

Introduction to Machine Learning (CS 412, CRNs: 43875, 43876¹)
Fall 2020

3 credit hours (undergraduates) or 4 credit hours (graduates)
Prerequisites: CS 251; and IE 342 or STAT 381; or consent of instructor

Course Modality

Online only, no in-person meeting

- Asynchronous (recorded lectures) + synchronous sessions (**5-6:15 PM**, Tuesday, Chicago time).
- Synchronous sessions will include recitation and Q & A. You are strongly encouraged to attend (join by Zoom), but attendance is **NOT mandatory**. It will be recorded and shared.

Instructor:	Prof. Xinhua Zhang, zhangx@uic.edu , (312) 413-2416			
Teaching Assistants:	Yingyi Ma (yma36@uic.edu) Mao Li (mli206@uic.edu) Ye Liu (yliu279@uic.edu)			
Synchronous sessions:	5 - 6:15 PM, Tuesday (on Zoom , see below)			
Office hours: (all on Zoom, not recorded may adjust slots, TA office hour from week 2)	Tuesday	Thursday	Saturday	
	1-2 PM	Mao Li (TA)	Instructor Yingyi Ma (TA)	Instructor
	10-11 PM			

Notes on registration:

1. The Spring 2021 semester will *probably* still be online [[news](#)]. I suggest you take this course in Fall 2020, as it may also help you find a summer internship.
2. **Non-CS students** can register at any time. Please send an email to cs-grad@uic.edu to get registration approval. It will be approved.
3. If you are unable to register on my.uic.edu due to **conflicts** arising from the *nominal* meeting time of 5-6:15 PM (Tu and Th), please send an email to cs-grad@uic.edu to receive a time conflict override in the system.

Course links (bookmark whenever appropriate):

Piazza	https://piazza.com/uic/fall2020/2020fallcs41243875 You can register using your UIC netid with no need of instructor's approval. Please complete the poll of time zone there as soon as possible.
Master Schedule	https://piazza.com/class/kducejptcqius?cid=7 (bookmark it as an entry point!) First register on Piazza as above. This schedule is pinned on top of the left list.

¹ The two sections have been merged on Blackboard and everywhere else. So graduate students can also follow 43875 if 43876 cannot be found. Just look for CS 412 and ignore the CRN.

Panopto	List.aspx?folderID=df9bd589-e913-4f84-8135-ac1700efd437">https://uic.hosted.panopto.com/Panopto/Pages/Sessions>List.aspx?folderID=df9bd589-e913-4f84-8135-ac1700efd437 (tip: you can watch at 1.25x speed if you like)
Gradescope	https://www.gradescope.com/courses/160760 (all your grades will be here) Most likely, you can't log in through this link directly. Instead, go to Blackboard, then see left column. Only registered students need it after week 2.
Zoom	https://uic.zoom.us/j/95361746472?pwd=K01VTE5hOWRIQUt6TjhiaWR2OFRyUT09 We will use this link for sync sessions and office hours (instructor's and TAs').
VoiceThread	Go to Blackboard, then see left column.
Blackboard	Announcement
Google Drive	Questions, rubrics, and solutions to Assignments and Labs
Google doc	Just to form groups. Link will be shared in week 3 after registration is finalized.

Use of Online Tools (see URLs on the first page, or just click)

What	Where	Who can access	
		Week 1 and 2	Week 3 and onwards
Asynchronous lecture video	Panopto	Anyone ¹	Registered + Auditing
Slides and other documents	Piazza	Anyone	Registered + Auditing
Technical discussion, Q&A	Piazza ^{2,3}	Anyone	Registered + Auditing
Announcement	Blackboard	Anyone	Registered + Auditing
Synchronous session with recording	Zoom ⁵	Anyone	Registered + Auditing
Office hours (not recorded)	Zoom	Anyone	Registered + Auditing
Assignment + Grade + Solution	Gradescope	N.A.	Registered only
Video self-introduction (week 1)	VoiceThread	Anyone	N.A.
Course project presentation	VoiceThread	N.A.	Registered only
Exam proctoring (to be determined)	Zoom	N.A.	Registered only

for collaboration within groups (but not for instructions)

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.

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

5. Both the audio and the video of the synchronous sessions will be recorded. Your video will NOT be record if you do not talk. So turn off your camera if you have privacy concerns.

Software and hardware

1. Python for Lab 1-3. You can use any language for the course project.
2. Matlab for some illustrations from the textbook. You can get free Matlab from Webstore <https://webstore.illinois.edu/shop/product.aspx?zpid=3819>.
3. 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).
4. 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 and it will remove a watermark.
5. Time management by Google calendar, Outlook, etc.
6. Keep the software on your compute or mobile device up to date at spare time.
7. Get a reliable internet connection at home (e.g., by using wired connection, closer to wifi router).
8. Get a mobile device (phone or pad), and a good headphone with a good mic. Test them before the semester and buy a new one if not good. A wired headset should be good enough.

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.

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.

Evaluation Criteria

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

1. About twelve (12) 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.
2. 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 on Oct XX), a **report** (due in week 15, and maximum 6 pages with template provided), and a video **oral presentation** (due in week 14, with peer review and rebuttal for it in week 15).
3. Three (3) labs, one for individual work and two for group work (same groups as in the course project). The language will be **Python** only. The topics and tasks will be given.
4. A final examination which will last 2 hours.

In both 2 and 3, there will be no extra workload for graduate students. So it is fine to form groups that mix graduate and undergraduate students. You can find group mates by going to

<https://piazza.com/class/kducjptcqiusy?cid=5>.

As organizing an exam can be highly complicated at this trying time, there won't be a midterm exam.

Partial grading will be used in all assessments.

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

1. Written assignments: 30%
2. Labs: 25%
3. Course project: 15%
4. Final exam: 30%

All the evaluations, including each of the 12 assignments, will be graded out of 100, and will be allocated a weight (e.g., 3% for assignment 4, 9% for Lab 3, etc).

Their **weighted average** will be used to determine the final letter grade (A/B/...) **based on threshold**:

- A: 80 - 100
B: 70 ~ 79
C: 60 ~ 69
D: 0 ~ 59

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

Important Note: Your final exam score must be above 50 in order to pass the course.

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 experiments:** all members of each group should make nearly the same amount of contribution to each experiment 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. Final exam: All work on the final exam must be individually performed.

Policy for Missed or Late Evaluation

1. **Late submissions:** All deadlines will be AOE (anywhere on earth). So if it is Oct 25 (AOE), then you can submit it as long as anywhere on earth is still Oct 25. 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 item in question has been returned.
3. **Missed exams:** Missed (final) exams cannot be made up unless there are extenuating circumstances (death, severe illness, etc.) **and** the student has e-mailed instructor **in writing before** 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.)

A (Very) Tentative weekly schedule

See a detailed schedule at <https://piazza.com/class/kducjptcqiusy?cid=7>

Week	Topic	Reading
1	Tour of IT resources	
1	Tour of syllabus	
1	Introduction	Ch 1 of Alpaydin
1, 2	Basic linear algebra, probability and statistics	Ch 2.1-2.3 of Murphy Ch 2.1-2.3 of GBC Statistics for ML Math for ML
2, 3	Supervised learning	Ch 2 of Alpaydin
3	Experiment design with cross validation	Ch 19 of Alpaydin Ch 18.4-18.4.1 of AIMA [PDF]
4	Parametric methods for classification	Ch 4.1-4.5 of Alpaydin
4, 5	Naïve Bayes	Ch 5.7 of Alpaydin Ch 3.5.1, 3.5.2 of Murphy

5	Bayesian estimation applied to naïve Bayes	Ch 16.1, 16.2 of Alpaydin
6	Course project introduction	
6, 7	Multivariate methods	Ch 5 of Alpaydin
8	Clustering	Ch 7 of Alpaydin
9	Nonparametric methods	Ch 8 of Alpaydin
10	Decision trees	Ch 9 of Alpaydin
11	Linear discrimination	Ch 10 of Alpaydin
12	Kernel machines	Ch 13 of Alpaydin
13	Combining multiple learners	Ch 17 of Alpaydin
14	Multilayer perceptrons	Ch 11 of Alpaydin
15	Convolutional networks	Ch 9 of GBC