Introduction to Machine Learning (CS 412) [1]
Fall 2022
3 credit hours (undergraduates) or 4 credit hours (graduates)
Prerequisites: CS 251; and IE 342 or STAT 381; or consent of instructor

Course Modality: In person only

Instructor:	**Prof. Xinhua Zhang** (zhangx@uic.edu)
Teaching Assistant:	**Siteng Kang** (skang98@uic.edu)
Lecture time:	**3:30-4:45 PM, Tuesday and Thursday**
Lecture venue:	**SES 130**
Office hours: (**Zoom + In-person**)	Instructor: 12-1 PM, Tuesday (North End, Level 3, Daley Library) TA: 11-12 AM, Thursday (3-190 Daley Library, Level 3)
Recitation (exercises)	**2-3 PM Mondays** on Zoom (recorded, attendance **not** required)

Course links (bookmark whenever appropriate):

Piazza	https://piazza.com/class/l6pouhg0a7945t You can register using your UIC netid with no need of instructor's approval.
Master Schedule	https://piazza.com/class/l6pouhg0a7945t/post/6 First register on Piazza as above. This schedule is pinned on top of the left list.
Gradescope	https://www.gradescope.com/courses/417726 All your grades and feedback will be here. You can log in with School Credentials or via Blackboard (left banner).
Zoom	https://uic.zoom.us/j/83380974884?pwd=Tm0zZVhJejQydnRnR1Y2dHB2dTBudz09 If needed, the password is: Kb4B0HHj The meeting ID is: 833 8097 4884 This is for recitation and office hour. In addition to in-person visits, you can also join **office hours** by Zoom, for both the instructor **and** the TA. First come first served on a joint queue.
Blackboard	Announcement only
Onenote for recitation	https://uic365-my.sharepoint.com/:o:/g/personal/zhangx_uic_edu/EqUqPDgm4F9Hs4ou2mj01_EB8QrsMCPcDF5XYEhqUmU7aQ
Google Drive	https://drive.google.com/drive/u/1/folders/1WWQYi3SJHt1zKVTxizl-9dRmWrEs8duc Questions of assignments and labs, along with rubrics and solutions. To access it, make sure you log in with your UIC netid (**not** your personal account).

---

[1] The course has two CRNs: 43875, 43876. They have been merged on Blackboard. So just look for CS 412 and ignore the CRN.

Use of Online Tools (see the URLs on the first page)

What	Where	Who can access	
		Week 1 and 2	Week 3 and onwards
Slides and other documents	Piazza	Anyone[1]	Registered + Auditing
Technical discussion, Q&A	Piazza [2, 3]	Anyone	Registered + Auditing
Announcement	Blackboard [4]	Registered + Auditing	Registered + Auditing
Echo 360 of lecture recordings (also streaming in real time) [5]	Blackboard	Registered + Auditing	Registered + Auditing
Office hour (not recorded)	**Zoom + In person**	Anyone	Anyone
Recitation (recorded)	**Zoom**	Registered + Auditing	Registered + Auditing
Assignment submission + grading	Gradescope	N.A.	Registered only
Whiteboard for sync sessions	Onenote	Anyone	Registered + Auditing
Assignment questions + solutions	Google Drive	Anyone	Registered only
iClicker	App	N.A.	Registered + Auditing

for collaboration within groups (but not for instruction)		
Teams, Slack	Github, Bitbucket	Box.com, Onedrive, Google drive

1. Anyone refers to anyone with a UIC netid. Contact the instructor if you do not have it.

2. Piazza is highly catered to getting you help fast and efficiently from classmates and the instructing team (professor and TA). Rather than emailing questions to the instructing team, you are encouraged to post on Piazza your **technical questions**, general questions about the course content, assignments, grading rubric, etc. If you have any problems or feedback for the Piazza developers, email team@piazza.com.

3. If you have any personal or **non-technical questions** such as medical considerations, please send an email directly to the instructor (zhangx@uic.edu) or the TA (if instructed to do so, e.g., dispute plagiarism suspect). Re-grading of assignments should be requested on Gradescope (not email).

4. For general announcements and notifications, I will send emails to the whole class via **Blackboard**. Please check your email frequently, especially around deadlines (homework and exam). The message will also be recorded on Blackboard in the announcement section.

5. We strongly recommend attending the lectures in person. The online option is only for those we really cannot make it due to health reasons, etc. In any case, we will use iClicker for quick quizzes at some random time of each lecture. This will ensure constant engagement.

It is your responsibility to check emails frequently (at least once a day).

Software and hardware

1. Python for labs. You can use any language for the course project.

2. You are required to use collaborative tools for labs and course project, including

   a. Slack or Teams for messaging. UIC does not have license for slack, but the free version might be sufficient.

   b. Github or Bitbucket for code maintenance (free with their respective education packs).

   c. Box, Onedrive, or Google-cloud for file sharing (all free from UIC).

3. Scanner on phone/pad: scan your answer, **NOT just take a photo**. Use a scanner app on your mobile device, e.g., **CamScanner** (https://www.camscanner.com/). Register using your UIC email address and it will remove a watermark. Feel free to use similar apps.

4. Time management by Google calendar, Outlook, etc.

5. Keep the iClicker app on your computer or mobile device up to date at spare time.

Eligibility

CS 251, IE 342 or STAT 381, or consent of instructor.

To ensure that all students understand the math required in course, a self-evaluation will be posted on Piazza, along with some background math readings.

How much linear algebra is needed?
Linear algebra is *not* a prerequisite because we will only use a fraction of it.
Make sure you can understand
1) Sections 2.1, 2.2, 2.3, 2.5, 2.6 of [GBC] (see textbook below)
2) Section 4.1 of this book up to Equation 4.8.

If you would like to audit, please send an email to the instructor (zhangx@uic.edu).

Course Description

Machine learning is the study of systems that improve automatically based on past experience. This course will introduce common machine learning tasks (e.g., classification, density estimation, clustering) and some of the successful machine learning techniques and broader paradigms that have been developed for these tasks. The course is programming- intensive and a large emphasis will be placed on tying machine learning techniques to specific real-world applications through hands-on experience.

Roadmap: see the master schedule on Piazza (link available on the first page of this syllabus).

Course Materials:

Required Texts (available electronically via UIC library or web)

Ethem Alpaydin, *Introduction to Machine Learning*, 3rd edition. MIT Press, 2014.

Reference:

[GBC] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. MIT Press, 2016.

[Murphy] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. (also available from UIC library)

Besides lecture notes, there will also be four tutorials. They are meant to help you understand the subject and work out the assignments. It only for reference; nothing needs to be submitted for them. The scope of the final exam will be the union of lectures, assignments, and tutorials. Some questions in the tutorial are explicitly marked as "only to understand, not a requirement to derive by yourself"; they won't appear in the exam.

You will need to use numpy a lot in the labs. Here is a numpy primer: Python Data Science Handbook, covering numpy, Pandas, Matplotlib. You should at least know that y = xMatrix[0] is a shallow copy, where xMatrix is a 2-D numpy array. Understand how to make a deep copy. The book provides many notebooks for learning. You can create your Jupyter notebook to run on Google Cloud, or locally on your own machine via VS Code and Anaconda (as opposed to directly downloading Python from the official site). Anaconda is the choice for data science. Data used in this lab has been processed for your convenience.

Evaluation Criteria

The understanding of course material and synthesis of knowledge will be evaluated using

1. In-class quizzes conducted by iClicker. Each class may have multiple quiz questions that you will need to answer on the iClicker app.

2. Eleven (11) light-weight individual written assignments will be due approximately every week. They focus on conceptual understanding. For simplicity, I will henceforth just call them "assignments". Graduate students will have slightly more questions to solve.

3. A course project. Groups will be formed to conduct a course project, with **four (4)** students in each group. The project, which is open-topic, aims to apply machine learning techniques to "real world" datasets. It will have a **proposal** due in week 10, an **oral presentation** in the lectures of week 15, and a **report** due in week 15.

   Each group can have a mix of graduate and undergraduate students. You can find teammates by going to Piazza, and a random grouping will be organized as the last resort. To find the email address (i.e., netid) by first and last name, try https://www.uic.edu/apps/find-people.

4. Seven (7) labs, due every 2 weeks. The language will be **Python** only. The topics and tasks will be given. Lab 1 is for individual work and is mandatory.

   For lab 2-7, you can either work individually or in groups (same groups as in the course project). To submit as a group, just specify your teammates on Gradescope. Undergraduates are expected to submit 5 individual or group labs (lab 1 included), and graduate students are expected to submit 6 individual or group labs. By expectation, it means how the total score of the entire course is calculated, based on which 85% will be A, etc; see below. So effectively, undergraduates have two labs to earn extra marks from, while graduates have one.

5. A 75-min **mid-term** exam and a 120-min **final** exam.

Partial grading will be used in all assessments.

Grading Policy. The overall course grade of each student will be determined based on:

1. (4%) In-class iClicker quiz from week 3 to 14: 10 points * 23 lectures
2. (21%) Eleven (11) written assignments: 100 points each for undergraduates, and there are five assignments of 160 points for graduate students.
3. (19%) Seven (7) labs: 200 points each. Undergraduate and graduate students are expected to submit 5 and 6 labs, respectively.
4. (15%) Course project: 800 points
5. (15%) Midterm exam: 800 points for undergraduates and 880 points for graduates.
6. (26%) Final exam: 1400 points for undergraduates and 1540 points for graduates.

The percentages in parenthesis are computed based on undergraduate's points. The total earned points will be used to determine the final letter grade (A/B/…) **based on threshold**:

	Total points	A (85%)	B (70%)	C (55%)	D (40%)	F (<40%)
Undergraduate	X = 5330	4530.5	3731	2931.5	2264	0
Graduate	Y = 6050	5142.5	4235	3327.5	2576	0

$X = 23*10 + 11*100 + 5*200 + 800 + 800 + 1400 = 5330$
$Y = 23*10 + 6*100 + 5*160 + 6*200 + 800 + 880 + 1540 = 6050$

Important Note: In order to pass the course, undergraduate students must earn at least **1100** points from midterm and final exams combined, while graduate students must earn at least **1210** points.

I reserve the right to **raise** your letter grade.

Course Policies

Academic dishonesty will not be tolerated. Please see the CS department policy below on the topic; this policy specifies penalties for violations. Academic misconduct will be handled according to UIC's Student Disciplinary Policy: http://dos.uic.edu/conductforstudents.shtml

What is academic dishonesty? To hand in any work which is not 100% the student's creation, unless you are explicitly allowed to do so.

Collaboration Policy

1. Written assignments: Discussion of homework assignments and solutions with other students is permitted. However, each student must submit his or her own write-up for the homework assignment and fully understand what he or she submits. Each student should be prepared to explain his or her homework assignment submissions to the instructor and teaching assistant if his or her "full understanding" is in doubt.

2. **Course project and labs:** all members of each group should make nearly the same amount of contribution to each lab and project. All members should be responsible for the whole submission of the team, not only his/her own contributed part. You are supposed to understand the work of your teammates inside out and be able to answer questions when asked. If one member plagiarized, then **all members** of the team will receive the **same** penalty. Find teammates for experiment & course project by "Search for Teammates!" on Piazza.

   For computer programs, if we cannot determine which team copied from which, we may, at our discretion, give failing grades to both groups. It is the responsibility of all engineering

and computer science professionals to safeguard their company's "trade secrets." An employee who allows trade secrets to be obtained by competitors will almost certainly be fired. So, YOU are responsible for making sure that your directories have permissions set so that only you can read your files, for being sure to log out at the end of working in the computer lab, etc.

3. **Exams: All work on the midterm and final exams must be individually performed.**

Policy for Missed or Late Evaluation

1. **Late submission**s: Unless specified otherwise, all deadlines will be Chicago time. Late submissions will not be accepted in any case, unless there is a **documented** personal emergency. Arrangements must be made with the instructor as soon as possible after the emergency arises, preferably well before the homework due date.

   Advice: If for whatever reason you don't manage to finish an assignment, hand in what you have. Partial credit will be given.

2. **Statute of limitations**: **Three weeks!** No grading questions or complaints — no matter how justified — will be listened to **three** weeks after the grade in question has been released.

3. **Missed exams**: Missed exams cannot be made up unless there are extenuating circumstances (death, severe illness, etc.) **and** the student has e-mailed instructor **in writing <u>before</u>** the exam. A mark of 0 is earned for the exam if the above policy is not adhered to.

Religious Holidays

Students who wish to observe their religious holidays shall notify the faculty member by the tenth day of the semester of the date when they will be absent. The faculty member shall make every reasonable effort to honor the request, not penalize the student for missing the class, and if an examination or assignment is due during the absence, give the student an exam or assignment equivalent to the one completed by those students in attendance. If the student feels aggrieved, he/she may request remedy through the campus grievance procedure.
http://oae.uic.edu/docs/ReligiousHolidaysFY20152017.pdf

CS department policy on academic dishonesty

The CS Department will not tolerate cheating by its students. The MINIMUM penalty for any student found cheating will be to receive an F for the course and to have the event recorded in a department and/or College record. The maximum penalty will be expulsion from the University.
Cheating includes all the following, though this is not a complete list:

- Copying or any other form of getting or giving assistance from another student during any test, quiz, exam, midterm, etc.
- Plagiarism–turning in writing that is copied from some other source.
- Obtaining solutions to homework by posting to the Internet for assistance, purchasing assistance, obtaining copies of solutions manuals for instructors, and obtaining copies of previous year's homework solutions.
- Computer programs: Any time you look at another student's code, it is cheating. (Exception: If you are EXPLICITLY told that you may do so by the instructor.)