CSC 311 Fall 2022: Introduction to Machine Learning

Table of Contents

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 AI as an academic field and industry. This course provides a broad introduction to some of the most commonly used ML algorithms. It also introduces vital algorithmic principles that will serve as a foundation for more advanced courses, such as CSC412/2506 (Probabilistic Learning and Reasoning) and CSC413/2516 (Neural Networks and Deep Learning).

We start with nearest neighbors, the canonical nonparametric model. We then turn to parametric models: linear regression, logistic regression, softmax regression, and neural networks. We then move on to unsupervised learning, focusing in particular on probabilistic models, but also principal components analysis and K-means. Finally, we cover the basics of reinforcement learning.
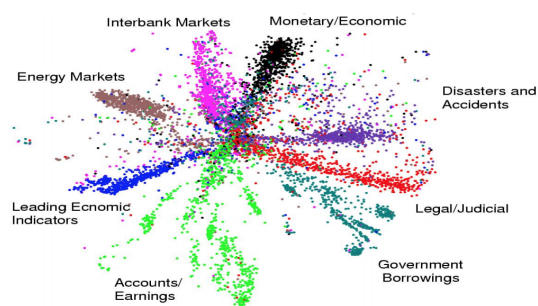
Instructors

Rahul Krishnan and **Alice Gao**

Lecture and Tutorial Times

As of Sep 2022, we plan to have in-person lectures, tutorials, office hours, midterm and final exam in the fall 2022 term. This may, given the COVID-19 situation, be subject to change by the university.

Please attend your assigned lecture section. We strongly encourage students to attend the tutorials although they are optional. Auditing is not allowed this term without express written permission by the instructor.

Section	Lecture	Tutorial
LEC0101, LEC2001	Wednesday 12 - 2pm at MC 254 by Rahul G. Krishnan	Friday 12 - 1pm at MC 254
LEC0201	Tuesday 3 - 5pm at NL 6 by Alice Gao	Thursday 3 - 4pm at AH 400

Prerequisites

- Programming Basics: CSC207/ APS105/ APS106/ ESC180/ CSC180
- Multivariate Calculus: MAT235/ MAT237/ MAT257/ (minimum of 77% in MAT135 and MAT136)/ (minimum of 73% in MAT137)/ (minimum of 67% in MAT157)/ MAT291/ MAT294/ (minimum of 77% in MAT186, MAT187)/ (minimum of 73% in MAT194, MAT195)/ (minimum of 73% in ESC194, ESC195)
- Linear Algebra: MAT221/ MAT223/ MAT240/ MAT185/ MAT188
- Probability: STA237/ STA247/ STA255/ STA257/ STA286/ CHE223/ CME263/ MIE231/ MIE236/ MSE238/ ECE286

Communication

There are many ways to get in touch with us.

- Piazza: piazza.com/utoronto.ca/fall2022/csc311h1f20229
- Course email address: csc311-2022-09@cs.toronto.edu
- Instructor and TA office hours

Please follow these rules when you contact us:

- If your question is **course related and doesn't give away answers**, please post on Piazza publicly so the entire class can benefit from the answer.
- If your question is **course related and may give away answers**, please post on Piazza privately.
- For **remark requests**, please submit on MarkUs (for assignments) or contact us via the course email: csc311-2022-09@cs.toronto.edu.
- For **special considerations requests**, please contact us via the course email: csc311-2022-09@cs.toronto.edu.

Instructor Office Hours

- Rahul Krishnan: Tuesdays 9-10:30am at [PT286](PT286)
- Alice Gao: Wednesdays 4 - 5pm, Fridays 3 - 4pm at [BA4240](BA4240)

TAs will hold office hours to help with assignments and project, as well as preparing for the midterm and the final exam. TA office hours will be posted in the respective sections.

COVID-19

Although the pandemic has diminished somewhat, all indications are that we are not yet out of the woods. The university no longer requires the use of masks on its premises but encourages it where it is impossible to maintain physical distancing, such as in classrooms and offices.

We strongly recommend that you continue to **wear masks** during lectures, tutorials, and office hours out of consideration for the health of others. We also strongly encourage you to **get vaccine booster shots** whenever possible. The instructors plan to wear masks when in close proximity with students, such as when answering questions after lectures or during office hours. However, we may take off our masks when lecturing if we are at a safe distance from students.

Marking Scheme

You must obtain a minimum grade of 40% on the final exam to pass this course.

Component	% Final Grade
3 Assignments	35% (~11.67% each)
Ethics	5%
Project	10%
Midterm	20%
Final	30%

5% of your course grade comes from assignments associated with the ethics module. All of these assignments will be short, and we expect that most of you will receive full marks. >>>>>>> Stashed changes

Assignment	% Final Grade	Marking
Pre-module survey	1%	Full credit for submitting.
Class participation	0.5%	You get this 0.5 point automatically.
Reflections on In-Class Activity	2%	Full credit for a good-faith effort.
Post-module survey	1.5%	Full credit for submitting.

Course Schedule

This course consists of 12 lectures, 3 assignments, a midterm, and a final exam. See the detailed course schedule below.

☰ csc311_fall2022_schedule	3 / 4	—	95%	+	▯ ◇

Assignments

There will be 3 assignments in this course, posted below. Assignments will be due at 11:59pm on Mondays or Fridays and submitted through MarkUs.

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

Late Submission Policy: Assignments will be accepted up to **3 days late**, but **10%** will be deducted for each day late, rounded up to the nearest day. No credit will be given for assignments submitted after 3 days.

Collaboration Policy: Collaboration on assignments is not allowed. Each student is responsible for his/her own work. Discussion of assignments should be limited to clarification of the handout itself, and should not involve any sharing of pseudocode or code or simulation results. Violation of this policy is grounds for a semester grade of F, in accordance with university regulations.

#	Materials	TA Office Hours
Assignment 1	TBD	TBD
Assignment 2	TBD	TBD
Assignment 3	TBD	TBD

Final Project

For your final project, you will attempt to solve a Netflix-Competition-style matrix completion problem. The goal is to predict, in the context of a personalized education platform, whether a student will correctly answer a diagnostic question. In groups of 2-3, you will implement and evaluate several algorithms from the course, and then propose and evaluate an extension to one of these algorithms. This will hopefully be a fun exercise that gives you a feel for what you'd do on a daily basis as a data scientist or machine learning engineer.

We will post the instructions and the start code at a later date.

TA Office Hours: TBD

Midterm

The midterm will occur in person during tutorial time slots. See the detailed course schedule above. As an aid, you can bring one 8.5' x 11' double-sided sheet of paper to the midterm.

TA Office Hours: TBD

Final Exam

The final exam schedule will become available around the end of November 2022. Check the A&S page for the final exam schedule. Please do not make travel plans before the final exam schedule is released.

TA Office Hours: TBD

Academic Integrity

Academic integrity is essential to the pursuit of learning and scholarship in a university, and to ensuring that a degree from the University of Toronto is a strong signal of each student's individual academic achievement. As a result, the University treats cases of cheating and plagiarism very seriously. The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic dishonesty and the processes for addressing academic offences. Potential offences include, but are not limited to:

In papers and assignments:

- Using someone else's ideas or words without appropriate acknowledgement;
- Submitting your own work in more than one course without the permission of the instructor;
- Making up sources or facts;
- Obtaining or providing unauthorized assistance on any assignment.

On tests and exams:

- Using or possessing unauthorized aids;
- Looking at someone else's answers during an exam or test;
- Misrepresenting your identity; and
- When you knew or ought to have known you were doing it.

In academic work:

- Falsifying institutional documents or grades;

- Falsifying or altering any documentation required by the University, including (but not limited to) doctor's notes; and
- When you knew or ought to have known you were doing so.

All suspected cases of academic dishonesty will be investigated following procedures outlined in the Code of Behaviour on Academic Matters. If students have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, they are expected to seek out additional information on academic integrity from their instructors or from other institutional resources.

Remark Request Policy

If you discover a marking error on an assignment or the midterm, you can submit a remark request. We will consider remark requests **up to two weeks** after we release the marks for an assignment or the midterm.

- For an **assignment**, please submit the remark request on MarkUs.
- For the **midterm**, please fill out this remark request form and send it to the course email.

Once the two-week period has passed, we will process all the requests as soon as possible.

Special Considerations Policy

If you are unable to complete an assignment on time or write a test due to extraordinary circumstances beyond your control, please apply for a Special Consideration by filling out this special considerations form and sending it to the course email with your supporting documentation. A special consideration request, particularly if it is not your first request in the course, would not be granted automatically.

Legitimate reasons to apply for a special consideration request:

- Late course enrollment
- Medical conditions (i.e., physical/mental health, hospitalizations, injury, accidents)
- Non-medical conditions (i.e., family/personal emergency)

A heavy course load, multiple assignments/tests scheduling during the same period, and time management issues are not appropriate reasons to grant special considerations. Such accommodations are meant for exceptional circumstances only and not as a means to catch up on term work. If you are having difficulty with stress and time management, please contact your college registrars, who can in turn suggest wellness counselling, academic advising, and/or learning strategists services.

Our special considerations policies are as follows.

- If you **miss the midterm**, we will shift the weight of the midterm to the final exam.
- If you **miss an assignment deadline**, we will shift the weight of the assignment to future assignments or to the final exam.
- If you are **registered with accessibility services**, your letter of accommodation will allow for an extension of up to 7 full days. However, due to the incremental nature of CS courses, granting such a long extension from the onset may cause you to fall behind and be at a disadvantage. As such, we will start by suggesting an initial 3-day extension. We will grant the 7-day extension later if necessary.

Student Support Resources

- CSC Wellness Resources
- Student Mental Health Resources
- Accessibility Services
- Student Life
- Recognized Study Groups

Recommended Textbooks

Suggested readings are optional; they are resources we recommend to help you understand the course material. All of the textbooks listed below are freely available online.

Bishop = Pattern Recognition and Machine Learning, by Chris Bishop.
ESL = The Elements of Statistical Learning, by Hastie, Tibshirani, and Friedman.
MacKay = Information Theory, Inference, and Learning Algorithms, by David MacKay.
Barber = Bayesian Reasoning and Machine Learning, by David Barber.
Sutton and Barto = Reinforcement Learning: An Introduction, by Sutton and Barto.
MML book = Math for ML, by Deisenroth, Faisal, and Ong.

Lecture and Tutorial Materials

We will post lecture and tutorial slides below as the term goes on.

#	Dates	Topic	Materials	Suggested Readings
1	9/13, 9/14	**Lecture**: Introduction, Nearest Neighbours **Tutorial:** Probability Review		ESL: 1, 2.1-2.3, 2.5
2	9/20, 9/21	**Lecture:** Decision Trees, Bias-Variance Decomposition **Tutorial:** Linear Algebra Review		Bishop: 3.2 ESL: 2.9, 9.2 Course notes: Generalization
3	9/27, 9/28	**Lecture:** Linear Models I **Tutorial:** Bias-Variance Decomposition		Bishop: 3.1 ESL: 3.1 - 3.2 Course notes: Linear Regression, Calculus
4	10/4, 10/5	**Lecture:** Linear Models II **Tutorial:** Optimization		Bishop: 4.1, 4.3 ESL: 4.1-4.2, 4.4, 11 Course notes: Linear Classifiers, Training a Classifier
5	10/11, 10/12	**Lecture:** Linear Models III, Neural Nets I **Tutorial:** PyTorch		
6	10/18, 10/19	**Lecture:** Neural Networks II **Tutorial:** Midterm Review		Bishop: 5.1-5.3 Course notes: Multilayer Perceptrons, Backpropagation
7	10/24, 10/25	**Lecture:** Probabilistic Models **Tutorial: midterm test**		ESL: 2.6.3, 6.6.3, 4.3.0 MacKay: 21, 23, 24 Course notes: Probabilistic Models
8	11/1, 11/2	**Lecture:** Multivariate Gaussians, GDA **Tutorial:** Linear Algebra Review II: Eigenvalues, SVD		Bishop: 12.1
9	11/15, 11/16	**Lecture:** Principal Component Analysis, Matrix Completion **Tutorial:** Final Project Overview		ESL: 14.5.1
10	11/22, 11/23	**Lecture:** Embedded Ethics Unit on Recommender Systems **Tutorial:** no tutorial this week		Recommended: Beyond Engagement
11	11/29, 11/30	**Lecture:** k-Means, EM Algorithm **Tutorial:** EM Algorithm		MacKay: 20 Bishop: 9 Barber: 20.1-20.3 Course notes: Mixture Modeling
12	12/6, 12/7	**Lecture:** Reinforcement learning **Tutorial:** Final Exam Review		Sutton and Barto: 3, 4.1, 4.4, 6.1-6.5

Computing Resources

For the assignments, we will use Python 3, and libraries such as NumPy, SciPy, and scikit-learn. You have two options:

1. The easiest option is probably to install everything yourself on your own machine.

    1. If you don't already have python 3, install it.

        We recommend some version of Anaconda (Miniconda, a nice lightweight conda, is probably your best bet). You can also install python directly if you know how.

    2. Optionally, create a virtual environment for this class and step into it. If you have a conda distribution run the following commands:

```
conda create --name csc311
source activate csc311
```

3. Use `pip` to install the required packages

```
pip install scipy numpy autograd matplotlib jupyter sklearn
```

2. All the required packages are already installed on the Teaching Labs machines.