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 past 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 popular statistical methods for supervised and unsupervised learning from data as well as important concepts used in the field, including: 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; neural nets; k-means clustering and nearest neighbour methods. Computational tutorials will support effective application of these methods.

Prerequisites

- **Statistics & probability:** STA302H1/ STA302H5/ STAC67H3
- **Multivariate calculus:** MAT235Y1 / MAT237Y1 / MAT257Y1 / (MATB41H3, MATB42H3) / (MAT232H5, MAT236H5) / (MAT233H5, MAT236H5)
- **Linear algebra:** MAT223H1 / MAT240H1 / MATA22H3 / MATA23H3 / MAT223H5 / MAT240H5
- **Programming basics:** CSC108H1 / CSC110Y1 / CSC120H1 / CSC148H1 / CSCA08H3 / CSCA48H3 / CSCA20H3 / CSC108H5 / CSC148H5

Instructor

Name: Xin Bing
Office: UY 9192
Email: xin.bing@utoronto.ca

Course Materials and Important Links

Course email Please, *do not* email the instructor or TAs on their personal or professional emails, unless for absolute emergency. Instead, use the course email, sta314@course.utoronto.ca, for special requests, such as: homework extension, regrading request, absence due to illness, etc. Questions about course material will not be addressed over email and these questions should be instead directed to the course Piazza site.

Course Website All the course materials (schedule, lecture and tutorial slides, readings, practical problem sets) can be found on the course website `http://courses.utstat.utoronto.ca/sta314_w26/`.

Quercus & Piazza Quercus will only be used to make annoucements. We will use Piazza for the course forum to which you need to sign up via `https://piazza.com/utoronto.ca/winter2026/sta314`. If your question is about the course material, logistics and clarification on homework & tutorial problems, please post to Piazza so that the entire class can benefit from the answer. All questions that give hint on *solving* homeworks should be exclusively asked during office hours.

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. There will be *no* synchronous online video stream or recordings of the lectures. Students should be enrolled in a lecture section and a tutorial section. The tutorial sessions are complementary to the lectures, and provide reviews and extension of the important concepts / methods in the lectures as well as helpful demonstrations on how to use computational software to conduct statistical analysis. Students are highly encouraged and expected to attend both lectures and tutorials. The first lecture will be on Monday, Jan 5th (no tutorials in the first week of the class).

Section	Category	Time	Location
LEC0101	Lectures	Mon (3-4pm)	PB B150
		Wed (3-5pm)	PB B250
	Instructor's Office Hours	Wed (9-10am)	Zoom
	Tutorials (101–104)	Mon (4-5pm)	See ACORN

The scheduled office hours and tutorial sessions of the TA's are listed below. Students are highly encouraged to choose the TA's office hours corresponding to their registered tutorial sessions. Unless necessary, do not attend sessions or use office hours for the other section.

Section	TA	Office Hour	Location
LEC0101	Morris Greenberg	TBD	TBD
	Xiaochuan Shi	TBD	TBD
	Haochen Song	TBD	TBD
	Groff William	TBD	TBD

Course Grading Scheme

Students are evaluated based on quizzes, tests and course project.

Item	Credit
Quizzes (taken during tutorials)	5%
Midterm (held during class)	35%
Final (held during the final period)	35%
Course Project (throughout the semester)	25%

There will be approximately 8 quizzes in total. The quiz grade will be calculated as the average after dropping the two lowest scores. Students who frequently answer questions on Piazza or actively participate in discussions during lectures will be given an extra 2% credits.

Tests

The course will have 2 mandatory 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 assessment period (see the dates and locations below). All tests will be closed-book. Students are responsible for the material covered in lectures, tutorials, and practical sets. More details on the tests will be provided during the term.

	Date + time	Location
Midterm	Feb 11th, 3-5pm	EM 119; PB B250; VC 206; VC 215
Final	TBA	TBA

Missed tests If you miss the midterm, its grading weight will be added up to the final, meaning that your final will be worth 70%.

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.

Regrading policy Regrading requests should be submitted to the course email sta314@course.utoronto.ca. Regrading requests must include student name, student number, and a justification for the request, which refers specifically to the problem and the student's answers. Requests without this justification will not be considered. Requests will be considered by the same TA who marked the problem. The deadline for requesting a regrading is one week after the marks are returned. Regrading requests may result in a decrease in the grade.

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/`).

Course Project

A course project will be initiated within the first two weeks and span the whole fall semester. The goal of the course project is to provide real world applications in which students shall use statistical methods, not only limited to the ones learned in this course, but those beyond the course material, to solve practical problems.

Students are encouraged to form groups of size 2 to 4 to finish the course project. Your final score will not depend on group size. (Please note that solo projects are NOT allowed.) The final delivery of the project should be one written report from each group. More details will be announced separately.

Every student needs to have a submitted report of the course project. If a final report is not submitted by the deadline (no late submission is allowed), then a 0 score would be assigned to the course project.

Practical Problem Sets

There will be (tentatively) 4 sets of practical problems in this course which will be released on the course webpage during this semester. The problem sets do not count into your final grades but will be very helpful to strengthen and deepen your understanding of the content in both lectures and tutorials. They could also

be related to the exam questions. Students are highly encouraged to carefully go over the problem sets and should be able to solve all questions. Solution to each practical set will be posted at the same time.