and a project where students apply the concepts discussed in class to a substantial open-ended problem.

Course Format and Teaching Methods

The course is taught in a lecture format, with lectures given by the instructor. Lectures will involve discussions making use of lecture slides, work on the white board, and input from students. Lectures will augment, enhance, and reiterate material from the relevant textbooks. Assignments and exams all involve individual work. Attending the lectures throughout the semester is required.

Instructor and Contact Information

Course D2L: https://d2l.arizona.edu/d2l/home/1403936

Instructor:
Xinchen Yu, xinchenyu@arizona.edu

Teaching assistants:
Yinan Li, yinanli@arizona.edu
Tugay Bilgis, tbilgis@arizona.edu

Office hours: TBA

Class Recordings

- The class recordings will be made for every class and available via D2L.
- The recordings will be available AFTER the class, and they **will not be live-streamed** because the course is design to be in-person, not an online class.
- If the student does not wish to be identified by name, please let the instructor know.
- For lecture recordings, which are used at the discretion of the instructor, students must access content in D2L only. Students may not modify content or re-use content for any purpose other than personal educational reasons. All recordings are subject to government and university regulations. Therefore, students accessing unauthorized recordings or using them in a manner inconsistent with UArizona values and educational policies (Code of Academic Integrity and the Student Code of Conduct) are also subject to civil action.

Course Objectives

An introduction to basic concepts in data science and machine learning. Topics include: descriptive statistics, basic data analysis, basic data visualization, predictive models and training, basic supervised and unsupervised learning models, evaluation measures.

Expected Learning Outcomes

A student who successfully completes this course will be able to:

1. Explain the difference between different measures of centrality and variability (means vs medians, variance vs interquartile range, etc.) (Part 1: remedial descriptive stats outcome)
2. Articulate the meaning of confidence intervals associated with statistical hypothesis tests (Part 1: remedial stats outcome)
3. Learn how to use probability and non-probability sampling to collect data from a population (Part 2: data collection outcome)
4. Learn how to identify potential sampling bias (Part 2: data collection outcome)
5. Convert a "raw" data source into a version appropriate for downstream analysis using Python (Part 2: data processing outcome)
6. Write appropriate visualizations for different sources and types of data (Part 2: basic data visualization outcome)
7. Explain why we seek to build machine learning models that generalize rather than memorize their inputs (Part 3: basic machine learning outcome)
8. Explain the different uses for training, validation, and testing datasets (Part 3 basic machine learning outcome)
9. Select appropriate evaluation measure for the dataset and task being solved (Part 3: basic machine learning outcome)
10. Articulate the difference between supervised and unsupervised machine learning, as well as select the appropriate methodology for a given problem (Part 3: basic machine learning outcome)

Makeup Policy for Students Who Register Late

Makeup can only be made for HW1 with the submission deadline being 8 days after the enrollment.

Course Communications

All announcements will be made via D2L. Assignments will be distributed and submitted via Gradescope. Off-class discussions among students and questions to the instructors/TAs will be conducted via Piazza. Office hours with the instructor will be conducted remotely, over Zoom.

- D2L: https://d2l.arizona.edu/d2l/home/1403936
- Gradescope: https://www.gradescope.com/courses/692188 with entry code: 8EZRBY **(Note: please sign up gradescope using the same email address you have on D2L)**
- Piazza: https://piazza.com/arizona/spring2024/csc380 access code: wildcats

Required Texts or Readings

- WJ: Watkins, J., "An Introduction to the Science of Statistics: From Theory to Implementation" (https://www.math.arizona.edu/~jwatkins/statbook.pdf)
- MK : Murphy, K. "Machine Learning: A Probabilistic Perspective." MIT press, 2012

(accessible online via UA library)
- WL: Wasserman, L. "All of Statistics: A Concise Course in Statistical Inference." Springer, 2004 (accessible online via UA library)
- ISL: James, G., Witten, D., Hastie, T., & Tibshirani, R. An Introduction to Statistical Learning with Applications in Python. New York: Springer (https://www.statlearning.com/).

Required or Special Materials

None

Required Extracurricular Activities

None

Assignments and Examinations: Schedule/Due Dates

	Description	Assigned/Date
HW1	Probability 1	Jan 18
HW2	Probability 2	Jan 30
HW3	Data Processing and Visualization	Feb 8
HW4	Statistics 1	Feb 20
HW5	Statistics 2	Feb 29
Midterm		Mar 14
HW6	Predictive Modeling and Classification	Mar 21
HW7	Linear Supervised Learning Models	Apr 2
HW8	Nonlinear Supervised Learning Models	Apr 11
Project	Participate in a data science competition	Apr 16
Final Exam		May 8

Homeworks will be due in 8 days. For example, if the homework was assigned on Thursday, then it is due by next Friday.

Final Examination

This course will have a final examination on **May 8, Wednesday, 3:30-5:30pm**.

See also: Final Exam Regulations, https://www.registrar.arizona.edu/courses/final-examination-regulations-and-information, and Final Exam Schedule, http://www.registrar.arizona.edu/schedules/finals.htm .

Final Project

The final project deadline is **May 2, Friday, 11:59pm**. It will be assigned Apr 16. The final project is to solve a data science competition problem that happened in the past. Students will have to solve a problem provided by the instructor, which may include a few open questions as well.

Grading Scale and Policies

University policy regarding grades and grading systems is available at http://catalog.arizona.edu/policy/grades-and-grading-system

- Assignments: 35%
- Midterm: 25%
- Project: 10%
- Final Exam: 25%
- Participation: 5%

Grade Distribution for this Course:

- A: 90% and above
- B: 80%-89.99%
- C: 70%-79.99%
- D: 60%-69.99%
- E: 59% and below

University policy regarding grades and grading systems is available at http://catalog.arizona.edu/policy/grades-and-grading-system

Policies

- Students can do the assignments individually or in pairs. If students choose to do the HWs in pairs, they are expected to **contribute equally for each individual question**. Please let me know if you believe you and your partner are contributing unequally in your HWs, and we will meet with both of you to discuss score adjustments.
- Late assignments or projects receive a grade of zero. There will be **no late days** for homework and project submissions. Plan to submit early to guard against unexpected delays.
- There are 7 assignments in total; the assignment with the lowest score will be dropped.
- Let us define 'academic days' to be those days that are not weekends and university holidays.
- The grading of homework, midterm exam, and project will be available in 7 academic days. The graded homework will be returned before the next homework is due.
- Requests for regrading of homework and midterm exam can only be done within 7 academic days of the grade release.
- Regarding policy for the final project and the final exam will be within 3 academic days since the grades are announced.
- All exams will be closed-book.
- Missed exams result in a grade of zero.
- The final exam grading will be available within 48 hours.
- Grading delays beyond promised return-by dates will be announced as soon as possible with an explanation for the delay.

Incomplete (I) or Withdrawal (W):

Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete and http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal respectively.

Dispute of Grade Policy:

Requests for regrading assignments or exams must be submitted within one week of receiving the graded homework or exam.

Honors Credit

Students wishing to contract this course for Honors Credit should e-mail me to set up an appointment to discuss the terms of the contact and to sign the Honors Course Contract Request Form. The form is available at http://www.honors.arizona.edu/honors-contracts

Scheduled Topics and Activities

Week 01: Course mechanics, introduction to data science Readings: ● WL (Ch. 1) ● WJ (Ch. 5)
Week 02: Probability Readings: ● WJ (Sec. 7.1-7.3.1)
Week 03: Probability Readings: ● WL (Ch. 2) ● WJ (Sec. 7.3.2-7.8, Ch. 9)
Week 04: Probability Readings: ● WL (Ch. 3) ● WJ (Ch. 8)
Week 05: Data Processing and Visualization Readings: ● WJ (Ch. 1, 2)
Week 06: Statistics Readings: ● WL (Sec. 9.1-9.2, 14.1-14.4, 15.1-15.2) ● WJ (Ch. 12)
Week 07: Statistics Readings: ● WL (Sec. 8.3)
Week 08: Statistics ● WJ (Sec. 16.1-16.2, 20.1-20.4)
Week 09: Statistics + Midterm
Week 10: Predictive Models Readings: ● MK (Sec. 1.1-1.3)
Week 11: Predictive Models Readings:

• MK (Sec. 1.4, 3.5)
Week 12: Linear Models Readings: • MK (Sec. 7.1-7.3, 7.5-7.6) • ISL (Sec. 6.1-6.2)
Week 13: Linear Models + Nonlinear Models Readings: • MK (Sec. 8.1-8.3) • ISL (Sec. 4.3.1-4.3.3, 7.1-7.3)
Week 14: Nonlinear Models Readings: • MK (Sec. 14.5, 14.4.3) • ISL (Sec. 10.1-10.7)
Week 15: Clustering Readings: • MK (Sec. 11.1-11.4.1)
Week 16: Course wrap-up, Project due
Week 17: Final Exam

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Students are asked to refrain from disruptive conversations with people sitting around them during lecture. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Notification of Objectionable Materials

This course will contain material of a mature nature, which may include explicit language, depictions of nudity, sexual situations, and/or violence. The instructor will provide advance notice when such materials will be used. Students are not automatically excused from interacting with such materials, but they are encouraged to speak with the instructor to voice concerns and to provide feedback.

Safety on Campus and in the Classroom

Familiarize yourself with the UA Critical Incident Response Team plans: https://cirt.arizona.edu/

Also watch the video available at https://ua-saem-aiss.narrasys.com/#/story/university-of-arizona-cert/active-shooter

University-wide Policies link

Links to the following UA policies are provided here, http://catalog.arizona.edu/syllabus-policies :

- Absence and Class Participation Policies
- Threatening Behavior Policy
- Accessibility and Accommodations Policy
- Code of Academic Integrity
- Nondiscrimination and Anti-Harassment Policy

Department-wide Syllabus Policies and Resources link

Links to the following departmental syllabus policies and resources are provided here, https://www.cs.arizona.edu/cs-course-syllabus-policies :

- Department Code of Conduct
- Class Recordings
- Illnesses and Emergencies
- Obtaining Help
- Preferred Names and Pronouns
- Confidentiality of Student Records
- Additional Resources
- Land Acknowledgement Statement

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.