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 10-11AM in AH400 and Tuesday 7-8PM in AH100
Email the professor: sta314-f25-prof@cs.toronto.edu
Email the entire course staff: sta314-f25-tas@cs.toronto.edu

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 emails (above). 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 date information, location information, and 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_f25/`.

Quercus We will make announcements and post non-public materials, e.g. homework and test solutions, on Quercus: `https://q.utoronto.ca/courses/404585`.

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/fall2025/sta314`.

Crowdmark We will use Crowdmark for assignment submission on Quercus.

Delivery Details

Lectures, Tutorials, and Office Hours Unless otherwise specified, lectures, tutorials, office hours, midterm and final exam will be delivered in-person. 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 one mandatory midterm test held during the scheduled class time, and a final exam, proctored by the Faculty of Arts & Science during the exam period.

You can attend any of the office hours, but please attend your assigned lecture and tutorial section. Students enrolled in LEC0101 must enroll in one of TUT0101-0104. Students enrolled in LEC5101 must enroll in one of TUT0201-0204. Please see the course website for times and building locations. The specific room information should be available on ACORN.

Recordings Lecture recordings will be generated and posted automatically on the OCCS Student App on Quercus. *Remember that course videos and materials belong to your instructor and the University and are protected by copyright.* You are permitted to download videos and materials for your own academic use, but *you may not copy, share, or otherwise distribute them* without explicit permission from the 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	7.5%
Homework 2	7.5%
Homework 3	7.5%
Homework 4	7.5%
Midterm Test (held during class)	30%
Final Test (proctored by FAS during exam period)	40%

See the course website 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 / MAT224H1 / MAT240H1 / MATA22H3 / MATA23H3 / MAT223H5 / MAT240H5 / MATB24H3 / MAT224H5
- **Statistics:** STA302H1 / STA302H5 / STAC67H3

Exclusions

- CSC311H1, CSC311H5, STA314H5, STA315H5, CSCC11H3

Format Homeworks must be submitted in PDF format through Crowdmark. 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, must include your student ID number, and the specific homework number.

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 You are welcome to work together with other students on the homework. You are also welcome to use any resources you find (online tutorials, textbooks, papers, etc.) to help you complete the homework. However, you must write up your submission *by yourself* without the help of someone else or generative AI. You must cite all the collaboration and resources you used to complete each assignment. Violation of this policy is grounds for a mark of 0 for that homework.

Remarks Remark requests must be submitted to the course TA email, sta314-f25-tas@cs.toronto.edu. 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 one midterm test and a final exam. The midterm will be 1 hour in duration and held during lecture hours. The final exam will be held during the final exam period and proctored by FAS. For both the midterm and final exam, you will be allowed to bring *only* one double-sided aid-sheet (8.5" by 11"). 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. See the course website for dates and timing details.

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

- *ACORN Absence Declaration.* You must qualify for an absence declaration and fill out the form on ACORN. Please notify the instructors as well to let them know.
- *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 redistributed to the final exam, which will be worth 70%.
- *Final Exam.* If you miss the final test with approval, a make-up exam 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/`).

Policies on Generative AI

General questions You are welcome to ask generative AI chatbots (GenAI for short) general questions about the course content. For homework assignments, you may ask general questions about concepts related to the homework questions. Be careful, though, as GenAI can make mistakes or false statements. You are responsible for the correctness of your assignments.

Homeworks You may not directly ask GenAI tools hints on homework assignment questions; e.g., you should not be copying and pasting directly from the assignment handout into the chat interface. Also, you may not directly use the outputs of AI chatbots in your homework solutions (even paraphrased) unless instructed to do so.

We will interpret the above policy leniently when judging whether the GenAI use is appropriate, though we reserve the right to update the GenAI policy if we find that students are using it in a way that reduces the educational value of the homeworks.

Ultimately, the assignments are designed to help you learn. That is, their primary purpose is to help you engage with and learn the course materials. If you rely too much on outside resources, you will likely not learn the material and will do poorly on the tests and exams, during which such resources will not be available.

Tests and exams GenAI tools are *strictly* not allowed for tests or exams, and you *may not* discuss the tests or exams with GenAI tools or 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.