Applied Machine Learning

Syllabus and logistics

Isabeau Prémont-Schwarz

COMP 551 (Fall 2023)

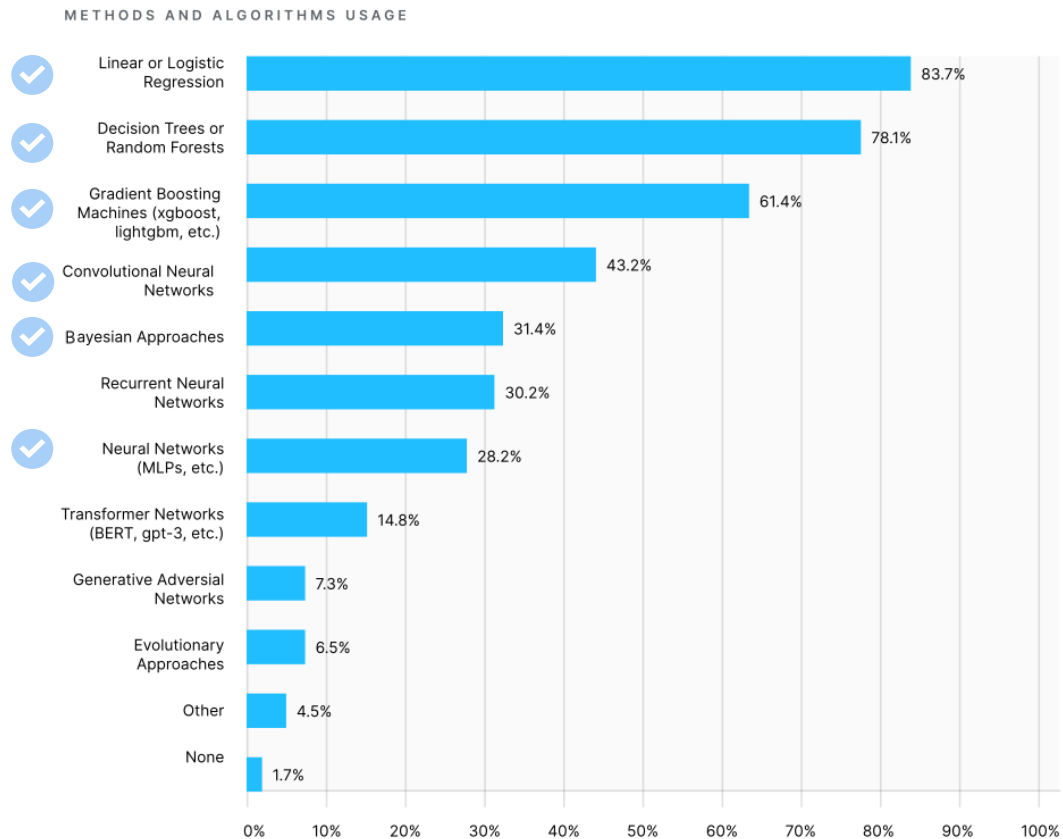
What is COMP 551?

- **Your one-stop shop for Machine Learning**

  (Pareto Principle ML Class: the 20% of knowledge that takes you 80% of the way)

- A difficult class where you will need to work hard

Most used ML Techniques



from 2020 Kaggle's survey on the state of Machine Learning and Data Science, you can read the full version here

What is COMP 551?

- **Your one-stop shop for Machine Learning**

  (Pareto Principle ML Class: the 20% of knowledge that takes you 80% of the way)

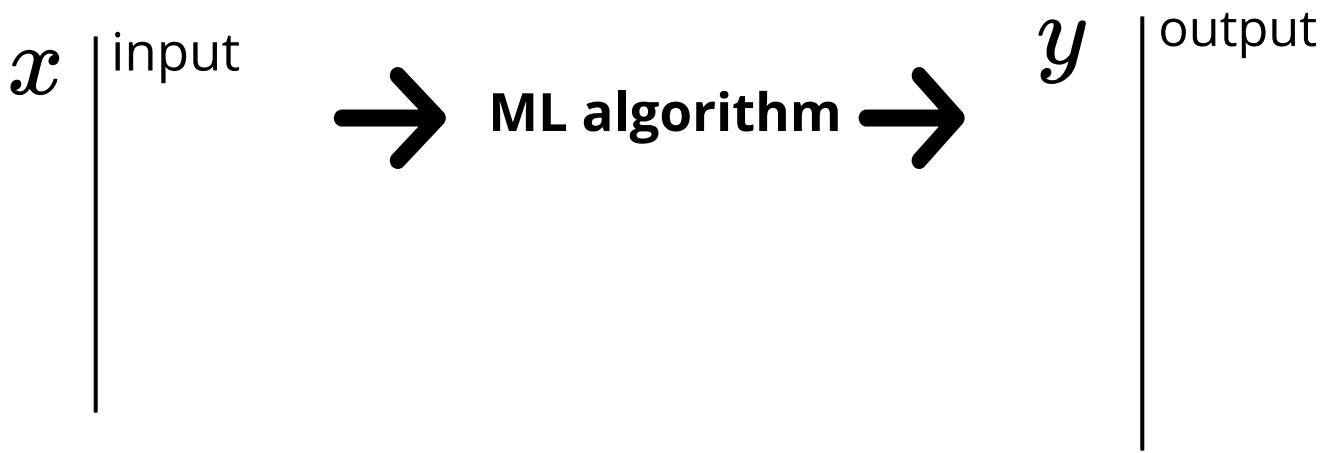
- A difficult class where you will need to work hard

What is Machine Learning?

What is Machine Learning?

- What the machine outputs in not pre-programmed
- A "training phase" where machine sees data/has experiences
- An "inference phase" where machine output what it learned

About this course



example

<tumorsize, texture, perimeter> = <18.2, 27.6, 117.5> → cancer = No

In-person Class

- **Lectures:** Tuesday & Thursdays, 11:35 - 12:55 (Montréal time)
  - Lectures will be **recorded** and uploaded in Mycourses
- **Course Website:** https://www.cs.mcgill.ca/~isabeau/COMP551/F23/
  - Syllabus, slides, deadlines, schedule, evaluation, etc.

Communications

- **Office Hours:**
  - Every day of the week!
  - Online (Zoom) and in person.
  - Hours and links on course website.

- **Course Email:** comp551mcgill@gmail.com
- **Course Discussion:** Ed through Mycourses

Prerequisites

- Strong linear algebra, probabilities, and Python programming is highly recommended: MATH 222, Math 113 and Comp 202
- How can I refresh my background knowledge to follow the lectures better? a lot of excellent online materials, see which one you can follow easier, you can also refer to these reviews on probability and linear algebra.
- Two quizzes on main concepts needed for lectures [with unlimited attempts allowed], due Sept. 12th, released already on mycourses.

Tutorials

Mid Sept.	Probability & Linear Algebra	
Mid Sept.	Python	https://www.python.org/
End Sept.	Scikit-learn	https://scikit-learn.org/
Early Oct.	Pytorch	https://pytorch.org/

About this course

Tentative Outline

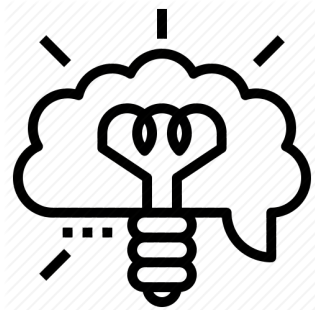
- Introduction
- Core concepts
- Linear regression
- Logistic and softmax regression
- Gradient descent methods
- Regularization
- Perceptrons & Multilayer Perceptrons
- Gradient computation and automatic differentiation
- Convolutional neural networks
- Nearest Neighbours
- Classification and regression trees
- Maximum likelihood and Bayesian Reasoning
- Naive Bayes
- Linear support vector machines
- Bagging & Boosting
- Unsupervised learning
- Dimensionality reduction
- Learning with graphs

About this course

complementary components

Theory

Lectures
Weekly Practice Quizzes
Midterm Exam
Understand the theory behind learning algorithms



Application

Codes in lectures
Mini-projects
Practice applying them in real-world

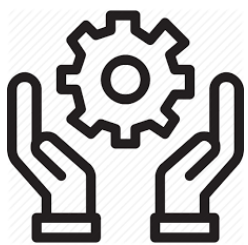
About this course:

Evaluation and grading

Regular Practice Quizzes - **10%**

Late Midterm exam - **40%**

Mini-projects - **50%** {group assignments}



About this course:

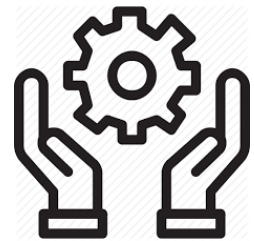
Evaluation and grading

Regular Practice Quizzes **- 10%**

- One per week to check the key concepts discussed in the last lecture
- Available until the start of the next Tue lecture
- Unlimited attempts are allowed

About this course:

Evaluation and grading



Mini-projects - **50%** {group assignments}

- Four programming assignments to be done in groups of three*, *no exception to this given the grading load on TAs
- Groups can stay the same between projects, you can also regroup when needed
- The goal is not to divide and conquer but to collaborate, do not wait for others to complete their tasks, help eachother do all the parts in the assignment
- All group members receive the same mark unless there are major complaints on not contributing, responding, etc. from group-mates, which will be resolved on a case-by-case basis. If a significant difficulty/conflict arises, please send an email to the course email, cc the group-TA and put 'Group-TA' in the title

Late submissions

All due dates are **11:59 pm** in Montreal unless stated otherwise. **No make-up quizzes** will be given. For mini-projects, 2^k% percent will be deducted per k days of delay.

If you experience barriers (including a covid related issue) to learning in this course, submitting the projects, etc., please do not hesitate to discuss them with me directly, and please make sure to put "551 special" in the header to make sure I see your email [for general course correspondence, please use the course email: comp551mcgill@gmail.com].

As a point of reference, you can reach the Office for Students with Disabilities at 514-398-6009

Code of Conduct

- Do not share or (re)post any of the course materials online. This includes: video lectures, codes, quizzes, zoom links, etc.
- Be respectful in the course forums and other communications
- Submit your own work for projects and quizzes

Academic Integrity

The ``McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offenses under the Code of Student Conduct and Disciplinary Procedures'' (see McGill's webpage for more information). (Approved by Senate on 29 January 2003)

Relevant Textbooks

No required textbook but slides will cover chapters from the following books, all available online, which can be used as reference materials.

- [Bishop] Pattern Recognition and Machine Learning *by Christopher Bishop (2007)*
- [Goodfellow] Deep Learning *by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016)*
- [Murphy] Machine Learning: A Probabilistic Perspective *by Kevin Murphy (2012)*
- Murphy'22] Probabilistic Machine Learning: An Introduction, by Kevin P. Murphy (2022)

Resources

Numerous great online resources at different levels, a selection is listed on the course website

Some may be more accessible than this course since they are designed for a different audience, but please note that this is a course designed for graduate students in computer science without ML background, with a heavy theory component.

Other Useful Online Resources

Learning plan

- metacademy
- https://github.com/afshinea/stanford-cs-229-machine-learning

Video Playlists

- StatQuest [🤩 Loved by many past students]
- FreeCodeCamp
- Essence of linear algebra and Neural Networks by 3Blue1Brown
- Mathematics for ML by David Rolnick

Courses with Playlist and/or Code

- Introduction to Machine Learning by Google
- Machine Learning by Stanford
- Deep Learning by UC Berkeley
- Hinton's Lectures on Neural Networks for Machine Learning
- Deep Learning & Linear Algebra courses by fastai
- Learning from Data by Caltech
- Deep Learning (with PyTorch) playlist and course by NYU
- Deep Learning by Stanford
- Deep Learning by deeplearning.ai
- Introduction to Deep Learning by MIT
- Information Theory, Pattern Recognition, and Neural Networks by David Mack

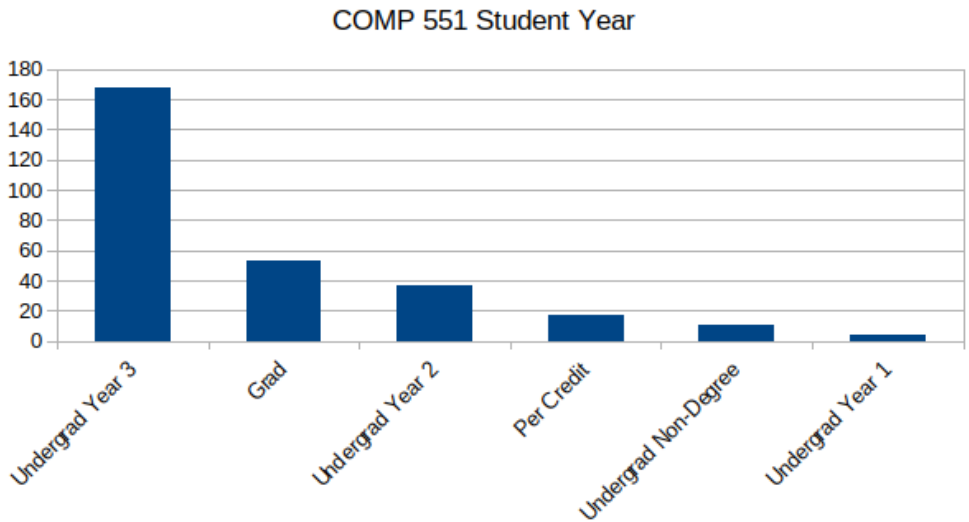
Good Blogs for Conceptual Understanding

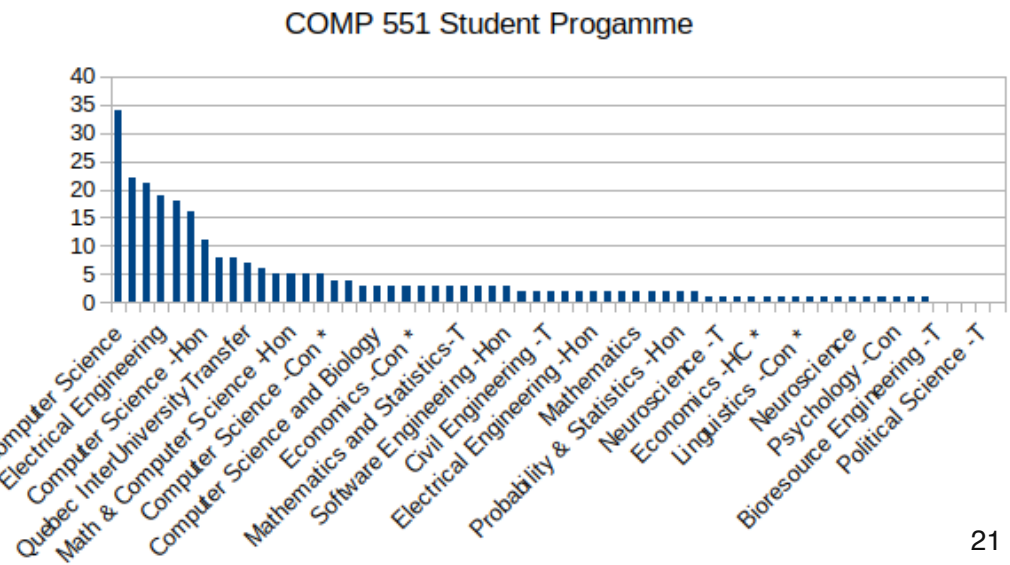
- Christopher Olah's Blog
- Fast.AI
- Distill.pub: Amazing Online and Interactive ML Journal, more accessible publications
- inFERENCE: Ferenc Huszár's Blog

Books with Code

- Probabilistic Machine Learning: An Introduction by Kevin Murphy (book 1)
- Dive into Deep Learning BY by Aston Zhang, Zachary Lipton, Mu Li, and Alexa
- Machine Learning Notebooks for O'Reilly book Hands-on Machine Lea TensorFlow

Who is in this class?





Questions?