# CSC 311 Fall 2021: Introduction to Machine Learning

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 in industry. This course provides a broad introduction to some of the most commonly used ML algorithms. It also serves to introduce key algorithmic principles which 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.
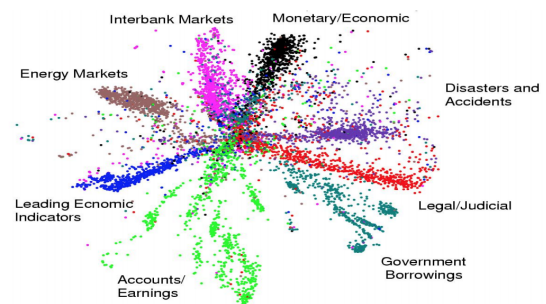
Announcements

- The course information handout (syllabus) is available.

Where and When

Unfortunately, due to the evolving COVID-19 situation, the specific class format is subject to change. As of this writing (9/2), we are required to have an in-person component to this class; we've decided to have in-person lectures. Fortunately, we're not required to force you to show up (with the exception that the final exam may need to be in person - more on that below). We are designing the course so that you do not need to show up in person if you don't want to, and so that not showing up will not put you at a disadvantage. We'll do everything we can to maintain a safe environment while following the letter of University policy.

The first two weeks of class will be entirely virtual, to accommodate international arrivals. After that point, the current plan is for lectures to be held in-person. As this would require in-person gatherings well beyond the size currently allowed by Ontario public health authorities for most purposes, we can make few assumptions about what will be permitted and safe by the time in-person instruction is set to begin. Please continue monitoring Quercus and this course web page for further updates.

Tutorials and office hours will be held virtually throughout the term. Students are encouraged to attend both lecture and tutorial each week, but attendance won't be taken.

We will also accommodate students who are unable to attend in person. Specifically, one of the four lecture sections will be held virtually. Additionally, students will have access to the lecture videos from Fall 2020, which follows roughly the same schedule as this year's class.

Most terms, we allow students to attend sections other than their assigned ones. However, this year, **if you choose to attend in-person, you must attend your assigned section**, since overcrowding during a pandemic would create an unsafe situation. Note that **different sections are held simultaneously in different rooms, so make sure you are in the correct room.** It's likely that many of you will want to switch between virtual and in-person formats. Here are the policies on that:

- If you are assigned to the virtual section and would like to switch to an in-person section, then please let us know by Sept. 16 (through a form which will be communicated through Quercus).
- If you are assigned to an in-person section and would like to attend the virtual section, then please let us know by Sept. 16 through the same form. The reasons are (1) so that we can manage capacity in the in-person sections to accommodate the former group, and (2) so we can make sure to have a Zoom license with sufficient capacity.
- If you are assigned to one in-person section and would like to switch to a different in-person section, let us know by Sept. 16 through the same form. This will require a justification, such as a scheduling conflict with another course.
- Auditing is not allowed this term.

Specifics about online delivery will be sent to enrolled students through Quercus.

Section	Lecture Time	Lecture Room	Instructor	Tutorial Time
LEC0101, LEC2001	Friday 11:00-13:00	KP 108	Roger Grosse	Friday 15:00-16:00
LEC0102	Friday 11:00-13:00	Virtual	Guodong Zhang	Friday 15:00-16:00
LEC0201	Thursday 16:00-18:00	ES 1050	Roger Grosse	Thursday 19:00-20:00
LEC0202	Thursday 16:00-18:00	SS 2102	Rahul Krishnan	Thursday 19:00-20:00

Modulo capacity constraints for the in-person sections, it will be up to you to decide whether the in-person lecture experience is worth the risk. We can't make the decision for you, but do consider that (as of this writing) most in-person gatherings of more than 25 people are banned, and that not only is the lecture a crowded indoor setting, but most of your fellow students will have been attending lots of other in-person lectures. In keeping with the theme of the course, where we use data to make decisions under uncertainty, you may find the MicroCovid Calculator helpful in reasoning about your own risk tolerance.

Course videos and materials belong to your instructor, the University, and/or other 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 the instructors.

Teaching Staff

Communication

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.
- If you have questions that may give away homework answers, please post privately to Piazza.
- For course administration matters (homework extensions, missed exams, etc.) please email the instructors (see below).

All office hours will take place via Gather Town. Details will be communicated via Quercus.

Instructors

	Roger Grosse	Rahul Krishnan	Guodong Zhang
Office Hours (Virtual)	Monday 10am-12	Monday 6-7pm	Monday 8-9pm
Email Instructors	csc311-f21-profs@cs.toronto.edu		

Teaching Assistants

	Homework	Test Prep	Project
Office Hours (virtual)	See homework schedule below.	**Midterm:** Tuesday 10/19 2-4pm Wednesday 10/20 7-9pm **Final:** Tuesday 12/7, 3:30-4:30pm Wednesday 12/8, 3-5pm, 8-9pm Thursday 12/9 9-11am	Thursday 11/18 2-4pm Friday 11/19 1-6pm Wednesday 11/24, 2-6pm Thursday 11/25, 2-4pm Friday 11/26, 8am-noon Tuesday 11/30, 3-5pm Wednesday 12/1, 1-5pm

Marking Scheme

We will use the following marking scheme:

- 3 homework assignments (**35%**, weighted equally)
- minor assignments for embedded ethics unit (**5%**)

  - Good faith effort = full credit

- project (**20%**)

- Due 12/3.

- 2 online tests (**40%**)

  - 1-hour online midterm test. Held during tutorial time on 10/21 or 10/22.
  - 2-hour online final exam, 14:00-16:00 on Friday, 12/10.
  - Weighting: higher of (15% midterm, 25% final) or (10% midterm, 30% final).

Homeworks

Homeworks will generally be due at 11:59pm on Wednesdays, and submitted through MarkUs. Please see the [course information handout](#) for detailed policies (marking, lateness, etc.). The detailed schedule will be posted soon.

#	Out	Due	Materials	TA Office Hours
1	9/16	9/29	[handout] [clean_fake.txt] [clean_real.txt] [clean_script.py]	Thursday, 9/23, 8am-noon Monday, 9/27, 3:30-5:30pm Tuesday, 9/28, 4-6pm and 8:30-10:30pm
2	9/30	10/13	[handout] [starter code]	Thursday, 10/7, 2-4pm Friday, 10/8, 4-6pm Monday, 10/11, 1-3pm Tuesday, 10/12, 9-11am and 4-6pm Wednesday, 10/13, 3-5pm
3	10/14	11/3	[handout] [code and data]	Thursday 10/28, 9:30am-11am Thursday 10/28, 3-4pm Friday 10/29, 12:30pm-2pm Friday 10/29, 2pm-5pm Monday 11/1, 12pm-2pm Tuesday 11/2, 2-4pm Wednesday 11/3, 2-3pm

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

Assignment	Due	% final grade	Marking
Initial Survey	10/29	0.5%	full credit for submitting
Class Participation (Nov. 18/19)	N/A	2%	You get these 2 points automatically.
[Reflections on In-Class Activity](#)	11/27	2%	A good-faith effort receives full credit.
Final Survey	12/15	0.5%	Full credit for submitting.

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. Here is the [project handout](#) and [starter code](#). The final report is due December 3.

Schedule

This is a tentative schedule, which will likely change as the course goes on.

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.

#	Dates	Topic	Materials	Suggested Readings
1	9/9, 9/10	**Lecture**: Introduction, Nearest Neighbours **Tutorial:** Probability Review	**Lecture:** [Slides] **Tutorial:** [Slides]	ESL: 1, 2.1-2.3, 2.5
2	9/16, 9/17	**Lecture:** Decision Trees, Bias-Variance Decomposition **Tutorial:** Linear Algebra Review	**Lecture:** [Slides] **Tutorial:** [Slides] [Worksheet] [Solutions]	Bishop: 3.2 ESL: 2.9, 9.2 Course notes: [Generalization](#)
3	9/23, 9/24	**Lecture:** Linear Models I **Tutorial:** Bias-Variance Decomposition	**Lecture:** [Slides] **Tutorial:** [Worksheet]	Bishop: 3.1 ESL: 3.1 - 3.2 Course notes: [Linear Regression](#), [Calculus](#)
4	9/30, 10/1	**Lecture:** Linear Models II **Tutorial:** Optimization	**Lecture:** [Slides] **Tutorial:** [Slides] [Worksheet] [Solutions]	Bishop: 4.1, 4.3 ESL: 4.1-4.2, 4.4, 11 Course notes: [Linear Classifiers](#), [Training a Classifier](#)
5	10/7, 10/8	**Lecture:** Linear Models III, Neural Nets I **Tutorial:** PyTorch	**Lecture:** [Slides] **Tutorial:** [Colab]	
6	10/14, 10/15	**Lecture:** Neural Networks II **Tutorial:** Midterm Review	**Lecture:** [Slides] **Tutorial:** [Slides]	Bishop: 5.1-5.3 Course notes: [Multilayer Perceptrons](#), [Backpropagation](#)
7	10/21, 10/22	**Lecture:** Probabilistic Models **Tutorial: midterm test**	**Lecture:** [Slides]	ESL: 2.6.3, 6.6.3, 4.3.0 MacKay: 21, 23, 24 Course notes: [Probabilistic Models](#)
8	10/28, 10/29	**Lecture:** Multivariate Gaussians, GDA **Tutorial:** Linear Algebra Review II: Eigenvalues, SVD	**Lecture:** [Slides] **Tutorial:** [Slides]	Bishop: 12.1
9	11/4, 11/5	**Lecture:** Principal Component Analysis, Matrix Completion **Tutorial:** Final Project Overview	**Lecture:** [Slides] **Tutorial:** [Slides] [Colab]	ESL: 14.5.1
10	11/18, 11/19	**Lecture:** Embedded Ethics Unit on Recommender Systems **Tutorial:** no tutorial this week	**Lecture:** [Part 1] [Part 2]	Recommended: [Beyond Engagement](#)
11	11/25, 11/26	**Lecture:** k-Means, EM Algorithm **Tutorial:** EM Algorithm	**Lecture:** [Slides] **Tutorial:** [Slides]	MacKay: 20 Bishop: 9 Barber: 20.1-20.3 Course notes: [Mixture Modeling](#)
12	12/2, 12/3	**Lecture:** Reinforcement learning **Tutorial:** Final Exam Review	**Lecture:** [Slides] **Tutorial:** [Slides]	Sutton and Barto: 3, 4.1, 4.4, 6.1-6.5

Computing Resources

For the homework 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.