

# 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; 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@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.

**Piazza** We will use Piazza for the course forum to which you need to sign up via <https://piazza.com/utoronto.ca/fall12023/sta314h1>. If your question is about the course material, logistics and clarification on homework 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.

**Course Website** Most of the course materials (schedule, lecture and tutorial slides, readings, homeworks) can be found on the course website [http://courses.utstat.utoronto.ca/sta314\\_f23/](http://courses.utstat.utoronto.ca/sta314_f23/).

**Crowdmark & Quercus** We use crowdmark for releasing, submitting and grading homeworks. Quercus will only be used to make announcements.

**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 of the important concepts in the lectures as well as helpful demonstrations on how to use computational software to conduct statistical analysis. Students are encouraged and expected to attend both lectures and tutorials.

Section	Category	Time	Location
LEC0101	Lectures	Wed (11AM-1PM)	MP103
	Instructor's Office Hours	Mon (2PM-3PM)	Zoom link
	Tutorials (101-103)	Mon (12PM-1PM)	SS1083, SS2105, UC52
LEC0201	Lectures	Wed (3PM-5PM)	SS2118
	Instructor's Office Hours	Mon (3PM-4PM)	Zoom link
	Tutorials (201-203)	Mon (4PM-5PM)	SS1072, SS1074, SS1087

The first lecture will be on Monday, September 11th. **However**, unless otherwise announced, there will be no future lectures on Mondays (11AM-12PM, 3PM-4PM) and all lectures will be delivered on Wednesdays as listed above. The slots on Mondays are reserved for make-up classes due to exams and holidays.

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	Tutorial Session
LEC0101	Liam Welsh	TBA	TBA	TBA
	Junhao Zhu	TBA	TBA	TBA
	Weizheng Zhang	TBA	TBA	TBA
LEC0201	Ziyi Liu	TBA	TBA	TBA
	Haochen Song	TBA	TBA	TBA
	Tong Li	TBA	TBA	TBA

## Course Evaluation

Students are evaluated based on homeworks and tests (see below for due dates).

Item	Credit
Homework 1	10%
Homework 2	10%
Homework 3	10%
Homework 4	10%
Midterm Test (held during class)	30%
Final Test (held during the final period)	30%

Students who frequently answer questions on Piazza or actively participate in discussions during lectures will be given an extra 2-3% credit.

## 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	Release	Due
Homework 1	Thursday, Sep 21st	Wednesday, Oct 4th
Homework 2	Thursday, Oct 5th	Wednesday, Oct 18th
Homework 3	Thursday, Nov 2nd	Wednesday, Nov 15th
Homework 4	Thursday, Nov 16th	Wednesday, Nov 29th

**Format** Homeworks must be submitted in PDF format through crowdmark. We encourage typesetting using  $\text{\LaTeX}$ , 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 least three days before the due date. Your written request must be made via the course email [sta314@utoronto.ca](mailto:sta314@utoronto.ca) to the instructors, must include your student ID number, the specific homework number and must follow the rule in the paragraph of **Absence declaration** below.

**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 to the course email [sta314@utoronto.ca](mailto:sta314@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 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 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 (its date and location are to be announced). Both 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 the test day will be rescheduled within the week.

	LEC0101		LEC0201	
Midterm Test	Wednesday, Oct 25th	11AM-1PM	Wednesday, Oct 25th	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.* You must follow the rule in the paragraph of **Absence declaration** below.
- *Prior approval from the instructor.* The request must be made at least one week before the test date.

For the midterm test that is missed in either of the above two cases, the policy will be the following.

- The weight of the midterm test will be added up to the final exam, meaning that the final exam will be worth 60%.

**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.

## Absence declaration

In case of illness, you need to provide at least one of the following documents

- Absence declaration via ACORN,
- U of T Verification of Illness or Injury Form (VOI), or a different form (like a letter from a doctor), as long as it is an original document, and it contains the same information as the VOI.
- College Registrars letter,
- Letter of Academic Accommodation from Accessibility Services.

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.

## 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 University's 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/>).