

CSC311 - Introduction to Machine Learning (Fall 2023)

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 neighbours, the canonical non-parametric model. We then turn to parametric models: linear regression, logistic regression, soft max 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.

This course will delve deep into the mathematical foundations of machine learning and will focus on implementing, understanding and debugging software programs for building machine learning models from data. This class will make heavy use of the mathematical and statistical pre-requisites for this class. Students are expected to be comfortable with technical material from multivariate probability theory, multivariable calculus and linear algebra enough to manipulate them towards the description, design and implementation of learning algorithms for machine learning models.

Announcements

- Welcome to CSC311! We will request all students to complete a math diagnostic prior to the class. .
- The first day of classes will be [Tue September 12, Wed September 13 2023].
- [Math diagnostic](#) released. Due on Friday, September 15 2023

Jump to:

[Assignments](#)

[Lectures](#)

Course Sections

There are two sections of this course being offered this term. As of September 2023, we plan to have in-person lectures, tutorials, office hours, midterm and final exam.

Please attend your assigned lecture and tutorial section. Auditing is not allowed this term without express written permission by the instructor.

Section	Instructor	Lecture	Tutorial
0101	Rahul G. Krishnan	Wed 13:00 - 15:00, AH400	Fri 13:00 - 14:00, AH400
0201	Rahul G. Krishnan	Tue 15:00 - 17:00, MS2158	Thu 15:00 - 16:00, SF1105

A Note on COVID-19

The pandemic has diminished and the university no longer requires the use of masks on its premises. However, we encourage the use of masks 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 encourage you to **get vaccine booster shots** whenever possible. The instructors may 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.

ChatGPT-GPT4 and the use of generative AI tools for learning

There are several generative AI tools such as ChatGPT, GPT4, Bard, Claude available for public use as general purpose chatbots. These machine learning models are incredibly powerful and are capable of providing responses to arbitrarily complex text (and sometimes image) queries. You may use generative artificial intelligence (AI) tools, including ChatGPT and GitHub Copilot, as learning aids.

Generative AI is not required to complete any aspect of this course, and we caution you to not rely entirely on these tools to complete your coursework. Instead, we recommend treating generative AI as a supplementary tool only for exploration or drafting content. Ultimately, you (and not any AI tool) are responsible for your own learning in this course, and for all the work you submit for credit. It is your responsibility to critically evaluate the content generated, and to regularly assess your own learning independent of generative AI tools. Overreliance on generative AI may give you a false sense of how much you've actually learned, which will lead to poor performance on the midterm test and final exam, in later courses, or in future work or studies after graduation. You will not be permitted to use generative AI on the midterm test or final exam. Remember that you (and not any AI tool) are responsible for your own learning in this course, and for the final work you submit for this assignment. This assignments and homeworks are designed to be doable without the use of generative AI, using only concepts and skills we have covered in lecture and in the course readings.

While some generative AI tools are currently available for free in Canada, please be warned that these tools have not been vetted by the University of Toronto and might not meet University guidelines or

requirements for privacy, intellectual property, security, accessibility, and records retention. Generative AI may produce content which is incorrect or misleading, or inconsistent with the expectations of this course. They may even provide citations to sources that don't exist—and submitting work with false citations is an academic offense. These tools may be subject to service interruptions, software modifications, and pricing changes during the semester.

The final work you submit for assignments in this course must be your own, and may not include any verbatim content from generative artificial intelligence (AI) tools. You may, however, use generative AI to support your work on this assignment in the following ways:

- To answer general questions about high-level concepts covered in this course or assignment
- To provide examples on the usage of a library's API
- To summarize information
- To generate test cases for your code

- To assist with understanding and debugging errors

If you choose to use any generative AI tools while working on this assignment, you must acknowledge which generative AI tools you used and how you used them. It is an academic offence to not credit sources—including generative AI—in work that you submit. This acknowledgment should take the form of:

- A statement at the beginning of the solution pdf highlighting which, tool was used and how.
- You should submit a separate txt file listing all raw ChatGPT transcripts that record your interactions with the tool while working on each assignment.
- For code, you should use comments in your code acknowledging which functions, classes, or blocks of code were created in collaboration with generative AI. e.g. `#genai - dataloader was created with chatgpt`

Any uses of generative AI beyond the ones listed above are not allowed, and will be considered use of an unauthorized aid, which is a form of cheating. In summary:

- This course is designed with a view to not have to use any such tools to fully absorb the course material and content.
- The use of generative AI tools can be both useful and simultaneously harmful to learning the fundamentals of machine learning. Your knowledge of the concepts will be tested on the final exam and midterm exam where you will not have access to generative AI tools.
- Any use of GPT4/ChatGPT must be documented i.e. students should explicitly state which model they used, and list the queries they found helpful. Each submitted assignment will require submission of a list of queries used with generative AI tools.
- Students may use GPT4/ChatGPT to help understand and personalize concepts taught in homework and lectures on their own time. Warning: Please note that the model can hallucinate and can fail in unpredictable ways and students should expect this.
- Using GPT4/ChatGPT's output directly in any material handed in for homework constitutes an academic violation. Students are expected to write their own homework assignments even if such tools

were used as aids to learn concepts.

Prerequisites

This course has the following pre-requisites:

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

Note that these are the **minimum** pre-requisites; in practice you are expected to have an operational knowledge of the topics in the pre-requisites. CSC311 is a **fast paced course**. The preliminary quiz will help you self assess the degree to which your knowledge of the pre-requisites will be tested during the course of the semester. If you find yourself falling behind on the mathematical concepts, we strongly recommend investing time into self-study using the Mathematics for Machine Learning textbook.

Grading Scheme

We will use the following grading scheme for the course.

Evaluation	Weight
Math Diagnostic Quiz	3%
Assignments (3)	27%
Project	10%
Midterm Exam	20%
Ethics Module	5%
Final Exam	35%

Note that you must obtain a grade of **at least 40%** on the final exam to pass the course. For the Ethics Module, the 5% weight will be split into mini-assignments, as follows:

Mini-Evaluation	Weight	Marking
Pre-Module Survey	1%	Completion
Class Participation	0.5%	Automatically Given
Reflection on Class Activity	2%	Good-Faith Effort
Post-Module Survey	1.5%	Completion
In short, we expect almost all students to get the full 5%.		

Communication

There are many ways to get in touch with us.

- **Piazza (for course material related questions) :**
<https://piazza.com/utoronto.ca/fall2023/csc3112023fall> Fastest way to get answers to course related questions.
- **Course Email Address (for administrative questions) :** csc311-2023-09@cs.toronto.edu
- **Instructor Office Hours:** We will prioritize conceptual questions about the lecture material during instructor office hours.
 - PT286 Tuesday 9:30-11am
 - PT286 Wednesday 9:30-11am
- **TA Office Hours:** These will be held and are designed to answer questions about each assignment, the midterm exam, and final exam. Please monitor the respective sections for dates/times.

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 Crowdmark or Markus (for assignments).
- For **special considerations requests**, please contact us via the course email. There is a ticketing system which instructional support helps us process.

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.

Assignments

There will be 3 assignments in this course, posted below. Assignments will be due at 5pm on Tuesdays or Fridays and submitted through Crowdmark.

#	Materials	TA Office Hours
Assignment 1		Day Date Time Location TBD
Assignment 2		Day Date Time Location TBD
Assignment 3		Day Date Time Location TBD

[Math diagnostic](#). The Math for ML textbook is probably helpful to skim.

Anticipated due dates:

Assignment 1: TBD

Assignment 2: TBD

Assignment 3: TBD

Computational Resources: 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 scikit-learn
```

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

Late Submission Policy: Everyone will receive 3 grace days, which can be used at any point during the semester on the three assignments. No credit will be given for assignments submitted after 3 days. We will have a separate policy on the final project.

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.

Remark Policy: If you discover a marking error on an assignment, 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. Please submit your remark request via Crowdmark.

Midterm Examination

About half-way through the term, we will have a midterm examination. You will be allowed to bring **one** aid-sheet (one-sided 8.5" by 11"). The midterm will be held during the lecture slot on Feb 13. More details will be posted as we get closer to the midterm.

Final Examination

At the end of the term, we will have a FAS proctored final examination. You will be allowed to bring an aid-sheet (two-sided 8.5" by 11"). More details will be posted as we get closer to the final exam.

Course Project

For your final project, one option is to solve a 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. The other option is open-ended and both are described in the handout below.

Office hours

Day Date Time Location TBD

Lecture and Tutorial Materials

You can find all of the relevant lecture/tutorial materials below. The suggested readings are optional. We have provided them in case you need alternate resources to understand the material. All of the textbooks listed below are freely available online.

Lecture recordings are available on Piazza.

- Bishop: [Pattern Recognition and Machine Learning](#).
- Hastie, Tibshirani, and Friedman: [The Elements of Statistical Learning \(ESL\)](#).
- MacKay: [Information Theory, Inference, and Learning Algorithms](#).
- Barber: [Bayesian Reasoning and Machine Learning](#).
- Sutton and Barto: [Reinforcement Learning: An Introduction](#).
- Deisenroth, Faisal, and Ong: [Math for ML](#).
- Shalev-Shwartz and Ben-David: [Understanding Machine Learning: From Theory to Algorithms](#).
- Murphy: [Machine Learning: a Probabilistic Perspective](#).

Week	Topics Covered	Materials	Suggested Readings
1 S12, S13	Lecture: Introduction / Nearest Neighbours Tutorial: Probability Review	Lecture: Slides	Math for ML (skim) ESL: 1, 2.1-2.3, 2.5
2 S19, S20 HW 1 out	Lecture: Decision Trees, Bias-Variance Decomposition Tutorial: Linear Algebra Review		Bishop: 3.2 ESL: 2.9, 9.2
3 S26, S27	Lecture: Linear Models 1 (Linear Regression) Tutorial: Bias-Variance Decomposition		Bishop: 3.1 ESL: 3.1 - 3.2
4 O3, O4 HW 1 due HW 2 out	Lecture: Linear Models 2 (Logistic Regression) Tutorial: Optimization		Bishop: 4.1, 4.3 ESL: 4.1-4.2, 4.4, 11
5 O10, O11	Lecture: Linear Models 3 / Neural Networks 1 Tutorial: Pytorch		Bishop: 5.1-5.3
6 O17, O18 HW 2 due HW 3 out	Lecture: Neural Networks 2 Tutorial: Midterm Review		Neural Net Playground
7 O24, O25	Lecture: Probabilistic Models Tutorial: Midterm		ESL: 2.6.3, 6.6.3, 4.3.0 MacKay: 21, 23, 24
8 O31, N1 HW 3 due, Project out	Lecture: Multivariate Gaussians, GDA Tutorial: Linear Algebra Review II: Eigenvalues, SVD		Math for ML Bishop: 12.1
9 N14, N15	Lecture: Principal Component Analysis, Matrix Completion Tutorial: Final Project Overview		ESL: 14.5.1
10 N21, N22	Lecture: Embedded Ethics Unit Tutorial: No Tutorial		Beyond engagement Understanding Recommendation Algorithms
11 N28, N29 Project due	Lecture: k-Means, EM Algorithm		MacKay: 20 Bishop: 9

Week	Tutorial Topics Covered	Materials	Suggested Readings
12 D5, D6	Lecture: Reinforcement learning Tutorial: Final Review		Sutton and Barto: 3, 4.1, 4.4, 6.1-6.5 See slides for RL recs.

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.

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.

Student Support Resources

- [CSC Wellness Resources](#)
- [Student Mental Health Resources](#)
- [Accessibility Services](#)
- [Student Life](#)
- [Recognized Study Groups](#)

Special Considerations Policy

If you are unable to complete a course requirement 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.