CSC 380: Principles of Data Science

Location and Times: TuTh, 5:00 – 6:15pm, Modern Languages, Rm 311
Web: http://www.pachecoj.com/courses/csc380_fall21/

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 or Co-requisites

Major: COSC. CSC 210, CSC 245

Instructor and Contact Information

Prof. Jason Pacheco
pachecoj@cs.arizona.edu

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 email advising@cs.arizona.edu for academic advising related questions.
- **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 visit our CS Tutoring Page. **[NOTE: Tutor Center is closed in Summer; delete this for summer courses]**
- **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 520-621-2057 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.

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

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

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 Designee) will be honored. See: https://deanofstudents.arizona.edu/absences

Course Communications

All communications will be conducted via Piazza. Assignments will be distributed and submitted via D2L. Office hours with the professor will be conducted remotely, over Zoom.

Required Texts or Readings

There will be no required text for this course. All required readings will be made available on the course webpage.

Required or Special Materials

None

Required Extracurricular Activities

None

Assignments and Examinations: Schedule/Due Dates

	Description	Weight	Due	Graded
HW1	Probability	6	9/9	9/15
HW2	Statistics	6	9/16	9/22
HW3	Data Collection and Exploratory Analysis	7	9/23	9/29

HW4	Data Processing and Visualization	7	9/30	10/4
Midterm		20	10/5	10/27
HW5	Predictive Models	6	10/28	11/3
HW6	Linear Supervised Learning Models	7	11/4	11/15
HW7	Nonlinear Supervised Learning Models	7	11/16	11/22
HW8	Unsupervised Learning	7	11/23	12/6
HW9	Model Assessment	7	12/7	12/14
Final		20	12/15	12/18

Final Examination

This course will have a final examination on Wednesday, December 15, 6:00-8:00pm.

The date and time of the final exam or project, along with links to the 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

Grading Scale and Policies

Assignments: 60%
Midterm: 20%
Final Exam: 20%

Grade Distribution for this Course:
A: 90%
B: 80%
C: 70%
D: 60%
E: 59% and below

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

Late assignments receive a grade of zero.

All exams will be closed-book.

Without prior arrangement, missed exams result in a grade of zero.

All grading will be done within one week of the assignment due date.

Department of Computer Science Grading Policy:

Instructors will explicitly promise when every assignment and exam will be graded and returned to students. These promised dates will appear in the syllabus, associated with the corresponding due dates and exam dates.

Graded homework will be returned before the next homework is due.

Exams will be returned "promptly", as defined by the instructor (and as promised in the syllabus).

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.

Scheduled Topics/Activities

Week 01: Course mechanics, introduction to data science
Week 02: Applied Probability HW1 Assigned
Week 03: Applied Statistics HW1 Due, HW2 Assigned
Week 04: Data Collection and Exploratory Analysis HW2 Due, HW3 Assigned
Week 05: Data Processing HW3 Due, HW4 Assigned
Week 06: Data Visualization HW4 Due
Week 07: Midterm + Introduction to Machine Learning
Week 08: Linear Algebra Primer
Week 09: Predictive Models HW5 Assigned
Week 10: Supervised Learning (Linear Models) HW5 Due, HW6 Assigned
Week 11: Supervised Learning (Nonlinear Models) HW6 Due, HW7 Assigned
Week 12: Unsupervised Learning (Clustering)
Week 13: Unsupervised Learning (Dimensionality Reduction) HW7 Due, HW8 Assigned
Week 14: Model Assessment HW8 Due
Week 15: Data Science Ethics HW9 Assigned
Week 16: Courwe Wrapup HW9 Due

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.

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: http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity.

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

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.