CSC 580: Principles of Machine Learning

MW 3:30pm-4:45pm, Gould-Simpson 701

Description of Course

Students will learn why machine learning is a fundamentally different way of writing computer programs from traditional programming, and why this is often an attractive way of solving practical problems. Machine learning is about automatic ways for computers to collect and/or adapt to data to make better predictions and decisions or gain insight; students will learn both advantages and unique risks that this approach offers. They will learn the fundamental frameworks, computational methods, and algorithms that underlie current machine learning practice, and how to derive and implement many of them.

Course Prerequisites

- Linear algebra or equivalent
    - You will need to understand the relationship between linear operators, linear transformations, change of bases, and matrices
    - We will make repeated use of matrix decompositions such as the SVD. Often we will need to use properties of the eigendecomposition of a matrix.
- Multivariate calculus or equivalent:
    - You will need to understand the relationship between the total derivative, the gradient, and how to take advantage of the fact that the derivative is a linear operator
- Probability and statistics:
    - You will need to understand (conditional) expectation and (conditional) independence of random variables.
- Programming
    - You will need a good amount of programming experience: a programming maturity is expected.

Instructor and Contact Information

- Chicheng Zhang
- chichengz@cs.arizona.edu
- Gould-Simpson 720

- Office Hour: TBD or by appointment.
- Course homepage: https://zcc1307.github.io/courses/csc588sp22/
- Gradescope: https://www.gradescope.com/courses/421111, entry code: XVXNBZ
- D2L: https://d2l.arizona.edu/d2l/home/1196047

Course Format and Teaching Methods

In-person lectures, individual assignments, written exams, projects, in-class/online discussions.

Obtaining Help

- **Advising:** If you have questions about your academic progress this semester, or your chosen

degree program, consider contacting your graduate program coordinator and faculty advisor. Your program coordinator, faculty advisor, and the Graduate Center can guide you toward university resources to help you succeed.**Computer Science students** are encouraged to email gradadvising@cs.arizona.edu for advising related questions.

- **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.

Class Recordings

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

A successful student will be able to implement and explain the limitations of many of the central methods and techniques in machine learning:

- Basic binary classifiers: decision trees, logistic regression
- Supervised vs. unsupervised learning - what's possible in the absence of labels
- Reductions - how to handle imbalanced data; how to build multiclass classifiers
- Practical issues - how to detect overfitting and underfitting; how and when to use feature engineering
- Efficiency issues - how to create classifiers that work well in the presence of large training sets, and large feature sets
- Modern techniques - students will be introduced, via classroom materials and projects, to recent methods in machine learning (this could include, for example, deep learning, reinforcement learning, A/B testing, and multi-armed bandits)

For a more granular description of the learning objectives, see the week-by-week schedule and the description of the assignments below.

Machine Learning is a big field, and there is no way we can cover all of it in one course. With that said, this course covers a large amount of material, and the assignments are a central part of the course. **Students are expected to dedicate a significant amount of time on the course outside of the classroom, especially if they have background deficiencies to make up**.

Expected Learning Outcomes

The expected learning outcomes of the course is:

- To be able to explain why supervised learning is expected to generalize.
- To be able to detect overfitting and underfitting.
- To be able to explain the risk of test set reuse.
- To be able to compute the error bar and discuss statistical significance of a given evaluation result.
- To be able to list the common supervised and unsupervised learning algorithms and the practical values of each.
- To be able to explain the key challenge in reinforcement learning that supervised/unsupervised learning paradigms did not have to worry about.

Absence and Class Participation Policy

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 activities, please contact me as soon as possible. Students struggling to follow the absence and class participation policies are encouraged to speak with the Graduate Program Coordinator.

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.

Statement on compliance with COVID-19 mitigation guidelines: As we enter the semester, your and my health and safety remain the university's highest priority. To protect the

health of everyone in this class, students are required to follow the university guidelines on COVID-19 mitigation. Please visit www.covid19.arizona.edu.

Makeup Policy for Students Who Register Late

If you register late for this class, contact me as soon as you do. You will be expected to submit all missed assignments within a week of your registration. It is your responsibility to catch up to the class content.

Course Communications

We will use D2L for communications and discussion. Make sure your D2L account is up to date and your email address on D2L is reachable - class announcements are sent through the website.

Required Texts and Materials

The required textbook is Hal Daumé's Course in Machine Learning (http://ciml.info/), fully and freely available online.

Scheduled Topics/Activities

Week	Date	Description
1	8/22	Lecture: Introduction, motivation, course logistics *HW0: Calibration homework*
	8/24	Lecture: Basics - Decision Trees, algorithms for learning Learning Objectives. Explain the difference between memorization and generalization • Implement a decision tree classifier • Take a concrete task and cast it as a learning problem, with a formal notion of input space, features, output space, generating distribution and loss function.
2	8/29	Lecture: Limits - Optimal Bayes rate and classifier; overfitting and underfitting Learning Objectives. Define "inductive bias" and recognize the role of inductive bias in learning • Illustrate how regularization trades off between underfitting and overfitting • Evaluate whether a use of test data is "cheating" or not.
	8/31	Lecture: Geometry, nearest-neighbor classifiers, k-means (unsupervised learning preview) Learning Objectives • Describe a data set as points in a high dimensional space • Explain the curse of dimensionality •

		Compute distances between points in high dimensional space • Implement a K-nearest neighbor model of learning • Implement the K-means algorithm for clustering. *HW1: Decision Trees, Naive nearest neighbors, k-means*
3	9/5	Labor day holiday
	9/7	Lecture: The perceptron (1/2) Learning Objectives. Describe the biological motivation behind the perceptron • Classify learning algorithms based on whether they are error-driven or not • Implement the perceptron algorithm for binary classification • Draw perceptron weight vectors and the corresponding decision boundaries in two dimensions • Contrast the decision boundaries of decision trees, nearest neighbor algorithms and perceptrons • Compute the margin of a given weight vector on a given data set.
4	9/12	Lecture: The perceptron (2/2)
	9/14	Lecture: Practical Issues (1/2) - performance measures, underfitting, overfitting, cross validation, prediction confidence via statistical tests and bootstrapping, debugging ML models Learning Objectives. Translate between a problem description and a concrete learning problem • Perform basic feature engineering on image and text data • Explain how to use cross-validation to tune hyperparameters and estimate future performance • Compare and contrast the differences between several evaluation metrics.
5	9/19	Lecture: Practical Issues - (2/2) *HW2: Perceptron and Feature Selection*
	9/21	Lecture: Bias-variance decomposition, and friends Learning Objectives. Understand how classification errors naturally split in approximation error and estimation errors • Understand how error decompositions are useful for debugging.
6	9/26	Lecture: Reductions (1/3) Learning Objectives. Represent complex prediction problems in a formal learning setting • Be able to artificially “balance”

		imbalanced data • Understand the positive and negative aspects of several reductions from multiclass classification to binary classification • Recognize the difference between regression and ordinal regression.
	9/28	Lecture: Reductions (2/3)
7	10/3	Lecture: Reductions (3/3)
	10/5	Lecture: Linear Models (1/2) Learning Objectives. Define and plot four surrogate loss functions: squared loss, logistic loss, exponential loss and hinge loss • Compare and contrast the optimization of 0/1 loss and surrogate loss functions • Solve the optimization problem for squared loss with a quadratic regularizer in closed form • Implement and debug gradient descent and subgradient descent
8	10/10	Lecture: Linear Models (2/2)
	10/12	Midterm exam
9	10/17	Lecture: Kernel Methods (1/2) Learning Objectives. Explain how kernels generalize both feature combinations and basis functions • Contrast dot products with kernel products • Implement kernelized perceptron • Derive a kernelized version of regularized least squares regression • Implement a kernelized version of the perceptron • Derive the dual formulation of the support vector machine.
	10/19	Lecture: Kernel Methods (2/2) *Project proposal due*
10	10/24	Lecture: Probability, Naive Bayes, and Graphical Model Basics (1/3) Learning Objectives. Define the generative story for a naive Bayes classifier • Derive logistic loss with an l2 regularizer from a probabilistic perspective.
	10/26	Lecture: Probability, Naive Bayes, and Graphical Model Basics (2/3)

11	10/31	Lecture: Probability, Naive Bayes, and Graphical Model Basics (3/3)
	11/2	Lecture: Bias and Fairness Learning Objectives. Identify how disparity along training/test data can generate bias/unfairness • Understand how a bad choice of metric to optimize can cause bias/unfairness • Identify how careless data collection practices can perpetuate bad decisions • Identify how different assumptions about the world change the way data should be processed for an ML method • Understand how feedback loops can cause arbitrarily bad predictions *HW3: Reduction / Linear Models, Kernel Methods / Probability*
12	11/7	Lecture: Neural Networks and Back-Propagation (1/2) Learning Objectives. Explain the biological inspiration for multi-layer neural networks • Construct a two-layer network that can solve the XOR problem • Implement the back-propagation algorithm for training multi-layer networks • Explain the trade-off between depth and breadth in network structure • Contrast neural networks with radial basis functions with k-nearest neighbor learning.
	11/9	Lecture: Neural Networks and Back-Propagation (2/2)
13	11/14	Lecture: Ensembling Learning Objectives. Implement bagging and explain how it reduces variance in a predictor • Explain the difference between a weak learner and a strong learner • Derive the AdaBoost algorithm • Understand the relationship between boosting decision stumps and linear classification.
	11/16	Lecture: Computational Efficiency in Machine Learning Learning Objectives. Understand and be able to implement stochastic gradient descent algorithms • Compare and contrast small versus large batch sizes in stochastic optimization • Derive subgradients for sparse regularizers • Implement feature hashing. *HW4: Ensembling and Efficiency*
14	11/21	Lecture: Unsupervised Learning (1/2)

		Learning Objectives. Explain the difference between linear and non-linear dimensionality reduction • Relate the view of PCA as maximizing variance with the view of it as minimizing reconstruction error • Implement latent semantic analysis for text data • Motivate manifold learning from the perspective of reconstruction error • Understand K-means clustering as distance minimization • Explain the importance of initialization in k-means and furthest-first heuristic • Implement agglomerative clustering • Argue whether spectral clustering is a clustering algorithm or a dimensionality reduction algorithm.
	11/23	Lecture: Unsupervised Learning (2/2)
15	11/28	Lecture: Learning Theory (1/2) Learning Objectives. Explain why inductive bias is necessary • Define the PAC model and explain why both the "P" and "A" are necessary • Explain the relationship between complexity measures and regularizers • Identify the role of complexity in generalization • Formalize the relationship between margins and complexity
	11/30	Lecture: Learning Theory (2/2)
16	12/5	Lecture: Multi-armed bandits and reinforcement learning Objective: • Understand the Multi-armed bandit learning problem and the exploration-exploitation dilemma • Understand the intuition of optimism-based learning algorithm • Understand the Markov decision process model. Understand the exploration challenges of reinforcement learning. • Define the state and state-action value function • Understand Bellman evaluation equation and Bellman optimality equation.
	12/7	(Reserved for Catch-up)
17	12/13	**Final exam at 3:30-5:30pm**
	12/15	**Project final report due**

18		

Assignment due dates:

- HW0 is due in 7 days. HWs 1-4 are due in 10 days.
- 10/21: Project proposal due
- 12/13: Final exam at 3:30pm - 5:30pm
- 12/15: Final project due

Final Examination or Project

The final exam will take place on Tuesday, Dec 13, 3:30pm-5:30pm at Gould-Simpson 701.

Links to the Final Exam Regulations and Final Exam Schedule:
https://registrar.arizona.edu/faculty-staff-resources/room-class-scheduling/schedule-classes/final-exams

Grading Scale and Policies

As mentioned above, you will be assessed based on your performance on assignments (written and programming), one final exam, and one project.

The instructing staff will assign grades on a scale from 0 to 100, with the following weights:

- Assignments: 40%
- Project: 20%
- Midterm Exam: 15%
- Final Exam: 15%
- Participation: 10%

Your final grade in the course of be the best of a per-class grading curve and overall performance:

- 90% or better: A;
- 80% or better: B;
- 70% or better: C;
- 60% or better: D;
- below 60%: E.

HW0 will not be part of the homework evaluation but will be part of the participation score as it serves as information on the students' background (the participation score will be deducted if the student's submission does not show nontrivial effort to solving it). There will be a total of 5 math/programming assignments and 1 project proposal assignment (total 6). Three of them (total 30%) and the project proposal (5%) along with participation (5%) will provide 40% of the grade by the final drop date.

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

Graded homework will be returned before the next homework is due. For due dates, see "Assignments and Examinations". The homework will be returned to students before the next homework is due. Grading delays beyond promised return-by dates will be announced as soon as possible with an explanation for the delay. As a rule, homework will not be accepted late except in case of documented emergency or illness.

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: If you wish to dispute your grade for an assignment, you have two weeks after the grade has been turned in. In addition, even if you only dispute one portion of the grading for that unit, I reserve the right to revisit the entire unit (assignment, midterm, or project).

Bibliography

We recommended the following texts for optional reading:

Moritz Hardt and Benjamin Recht, Patterns, predictions, and actions: Foundations of machine learning, 2022 (https://mlstory.org/)

Shai Shalev-Shwartz and Shai Ben-David, Understanding machine learning: from theory to algorithms (https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/)

Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The elements of statistical learning (2nd edition).

Tom Mitchell, Machine Learning, 1997.

Michael Kearns and Umesh Vazirani, Introduction to Computational Learning Theory, 1997.

Kevin Murphy, Machine learning: a probabilistic perspective.

Richard Duda, Peter Hart, and David Stork, Pattern classification (2nd edition).

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.

Some learning styles are best served by using personal electronics, such as laptops and iPads. These devices can be distracting to other learners. Therefore, students who prefer to use electronic devices for note-taking during lectures should use one side of the classroom.

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 https://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.

Pronouns and Preferred Names

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. The instructor (Chicheng Zhang) has a pronoun of he/him/his. 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/

Department of Computer Science Evacuation Plan for Gould-Simpson:
https://drive.google.com/file/d/1iR1IcGcV_BgbGnEFBzZ2-do0FbLC3cvo/view?usp=sharing

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/personal-information/family-educational-rights-and-privacy-act-1974-ferpa?topic=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.