

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; 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.

Instructor

Chris Maddison

Office Hours: Monday 3-4PM and Friday 12-1PM via Zoom

Email: sta314@utoronto.ca

Course Materials and Important Links

Course email Please, *do not* email the instructor or TAs on their personal or professional emails. Instead, use the course email: sta314@utoronto.ca. Questions about course material will not be addressed over email and these questions should be instead directed to the Course Piazza site.

Course Website Most of the course materials (schedule, lecture and tutorial slides, readings, homeworks) can be found on the course web site: https://www.cs.toronto.edu/~cmaddis/courses/sta314_f21/.

Quercus Information on attending lectures, tutorials, or office hours; submitting homeworks; or taking tests can be found on Quercus: <https://q.utoronto.ca/courses/236213>.

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: <https://piazza.com/utoronto.ca/fall2021/sta314>.

Delivery Details

Lectures and Tutorials Unless otherwise specified, lectures and tutorials will be held synchronously, either online via Zoom or in-person. Whether we are in-person or online will change throughout the term, and the course website will have the most up to date information. Students should be enrolled in a lecture section and a tutorial section, and they are expected to attend both lectures and tutorials. There will be two mandatory tests held during the scheduled class time.

Friday Classes	LEC0201	10AM-12PM
	Office Hours	12PM-1PM
	TUT0201, TUT0202, TUT0203, TUT0204	1PM-2PM
Monday Classes	LEC0101	1PM-3PM
	Office Hours	3PM-4PM
	TUT0101, TUT0102, TUT0103, TUT0104	4PM-5PM

Recordings All course videos and materials belong to your instructor, the University, and/or another source depending on the specific facts of each situation, and are protected by copyright. In this course you are permitted to download session videos and materials for your own academic use, but you should not copy, share, or use them for any other purpose without the explicit permission of the instructor. For questions about recording and use of videos in which you appear please contact your instructor.

Course References

There is no textbook for the course. Students are only responsible for the material covered in lectures, tutorials, and homeworks. Nevertheless, there are many publicly available references that you may find useful.

- Chris Bishop. *Pattern Recognition and Machine Learning*.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*.
- David MacKay. *Information Theory, Inference, and Learning Algorithms*.
- David Barber. *Bayesian Reasoning and Machine Learning*.
- Kevin Murphy. *Machine Learning: a Probabilistic Perspective*.
- Shai Shalev-Shwartz & Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*.

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.

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

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	Friday 17 September	Thursday 30 September
Homework 2	Friday 1 October	Thursday 14 October
Homework 3	Friday 29 October	Thursday 11 November
Homework 4	Friday 12 November	Thursday 25 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 email (sta314@utoronto.ca), 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 registrar's office to notify them as soon as possible.

Collaboration policy Students may work on the assignments alone or in pairs. Each student or pair of students is responsible for their own work. The marking scheme will not be adapted depending on the size of the team. If you choose to work in a pair, then you:

- Must register the pair with the instructor before the due date.
- Must include a contributions statement in your submission that describes the contribution of each team member.

If you choose to discuss the assignment with students that are not your teammate, then you:

- Must not share proofs, pseudocode, code, or simulation results.
- Must do your own work and contribute equally to the assignment.

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

Remarks Remark requests should be submitted to the course email (sta314@utoronto.ca). Remark 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 remark 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 1 hour and held during the normal class time. Tests will be closed-book and held during class hours. 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 schedule on both days will be somewhat unusual.

	LEC0201		LEC0101	
Midterm Test	Friday 22 October	10AM-11AM	Monday 25 October	1PM-2PM
Final Test	Friday 3 December	10AM-11AM	Monday 6 December	1PM-2PM

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 *may 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 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/>).

Canadian Institute of Actuaries (CIA)'s University Accreditation Program (UAP)

STA314 is an accredited course under the UAP program. The minimum grade needed to apply for an exemption is 80. For detailed information on UAP, please visit the following webpages:

- University Accreditation Program description
- List of accredited courses offered by University of Toronto
- How to apply for CIA exemptions

Note: The CIA will grant credits to students for SOA/CAS examinations based on the achievement of the minimum Grade towards Associateship (ACIA) and Fellowship (FCIA) in the CIA. At the time of this agreement, CIA credits are recognized by the following actuarial organizations towards their respective designations:

- Casualty Actuarial Society (CAS): ACAS, FCAS
- UK Institute and Faculty of Actuaries (IFoA): FIA, AIA
- Institute of Actuaries of Australia (IAA): AIAA, FIAA
- Actuarial Society of South Africa (ASSA): AMASSA, FASSA
- American Academy of Actuaries (AAA): MAAA

The CIA does not guarantee that credits granted to students under the CIA UAP will be recognized by any other actuarial organizations towards their actuarial designations.