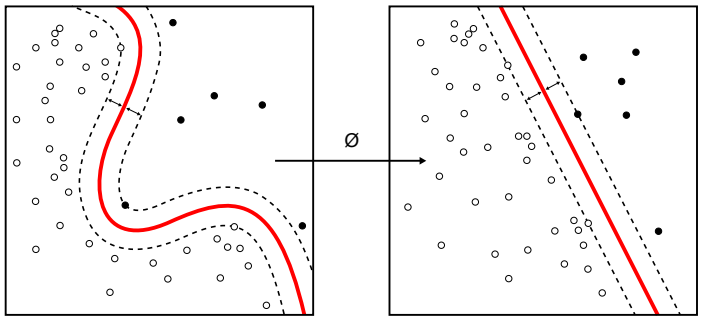
COMP6321 Machine Learning Fall 2019



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

What you will learn

- The core ideas behind standard learning algorithms, why they work, and why they sometimes don't.
- How to train models on a data set, and how to evaluate performance.
- Recent advances and where the field is heading.

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>
Office: EV 3.129
Office hours: Thursdays 10:00–12:00pm *by appointment through Moodle scheduler only*

Teaching assistants:
Yingcong Tan <t_yingco@encs.concordia.ca>
Farhood Farahnak <farhood.farahnak@concordia.ca>
Laya Rafiee <laya.rafiee@concordia.ca>

Locations

Lecture:
H507 Wednesday 5:45–8:15pm

Labs:
H831 Tuesday 5:45–7:30pm (Section I)
H917 Tuesday 5:45–7:30pm (Section J)
H903 Friday 5:45–7:30pm (Section K)

Tentative Schedule

Week	Lecture	Other
Sep 4	*Lec1* Regression, Classification, Linear Models	*Lab1* Python, Numpy, Matplotlib
Sep 11	*Lec2* Clustering, Mixture Models	*Lab2* Linear Models
Sep 18	*Lec3* Loss Functions, Support Vector Machines	*Lab3* Clustering
Sep 25	*Lec4* Decision Trees, Random Forests	*Lab4* Support Vector Machines
Oct 2	*Lec5* Bootstrap, Bagging, Boosting, Ensembles	*Lab5* Random Forests
Oct 9	*Lec6* Generalization, Cross Validation, Model Selection	*Lab6* Boosting
Oct 16	**Midterm 1**	Extra office hours in H831
Oct 23	*Lec7* Neural Networks	*Lab7* Hyperparameter Search
Oct 30	*Lec8* Convolutional Networks, Recurrent Networks	*Lab8* Neural Networks
Nov 6	*Lec9* Autoencoders, Generative Models	*Lab9* Convolutional Networks
Nov 13	*Lec10* Gaussian Processes, Bayesian Optimization	*Lab10* Gaussian Processes
Nov 20	*Lec11* Reinforcement Learning	*Lab11* Reinforcement Learning
Nov 27	**Midterm 2**	Extra office hours in H831
Dec 4	**Project presentations** (custom projects only)	Scheduled slots in H507

Information provided in this course outline may be subject to change.

Textbook

There is no required textbook for this course, but course material will refer to supplementary (optional) readings within key references in the field:

- *Pattern Classification* by Richard O. Duda, Peter E. Hart, and David G. Stork (2001)
- *Pattern Recognition and Machine Learning* by Christopher M. Bishop (2006)

- *Information Theory, Inference, and Learning Algorithms* by David J.C. MacKay (2003)
- *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016)

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

Course evaluation

30% Midterm 1
30% Midterm 2
10% Labs
30% Project

Prerequisites

Linear algebra, multivariable calculus, and a basic probability are all essential to understanding the material in this course. Prior experience with Python and with numerical optimization are very helpful.

Projects

Projects must be done in groups of 2–4 students. Groups can choose between doing the "default project" or proposing a custom project related to their research.

- The **default project** is designed to give a comprehensive learning experience for students who are encountering machine learning for the first time. It comes with a well-defined data set, a description of research questions, and suggests possible experiments. Think of it as a large assignment where you have some freedom to try your own ideas for models/analyses on the data.
- A **custom project** can be proposed by groups who have a specific plan to apply machine learning to their graduate research. Custom projects must be proposed in writing to the instructor by no later than October 10. This is done by filling the *custom project proposal form* (available from the course Moodle page) and sending it to the instructor by e-mail. Students are encouraged to schedule a meeting during office hours to get feedback prior to writing up their proposal.

For the default project, the 30% project grade breaks down further as follows:

- 15% **4 page report** written using the LaTeX paper template provided in the course Moodle page,
- 10% **code and data** so that the TAs can reproduce your results on the lab computers,
- 5% **novelty component** to assess novel models/analyses applied to the project data.

For a custom project, the breakdown is slightly different:

- 20% **4 page report** written using the LaTeX paper template provided in the course Moodle page,
- 5% **code and data** so that the TAs can reproduce your results on the lab computers,
- 5% **presentation** wherein each group gives a short talk about their project and results.

Details of the "default project" will be posted by September 27, tentatively.

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, attempted communication during an exam, and everything else in the list of offences. If you are struggling, don't try to cheat, ask for extra help right away! We want everyone to have a great experience and to get a decent grade in the course!

Course content. Lecture and lab content will be hosted on Moodle and posted on a weekly basis.

Office hours. Instructor office hours are by appointment only. Available office hour time slots will be posted on Moodle each week, allowing students to book 15-minute time slots. By default these slots will be Thursday 10:00–12:00pm; if there is any change the instructor will announce it beforehand.

Communication. Outside office hours & labs, students must communicate with staff through **MOODLE ONLY** except in urgent situations and for specific exceptions (marking corrections, custom project proposals).

- Questions regarding course material should be posted on the course discussion board on Moodle. Students are encouraged to try to answer each others' questions if a TA has not yet answered.
- TAs are reachable through the DISCUSSION BOARD ONLY. TAs will not respond to personal e-mails or make personal appointments, as this is outside their official duties.
- 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. Students should be respectful of other students and of the instructor. Cell phones should be on *silent*. Laptops are allowed only in the last row of class. *Talking, using social media, or surfing the internet distracts your peers.* Show respect to your peers by refraining from such inconsiderate behaviour.

Readings. Prior to each midterm, students will be assigned to read an important research paper in machine learning. Everyone reads the same paper. The midterm will test your conceptual understanding of the paper.

Midterms. Midterms are held in class and are 2:15 long. The midterm will cover lectures, labs, and readings. Midterms and midterm grades will be returned to students within approximately 2 weeks of writing.

- Midterms are closed book.
- University-approved calculators are permitted and encouraged.
- Cheat sheets are *not* permitted.
- Bags, coats, and cell phones must be left at the edge of the room.
- Cell phones must be placed in *silent* prior to being stowed away in your bag/coat.

Labs. There are 11 labs worth 1% each up to a maximum of 10% of the total grade. The lab supervisor takes attendance. If the supervisor deems the student as having 'attempted' the lab during the alloted time, *i.e.* has made some minimal progress through the exercises, then the student receives the full mark for having attempted that lab. *Students are still responsible for knowing the complete lab material even if they did not complete the lab during the allotted lab time.* In such cases, think of the lab as a mini-homework.

- You must attend your assigned lab section. If your friends are in a different section, too bad.
- Students are encouraged to help each other with labs, but copying code is forbidden. You should write answers yourself and understand those answers. Otherwise you will struggle on the midterm.
- A student can miss at most one lab (one "free pass") 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*.
- On midterm weeks, there will be no labs. The Tuesday lab slot in H831 will be used as extra office hours to for students to ask questions while preparing for the midterm.

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

- Students who form their own group must send the list of names to the instructor via direct message on Moodle **by September 26**.
- Students who do not form their own group by September 26 will be randomly grouped.
- Custom project proposals must be submitted to the instructor by e-mail **no later than October 10** using the *custom project proposal form* available on the course Moodle page.
- Students who do their projects in groups are expected to contribute equitably. If an individual group member fails to contribute, that person's individual project grade may be lowered.
- Quality and clarity of writing will influence the grade of the "4 page paper" component, so if your group isn't strong in technical writing, consider completing a writing workshop or getting other help, but the final writing must be by the group and not contributed by anyone outside the group.

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 midterms becoming available, either by slipping a printed copy under his office door (EV 3.129) or by sending a scanned copy by e-mail. (Yes, a photo taken on a cell phone is acceptable.)