CSC 380: Principles of Data Science

Tue/Thu, 3:30 – 4:45pm, C E Chavez Bldg, Rm 111

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 Prerequisites

Major: COSCBA or COSCBS. Completion of CSC 210 and CSC 244.

Instructor and Contact Information

Course D2L: https://d2l.arizona.edu/d2l/home/1325412

Instructor:
Xinchen Yu, xinchenyu@arizona.edu

Teaching assistants:
Saiful Salim, saifulislam@arizona.edu
Hui Ni, huini@arizona.edu
Shahriar Golchin, golchin@arizona.edu

Office hours: TBA

Obtaining Help

- **Academic advising:** If you have questions about your academic progress this semester, or your chosen degree program, consider contacting your department's academic advisor(s). Your academic advisor and the Advising Resource Center can guide you toward university resources to help you succeed. **Computer Science major students** are encouraged to visit https://www.cs.arizona.edu/undergraduate/advising for advisor contact information.
- **CS Tutor Center:** The Department of Computer Science offers FREE tutoring for students enrolled in CSC courses. You can view tutor schedules and sign up for tutoring sessions by visiting our CS Tutoring Page.
- **CS Help Desk:** The Computer Science IT team can help students with department technology issues including logging into/resetting your Lectura account, printing in the 930 lab, etc. You can submit a ticket for help by visiting the Computer Science Lab Helpdesk (requires UA login).
- **Life challenges:** If you are experiencing unexpected barriers to your success in your courses, please note the Dean of Students Office is a central support resource for all students and may be helpful. The Dean of Students Office can be reached at 520-621-7057 or DOS-deanofstudents@email.arizona.edu.
- **Physical and mental-health challenges**: If you are facing physical or mental health

challenges this semester, please note that Campus Health provides quality medical and mental health care. For medical appointments, call (520)-621-9202. For After Hours care, call (520) 570-7898. For the Counseling & Psych Services (CAPS) 24/7 hotline, call (520) 621-3334.

- **UA Ombuds**: The UA Ombuds Office (https://ombuds.arizona.edu/) helps with a wide variety of issues, concerns, questions, conflicts, and challenges. The primary mission of the Ombuds Program is to assist individuals in resolving conflict, facilitating communication, and assisting the University by surfacing issues and providing feedback on emerging or systemic concerns. Communications with the Ombuds Committee are informal and off-the-record. The Ombuds Committee is governed by the following standards: (1) Confidentiality; (2) Impartiality: (3) Informality; and (4) Independence.

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)

Absence and Class Participation Policy

If students desire to withdraw after the course withdrawal deadline, they are encouraged to see an advisor. Advisors will provide options and alternatives as appropriate for individual student situations.

Absence

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at https://catalog.arizona.edu/policy/class-attendance-and-participation

The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable: http://policy.arizona.edu/human-resources/religious-accommodation-policy.

Absences pre-approved by the UA Dean of Students (or dean's designee) will be honored. See https://deanofstudents.arizona.edu/policies/attendance-policies-and-practices

Participating in the course and attending lectures and other course events are vital to the learning process. As such, attendance is required at all lectures. Absences may affect a student's final course grade. If you anticipate being absent, are unexpectedly absent, or are unable to participate in class online activities, please contact me as soon as possible. To request a disability-related accommodation to this attendance policy, please contact the Disability Resource Center at (520) 621-3268 or drc-info@email.arizona.edu. If you are experiencing unexpected barriers to your success in your courses, the Dean of Students Office is a central support resource for all students and may be helpful. The Dean of Students Office is located in the Robert L. Nugent Building, room 100, or call 520-621-7057.

Illnesses and Emergencies

- If you feel sick, or may have been in contact with someone who is infectious, stay home. Except for seeking medical care, avoid contact with others and do not travel.
- Notify your instructor(s) if you will be missing up to one week of course meetings and/or assignment deadlines.
- If you must miss the equivalent of more than one week of class and have an emergency, the Dean of Students is the proper office to contact (DOS-deanofstudents@email.arizona.edu). The Dean of Students considers the following as qualified emergencies: the birth of a child, mental health hospitalization, domestic violence matter, house fire, hospitalization for physical health (concussion/emergency surgery/coma/COVID-19 complications/ICU), death of immediate family, Title IX matters, etc.
- Please understand that there is no guarantee of an extension when you are absent from class and/or miss a deadline.

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.

- Gradescope: https://www.gradescope.com/courses/567943 with entry code: ZZNJEN **(Note: please sign up gradescope using the same email address you have on D2L)**
- Piazza: https://piazza.com/arizona/fall2023/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	08/29
HW2	Probability 2	09/12
HW3	Statistics	09/21
HW4	Data Processing and Visualization, Hypothesis Testing	10/03
Midterm		10/12
HW5	Predictive Models	10/19
HW6	Linear Supervised Learning Models	11/02
HW7	Nonlinear Supervised Learning Models	11/14
Project	Participate in a Kaggle competition	11/16
Final Exam		12/13

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 **December 13, 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 **December 8, Friday, 11:59pm**. It will be assigned November 16. The final project is to solve a Kaggle competition problem that happened in the past. Students will have to solve a problem set 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: 36%
- Midterm: 20%
- Project: 14%
- Final Exam: 20%
- Participation: 10%

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.
- Participation may go over 10% for students who very actively participate in the discussions both in class and online (piazza).

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.

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/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: Statistics Readings: • WL (Sec. 9.1-9.2, 14.1-14.4, 15.1-15.2) • WJ (Ch. 12)
Week 06: Statistics Readings: • WL (Sec. 8.3) • WJ (Sec. 16.1-16.2, 20.1-20.4)
Week 07: Data Processing and Visualization Readings:

• WJ (Ch. 1, 2)
Week 08: Data Processing and Visualization + Midterm Readings: • WJ (Ch. 4)
Week 09: Data Processing and Visualization + Predictive Models Readings: • MK (Sec. 1.1-1.3)
Week 10: Predictive Models Readings: • MK (Sec. 1.4, 3.5)
Week 11: Linear Models Readings: • MK (Sec. 7.1-7.3, 7.5-7.6) • ISL (Sec. 6.1-6.2)
Week 12: Linear Models + Nonlinear Models Readings: • WK (Sec. 8.1-8.3) • ISL (Sec. 4.3.1-4.3.3, 7.1-7.3)
Week 13: Nonlinear Models Readings: • MK (Sec. 14.5, 14.4.3) • ISL (Sec. 10.1-10.7)
Week 14: Clustering Readings: • MK (Sec. 11.1-11.4.1)
Week 15: Clustering + Course wrap-up
Week 16: Course wrap-up, Project due
Week 17: Final Exam

Department of Computer Science Code of Conduct

The Department of Computer Science is committed to providing and maintaining a supportive educational environment for all. We strive to be welcoming and inclusive, respect privacy and confidentiality, behave respectfully and courteously, and practice intellectual honesty. Disruptive behaviors (such as physical or emotional harassment, dismissive attitudes, and abuse of department resources) will not be tolerated. The complete Code of Conduct is available on our

department web site. We expect that you will adhere to this code, as well as the UA Student Code of Conduct, while you are a member of this class.

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.

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-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.

Accessibility and Accommodations

At the University of Arizona, we strive to make learning experiences as accessible as possible. If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, https://drc.arizona.edu/) to establish reasonable accommodations.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See https://deanofstudents.arizona.edu/student-rights-responsibilities/academic-integrity .

Uploading material from this course to a website other than D2L (or the class piazza) is strictly prohibited and will be considered a violation of the course policy and a violation of the code of academic integrity. Obtaining material associated with this course (or previous offerings of this course) on a site other than D2L (or the class piazza), such as Chegg, Course Hero, etc. or accessing these sites during a quiz or exam is a violation of the code of academic integrity. Any student determined to have uploaded or accessed material in an unauthorized manner will be reported to the Dean of Students for a Code of Academic Integrity violation, with a recommended sanction of a failing grade in the course.

The University Libraries have some excellent tips for avoiding plagiarism, available at http://new.library.arizona.edu/research/citing/plagiarism.

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructor's express written consent. Violations to this and other course

rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA e-mail to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student e-mail addresses. This conduct may also constitute copyright infringement.

Nondiscrimination and Anti-harassment Policy

The University of Arizona is committed to creating and maintaining an environment free of discrimination. In support of this commitment, the University prohibits discrimination, including harassment and retaliation, based on a protected classification, including race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, or genetic information. For more information, including how to report a concern, please see http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy

Our classroom is a place where everyone is encouraged to express well-formed opinions and their reasons for those opinions. We also want to create a tolerant and open environment where such opinions can be expressed without resorting to bullying or discrimination of others.

Additional Resources for Students

UA Academic policies and procedures are available at http://catalog.arizona.edu/policies

Visit the UArizona COVID-19 page for regular updates.

Campus Health

http://www.health.arizona.edu/

Campus Health provides quality medical and mental health care services through virtual and in-person care. Voluntary, free, and convenient COVID-19 testing is available for students on Main Campus. COVID-19 vaccine is available for all students at Campus Health.

Phone: 520-621-9202

Counseling and Psych Services (CAPS)

https://health.arizona.edu/counseling-psych-services

CAPS provides mental health care, including short-term counseling services.

Phone: 520-621-3334

The Dean of Students Office's Student Assistance Program

https://deanofstudents.arizona.edu/support/student-assistance

Student Assistance helps students manage crises, life traumas, and other barriers that impede success. The staff addresses the needs of students who experience issues related to social adjustment, academic challenges, psychological health, physical health, victimization, and relationship issues, through a variety of interventions, referrals, and follow up services.

Email: DOS-deanofstudents@email.arizona.edu

Phone: 520-621-7057

Survivor Advocacy Program

https://survivoradvocacy.arizona.edu/

The Survivor Advocacy Program provides confidential support and advocacy services to student survivors of sexual and gender-based violence. The Program can also advise students about relevant non-UA resources available within the local community for support.

Email: survivoradvocacy@email.arizona.edu

Phone: 520-621-5767

Campus Pantry

Any student who has difficulty affording groceries or accessing sufficient food to eat every day, or who lacks a safe and stable place to live and believes this may affect their performance in the course, is urged to contact the Dean of Students for support. In addition, the University of Arizona Campus Pantry is open for students to receive supplemental groceries at no cost. Please see their website at: campuspantry.arizona.edu for open times.

Furthermore, please notify me if you are comfortable in doing so. This will enable me to provide any resources that I may possess.

Preferred Names and Pronouns

This course affirms people of all gender expressions and gender identities. If you prefer to be called a different name than what is on the class roster, please let me know. Feel free to correct instructors on your pronoun. If you have any questions or concerns, please do not hesitate to contact me directly in class or via email (instructor email). If you wish to change your preferred name or pronoun in the UAccess system, please use the following guidelines:

Preferred name: University of Arizona students may choose to identify themselves within the University community using a preferred first name that differs from their official/legal name. A student's preferred name will appear instead of the person's official/legal first name in select University-related systems and documents, provided that the name is not being used for the purpose of misrepresentation. Students are able to update their preferred names in UAccess.

Pronouns: Students may designate pronouns they use to identify themselves. Instructors and staff are encouraged to use pronouns for people that they use for themselves as a sign of respect and inclusion. Students are able to update and edit their pronouns in UAccess.

More information on updating your preferred name and pronouns is available on the Office of the Registrar site at https://www.registrar.arizona.edu/.

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

Confidentiality of Student Records

http://www.registrar.arizona.edu/ferpa

Land Acknowledgement Statement

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O'odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

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.