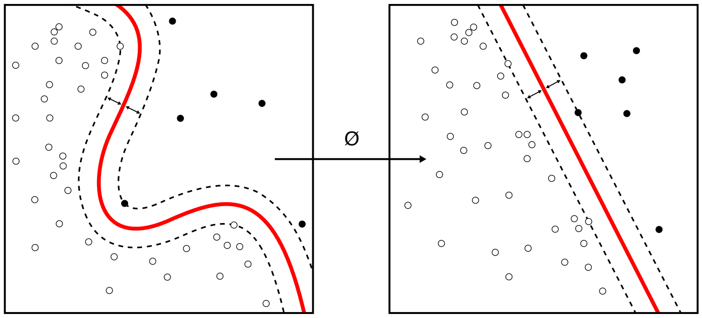
COMP 6321 Machine Learning Fall 2020



This course introduces conceptual and practical aspects of machine learning. Concepts include regression, classification, maximum likelihood estimation, discriminative vs generative modeling, generalization, supervised learning, unsupervised learning, semi-supervised learning. Methods include linear models, mixture models, nearest neighbours, support vector machines, random forests, boosting, ensembles, Gaussian processes, and deep learning.

What you will learn

- The "big ideas" behind many famous learning algorithms, why they work, and why they sometimes don't.
- How to use popular packages like scikit-learn and PyTorch effectively.
- How to train lots models on data, and how to evaluate success and failure.

By the end of the course, students should be conversant in machine learning concepts and terminology, capable of implementing basic learning algorithms from scratch, capable of using the scikit-learn and PyTorch libraries, and be effective at incorporating machine learning into their own research.

Staff

Instructor:
Andrew Delong <andrew.delong@concordia.ca> (but please **use Moodle**)
Office: Zoom links will be announced for office hours
Office hours: Thursdays 9:00–11:00am (by appointment through the **Moodle scheduler** only)

Teaching assistants:
Paras Kapoor <paras.kapoor@concordia.ca>
Soroush Saryazdi <soroush.rahimi@concordia.ca>
Laya Rafiee <laya.rafiee@concordia.ca>

Weekly Schedule

Lecture:
Wednesday 5:45–8:15pm in Zoom

Labs:
`I` : Tuesday 5:45pm–7:35pm in Zoom (Paras Kapoor)
`J` : Tuesday 5:45pm–7:35pm in Zoom (Soroush Saryazdi)
`K` : Friday 5:45pm–7:35pm in Zoom (Laya Rafiee)

Tentative Term Schedule

Materials are being updated as the course progresses. The tentative schedule is below, but may be subject to small changes.

There will be one practice quiz (Q0), four graded quizzes (Q1, Q2, Q3, Q4), two readings (R1, R2), and one graded assignment (A1).

- Quiz Q0 will be a brief practice quiz (ungraded) at the start of the second lecture.
- Quizzes Q1, Q2, Q3, Q4 will be written in the first 20 minutes of the corresponding lecture. The quiz may cover any previously covered material from lectures, labs, or course notes. Don't be late!
- Lab files are due on the Sunday that follows the lab. See "Policies."

The timing of quizzes, readings, and assignments is indicated below.

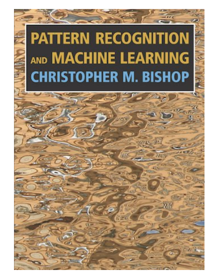
Week	Lecture topics	Lab	QAR
Sep 7	Intro, Linear Regression, Maximum Likelihood	*Lab1* Python, Numpy, Plots	
Sep 14	Logistic Regression, Clustering, K-Means	*Lab2* Linear Models	Q0
Sep 21	Gaussian Mixtures, Kernel Density	*Lab3* Clustering	Q1
Sep 28	Support Vector Machines	*Lab4* Support Vector Machines	R1
Oct 5	Multi-class, Decision Trees, Random Forests	*Lab5* Random Forests	
Oct 12	Bootstrap, Bagging, Boosting, Ensembling	*Lab6* Boosting	Q2
Oct 19	Loss Functions, Generalization, Cross Validation	*Lab7* Hyperparameter Search	A1
Oct 26	Neural Networks, Backpropagation	*Lab8* Neural Networks	
Nov 2	Convolutional Networks, Recurrent Networks	*Lab9* Convolutional Networks	Q3
Nov 9	Dimensionality Reduction, PCA, Autoencoders	*Lab10* Dimensionality Reduction	R2
Nov 16	Generative Models, Naive Bayes, Variational AEs	*Lab11* Transfer Learning	
Nov 23	Generative Adversarial Net, Gaussian Processes	*Lab12* Gaussian Processes	Q4
Nov 30	**Project presentations**	No lab	

Key dates (EST):

- *Sep 16 @ 5:45pm* – Q0
- *Sep 23 @ 5:45pm* – Q1
- *Sep 24 @ 11:59pm* – groups due
- *Sep 30 @ 4:00pm* – R1 due
- *Oct 10 @ 11:59pm* – project proposal due
- *Oct 14 @ 5:45pm* – Q2
- *Oct 23 @ 3:00pm* – A1 due
- *Nov 4 @ 5:45pm* – Q3
- *Nov 11 @ 4:00pm* – R2 due
- *Nov 25 @ 5:45pm* – Q4
- *Dec 1 @ 3:00pm* – project report, code, and presentation due

If absolutely necessary, small changes to these dates may be announced.

Textbook



The required textbook for this course is known as "the PRML book" or "the Bishop book":

- *Pattern Recognition and Machine Learning* by Christopher M. Bishop (2006)

Other recommended books include:

- *Pattern Classification* by Richard O. Duda, Peter E. Hart, and David G. Stork (2001)
- *Applied Machine Learning* by David Forsyth (2019)
- *Information Theory, Inference, and Learning Algorithms* by David J.C. MacKay (2003)
- *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016)

The Bishop book does not adequately explain certain topics. For example, the *Pattern Classification* book covers decision trees, and the *Deep Learning* book explains convolutional and recurrent networks. The *Applied Machine Learning* book covers support vector machines and decision trees and may make a good supplement. However, acquiring the "recommended" books is not required for the course. *The instructor will provide lecture slides, labs, and course notes for topics not covered by Bishop.*

The lectures and course notes may occasionally refer to an online reading via hyperlink. If any such reading is *required*, it will be indicated as such.

Course evaluation

The following assessment is tentative, subject to approval by the department.

10% Labs
16% Assignment (A1)
16% Quizzes (4% each Q1, Q2, Q3, Q4)

8% Readings (4% each R1, R2)
50% Project

There will be no midterm or final exam for the online version of this course.
See "Projects" and "Policies" sections for more details on assessment.

Prerequisites

This course requires good Python programming skills and a good understanding of basic linear algebra, basic multivariable calculus, and basic probability. If you tend to struggle with those topics, this course may not be for you. Prior experience with numerical computing and/or optimization is also very helpful.

Projects

Projects must be done in groups of 2–4 students. The 50% project grade breaks down as follows:

- 5% **1 page proposal** explaining your project goals and plans.
- 25% **4 page report** describing your project data, methods, and conclusions.
- 10% **code and data** to be submitted with the report; the instructor and TAs will inspect these files.
- 5% **originality** to be assessed by the instructor, to reward groups that propose interesting projects.
- 5% **presentation** to be submitted with the report; the instructor will play these in final lecture.

See "Policies" section for more details.

Policies

Academic integrity. Your instructor takes academic integrity very, very seriously. Students who violate the Code of Conduct in will be reported. This includes plagiarism of either code or written text, attempted communication during a quiz or exam, and everything else in the list of offences. Please, if you are struggling with course material, ask for help from TAs or the instructor right away.

Course content. Lectures, lab files, quizzes, and assignments will be hosted on Moodle and posted on a weekly basis.

Zoom profiles. When you join a Zoom session, your Zoom profile must use your real name. The instructor and lab supervisors will use names to identify students, and will remove students who do not appear to be enrolled or who have joined the wrong lab session.

Office hours. Instructor office hours are by appointment only. Each week students will be able to book up to one 15-minute time slot at a time. By default these slots will be Thursdays 9:00am–11:00am EST.

Communication. Communicate with the instructor through **MOODLE ONLY**, except in urgent situations. The instructor is teaching over 200 students, and communicating in Moodle helps him stay organized!

- Questions regarding course material should be posted on the forum on Moodle. Students are encouraged to try to answer each others' questions if the instructor or TAs have not been able to answer immediately.
- Personal matters such as "I will miss lab 4 because I have the flu" should be sent to the instructor as a direct message on Moodle, NOT through e-mail unless it is very urgent.

Lectures. Lectures will be held on Zoom. You will receive a Zoom link via email prior to each lecture. The format of each lecture will be to alternate between a 20-30 min block of course material (recorded) followed by a 5-8 min break (not recorded). During each break, students can ask questions by raising their hand (using Zoom's "raise hand" feature) or simply stretch their legs. Students should be respectful of other students and of the instructor. Recorded portions will be posted on Moodle the following day.

Quizzes. Quizzes will be written as timed Moodle quizzes at the start of lecture. There is one practice quiz (Q0) and four graded quizzes (Q1,Q2,Q3,Q4). Your lowest quiz grade will be increased to match your second-lowest quiz. Therefore, if you miss a quiz, you will not receive zero for the missed quiz. The purpose of the practice quiz is to ensure you do not have confusion or technical problems on the graded quizzes, so, combined with the "lowest quiz gets increased" rule, there will be no further exceptions made for "technical difficulties." Students are encouraged to attempt all quizzes, just in case.

Readings. Over the term, you will be assigned two research papers to read. You will be asked to read each paper and write long-form answers to questions about the paper. Your answers will be assessed by the instructor. The purpose is to strengthen your reading comprehension and technical writing skills regarding important machine learning concepts. Also, the papers will hopefully be interesting to you!

Assignment. The assignment will involve programming in Python and possibly in C. Assignments will be collected via a Moodle "assignment". Assignment instructions will be handed out later, in September.

Labs. There are 12 labs worth 1% each up to a maximum of 10% of the total grade. Labs are a major learning opportunity in this course—please take advantage of them. Scheduled Zoom lab sessions are primarily you to get help from your lab supervisor or from peers. *Students are responsible for knowing the complete lab material.*

- Lab files (`labX.ipynb`) will be collected in Moodle on the Sunday after the lab. The TA for your section will give you full credit for a lab if you tried to answer every question, even if they are not all correct.
- You must attend your assigned lab section ONLY. Please respect this rule, it helps the TAs.
- Students are encouraged to help each other with labs, but copying code is forbidden. You must be the author of all your submitted answers. If you do not complete a lab, you risk struggling on certain quizzes.
- A student can miss at most two labs (two "free passes") without losing marks and without giving a reason. Each additional missed lab requires a doctor's note, or the corresponding lab mark (1% of final mark) will be zero. Again, *students are still responsible for knowing the material from any missed lab.*

Projects. The projects have specific requirements around group assignment, equal contribution, and code reproducibility.

- Students who form their own group must send the names to the instructor via Moodle direct message by the deadline. Otherwise they will be randomly assigned a group.

- Project proposals must be written using the $\LaTeX$ template provided on Moodle. Proposals will be collected as a Moodle "assignment."
- Students are expected to contribute equitably. If an individual group member fails to contribute meaningfully, that person's individual project grade may be lowered.
- The "originality" grade is subjectively assessed by the instructor, with the goal of having students put time into designing their own project and not taking one "off the shelf". For example, if you simply attack a standard Kaggle competition, for which many project codes exist on the web, then you will receive an originality score of 0% even if you did a great job at implementing and writing the report.
- Presentations must be 4 minute recorded videos, like if you were submitted a "highlight video" your project to a virtual conference. Consider using Zoom local recording to record audio over slides.

Marking corrections. If a students believes that a mistake was made in marking, a *marking correction request form* (available on the course Moodle page) must be printed, filled, and sent to the instructor within 72 hours of the quiz or assignment grade becoming available, A scan or a photo of the filled form is acceptable.