STA 314: Statistical Methods for Machine Learning I

Overview

Machine learning (ML) is a set of techniques that allow computers to learn from data and experience, rather than requiring humans to specify the desired behaviour by hand. ML has become increasingly central both in statistics as an academic discipline, and in the data science industry. This course provides a broad introduction to commonly used ML methods, as well as the key statistical concepts underlying ML. It serves as a foundation for more advanced courses, such as STA414 (Statistical Methods for Machine Learning II).

We will cover statistical methods for supervised and unsupervised learning from data: training error, test error and cross-validation; classification, regression, and logistic regression; variable selection; penalized regression; principal components analysis; stochastic gradient descent; decision trees and random forests; k-means clustering and nearest neighbour methods. Computational tutorials will support the efficient application of these methods.

Prerequisites

- **Programming basics:** CSC108H1 / CSC110Y1 / CSC120H1 / CSC148H1 / CSCA08H3 / CSCA48H3 / CSCA20H3 / CSC108H5 / CSC148H5
- **Multivariate calculus:** MAT235Y1 / MAT237Y1 / MAT257Y1 / (MATB41H3, MATB42H3) / (MAT232H5, MAT236H5) / (MAT233H5, MAT236H5)
- **Linear algebra:** MAT223H1 / MAT240H1 / MATA22H3 / MATA23H3 / MAT223H5 / MAT240H5
- **Statistics & probability:** STA238H1 / STA248H1 / STA255H1 / STA261H1 / STAB57H3 / STA260H5 / STA258H5 / ECO227Y1

Instructor

Xin Bing
Office Hours: Monday, 2-4pm, via zoom.
Email: xin.bing@utoronto.ca

Course Materials and Important Links

Quercus Information on attending lectures, tutorials, or office hours; submitting homeworks; or taking tests can be found on Quercus: `https://q.utoronto.ca/courses/278572` (LEC0101) & `https://q.utoronto.ca/courses/278580` (LEC0201).

Piazza We will use Piazza for the course forum. If your question is about the course material and doesn't give away any hints for the homework, please post to Piazza so that the entire class can benefit from the answer: `piazza.com/utoronto.ca/fall2022/sta314h1`.

Textbooks We will mainly use the following textbook for the course.

- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. `https://www.statlearning.com`.

Students are only responsible for the material covered in lectures, tutorials, and homeworks. There are many other publicly available references that you may find useful, such as

- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*.
- Chris Bishop. *Pattern Recognition and Machine Learning*.
- Kevin Murphy. *Machine Learning: a Probabilistic Perspective*.

Delivery Details

Unless otherwise specified, lectures and tutorials will be held in-person. Students should be enrolled in a lecture section and a tutorial section. Since lectures and tutorial are complementary, students are encouraged and expected to attend both lectures and tutorials. There will be two mandatory tests held during the scheduled class time.

Section	Category	Time	Location
LEC0101	Lectures	Wed (11AM-1PM)	MC102
	Office Hours	Mon (2PM-4PM)	Zoom
	Tutorials (0101–0103)	Mon (12PM-1PM)	WB119, WB219, AB107
LEC0201	Lectures	Wed (3PM-5PM)	SS2118
	Office Hours	Mon (2PM-4PM)	Zoom
	Tutorials (0201–0203)	Mon (4PM-5PM)	GB248, WB119,WB219

Course Evaluation

Students are evaluated based on homeworks and tests.

Item	Credit
Homework 1	15%
Homework 2	15%
Homework 3	15%
Homework 4	15%
Midterm Test (held during class)	20%
Final Test (held during class)	20%

See below for due dates.

Homeworks

There will be 4 assignments in this course. The assignments will be released on the course webpage and are due at 11:59PM on the day that they are due.

Item	Released	Due
Homework 1	Thursday 22 September	Wednesday 5 Octorber
Homework 2	Thursday 6 October	Wednesday 19 October
Homework 3	Thursday 3 November	Wednesday 16 November
Homework 4	Thursday 17 November	Wednesday 30 November

Format Homeworks must be submitted in PDF format through Quercus. We encourage typesetting using LaTeX, but scans of handwritten solutions are also acceptable as long as they are legible.

Lateness Homeworks will be accepted up to 3 days late, but 10% of the total credit for the assignment will be deducted for each day late, rounded up to the nearest day. No credit will be given for assignments submitted after 3 days. Extensions will be granted only in special situations, and you will need a written request approved by the instructor at most one week after the due date. Your written request must be made via Piazza to the instructors, must include your student ID number, the specific homework number and must include the following declaration:

I affirm that I was experiencing an illness or personal emergency that has prevented me from submitting this homework and I understand to falsely claim so is an offence under the Code of Behaviour on Academic Matters.

Please note that if you are missing more than one week of class due to illness or emergency then please reach out to your registrars office to notify them as soon as possible.

Collaboration policy Students are required to work on the assignments and to submit their handout alone. Discussion with instructors and other students are allowed. If you choose to discuss the assignment with other students, then you:

- Must include a statement in your submission that includes the name of the student that you discussed with and what part of your submission is involved.
- Must not share proofs, pseudocode, code, or simulation results.
- Must do your own work.

Violation of this policy is an academic offence and will be investigated and reported as such.

Regrading policy Regrading requests should be submitted via Quercus. Regrading requests must include student name, student number, and a justification for the request, which refers specifically to the student's answers and the course materials. Requests without this justification will not be considered. Requests will be considered by the same TA who marked the assignment. The deadline for requesting a regrading is one week after the marked assignments are returned. Remarks may result in a decrease in the grade.

Tests

The course will have 2 tests, each with a duration of 2 hours. The midterm test is held during the normal class time while the final test is held in the final period (its date and location are to be annouced). Tests will be closed-book. Students are responsible for the material covered in lectures, tutorials, and homeworks, but focus will be placed on material introduced during lecture. More details will be provided during the term.

Students must take the test with their assigned section, unless they have prior permission from the instructor. Please note, the lecture on both days will be rescheduled within the week.

	LEC0101		LEC0201	
Midterm Test	Wednesday 26 October	11AM-1PM	Wednesday 26 October	3PM-5PM
Final Test	TBA		TBA	

Missed tests Missed tests will get a score of 0 except in the following two cases.

- *Valid medical reason.* In case of illness, you should complete the absence declaration form on ACORN and notify the instructors to request special consideration.
- *Prior approval from the instructor.* The request must be made at least one week in advance of the test date.

For tests that are missed with approval, the policy will be the following.

- *Midterm Test.* If you miss the midterm test with approval, then the weight will be distributed across the other assessments: the homeworks will be worth 18% each and the final exam will be worth 28%.
- *Final Test.* If you miss the final test with approval, a make-up test will be scheduled after classes end and the format of this make-up final will be at the discretion of the instructor. For example, it may be an oral assessment.

Collaboration policy Collaboration on the tests is *strictly* not allowed, and you *must not* discuss the test with anyone other than the instructor or TAs. Each student is responsible for his/her own work. Violation of this policy is an academic offence and will be investigated and reported as such.

Academic Integrity

The University supports acting in honesty, trust, fairness, respect, responsibility, and courage in all academic matters. Students are responsible for knowing the content of the Universitys Code of Behaviour on Academic Matters. All suspected cases of academic dishonesty will be investigated following procedures outlined in the Code of Behaviour above. If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, you are expected to seek out additional information on academic integrity from your instructor or from other institutional resources (`http://academicintegrity.utoronto.ca/`).

CIA's University Accreditation Program and Pathway to Actuarial Credential

This course is one of the mandatory courses under Canadian Institute of Actuaries (CIA)'s University Accreditation Program (UAP). UAP has moved away from the course-by-course accreditation method and towards program accreditation method (the "Pathway 1 of CIA qualification"). Under the new pathway, in order to obtain ACIA (Associate of CIA) professional credential, students need to:

1. Complete a degree from an actuarial program (ACT Specialist or Major) at University of Toronto and pass a list of mandatory courses. No minimum course grade or GPA is required as long as students pass all the mandatory courses. The full list of UofT's 16 mandatory courses are: ACT240, ACT245, ACT247, ACT348, ACT349, ACT370, ACT451, ACT452, ACT466, STA257, STA261, STA302, STA314, ECO101, ECO102, MGT201/RSM219.

   For transition: CIA will accept an actuarial degree from UofT completed between June 30, 2015 and October 31, 2023 without all the specified mandatory courses.

2. Complete the ACIA Module (administered by CIA, projected Spring 2023).

   For transition: a student can be exempt from the ACIA Module if they complete SOA exam PA and the 8 FAP Modules and assessments by December 31, 2023.

3. Complete an open-book ACIA Capstone Exam (administered by CIA, projected Fall 2023).

   For transition: a student can be exempt from the capstone exam by completing any combination of UAP credits or exams for P, FM, IFM, LTAM, STAM and SRM by October 31, 2023. The deadline to apply for UAP credits is September 30, 2023.

Details on the new pathway for students can be found here: `https://education.cia-ica.ca/acia/acia-for-accredited-university-students/`.

Your participation in the UAP program is, as always, voluntary. For further questions, please email education@cia-ica.ca.