

# MATH/COMP 562: Theory of Machine Learning

## Honours Mathematics for Machine Learning

Professor: Adam Oberman <https://www.adamoberman.net/>

Teaching Assistant: Viet Nguyen [baviet.nguyen@mail.mcgill.ca](mailto:baviet.nguyen@mail.mcgill.ca)

Class time: Tuesday and Thursday 2:35pm - 3:55 pm. [Lea 14](#)

Office Hours: TBD, and by appointment.

Lecture notes and assignments: <https://adam-oberman.github.io/>

Assignment submission: <https://mycourses2.mcgill.ca/>

### Prerequisites

- [MATH 462](#) <https://adam-oberman.github.io/>
- or [COMP 451](#) : <https://www.siamak.page/courses/COMP451F22/index.html>
- or ([COMP 551](#), [MATH 222](#), [MATH 223](#) and [MATH 324](#)) <http://www.reirab.com/Teaching/AML23/index.html>
- or [ECSE 551](#) <https://ismart.ece.mcgill.ca/teaching/>

### Course Description

A mathematically rigorous approach to Machine Learning (ML), with a focus on a rigorous presentation of ML models. Proofs of in-distribution generalization bounds.

- This course assumes familiarity with the ML models discussed, and experience implementing the models. It provides a deeper theoretical analysis,
- There currently is no useful theory for deep learning methods, they are still primarily achieving success using engineering methods. This course aims to cover theory topics from machine learning which experts believe will be relevant to the to-be-developed theory for deep learning.

### Audience

- Math and Stats and theory oriented Computer Science graduate students.
- Advanced undergraduate students who have prior background in implementing Machine Learning methods, and who plan to go on to graduate school.

Students will be expected to have seen and coded ML models before. Experience with mathematical proofs and with probability is expected.

### How is this course different from other courses?

- Some other courses focus on implementation, and assignments can be kaggle competitions to train the most accurate model.
- Other courses mix theory with algorithms, and spend more time on deep learning heuristics and implementation

- Some courses have statistical machine learning textbooks/Bayesian approach as a foundation, for example, [Pattern Recognition and Machine Learning](#) by Christopher Bishop (2007), or [Machine Learning: A Probabilistic Perspective](#) by Kevin Murphy (2012)
- Older ML theory courses cover topics such as VC dimension, which is less relevant to differentiable models, such a neural networks.

This course is a theory course covering material relevant to deep learning, but it spends very little time on deep learning. The reason for this is that the deep learning theory is still missing/in very preliminary stages. Instead, we cover established material which will be useful in the future. We avoid trendy theoretical topics which may not be important a short time from now.

This is a deliberate choice: the list of trendy theory topics in deep learning has changed every year for the past five years. Instead we cover material which is expected to remain relevant in the future.

## Grading

- Final exam: 30%
- HW assignments : 30% (5-6, approximately biweekly)
- Project: 30%
- Attendance: 5%.
- Participation: 5%.

The material covered in class will form the basis for homework problems. Relevant textbook sections will be referenced on the course page. There will also have a programming problems designed to improve understanding of the material. However, it is assumed that students will have programmed before, and implementation is not the focus of the course.

The final exam will have questions which are similar to a selection of problems from the homework, so there should be no surprises about the material to study. Completing and understanding the HW will be essential for the final exam.

Additional topics covered at the end of the courses will not be on the HW or the final exam.

In person attendance of lectures is a requirement for this course. Participation points come from refereeing other student projects.

## Class Schedule and key dates

Refer to <https://www.mcgill.ca/importantdates/key-dates>.

- First Class: Thursday, Jan 5
- Winter Reading Break (no class): Feb 28, March 2
- Last Class: Tuesday, April 11

## Textbook:

Learning Theory from First Principles, (January 1, 2023 edition) [Free PDF](#), by [Francis Bach](#)

This a very recent book by a highly respected researcher. It has a modern approach to the topics covered, and does a good job of addressing confusing aspects of the theory. It's concise and very mathematical, so may be hard to read for students. The lectures are designed to make the material more understandable.

- Feel free to read ahead in the textbood before lectures
- Expect to refer to textbook for details.
- Avoid the advanced sections marked with a symbol until a second reading.

## Additional references

- [Mathematics for Machine Learning by Diesenroth](#) Review of prerequisites, with a ML focus.
- [Understanding Machine Learning: From Theory to Algorithms](#) by Shalev-Shwartz and Ben David
- [Foundations of Machine Learning](#) by Mohri, Rostamizadeh, Talwalkar.

Use Shalev-Shwartz for introduction and definitions. Use Mohri for the proofs, which are more precise.

- [High dimensional statistics, a non-asymptotic viewpoint](#) by Martin J Wainwright. Advanced book on concentration of measure topics in machine learning. The first chapters are a good source of advanced material on concentration of measure. However, this material is covered in Bach Ch 1.

## Project

Instructions [MathComp 562 Project Description.pdf](#)

- Stage 1: choose topics and groups, TODO Tues Jan 31st. Project suggestions will be provided: [Previous Suggested Paper Links.pdf](#)
- Stage 2: project outline and work plan: due Thurs Feb 3rd.
- Stage 3: hand in write up. Optional spotlight presentations.

## Lecture Plan

(Schedule may be adjusted according to the class needs. Refer to course page)

Introduction

- Introduction to supervised learning, Bach Ch2, (3 lectures)
- Linear least-squares regression theory, Bach Ch 3, implementation, (3 lectures)

Generalization bounds for learning algorithms

- Empirical Risk Minimization,
  - Bach Ch 4, (3 lectures)
  - Concentration of Measure, McDiarmid's inequality, Bach Ch1, (1 lecture)
- Stability approach, Shalev-Shwartz Ch 13, Mohri Ch 14 (1 lecture)

- See also Bach Ch 5 for definitions of convex, strongly convex, smooth functions
- Optimization for Machine Learning, Bach Ch 5, 3 lectures
- Local Averaging Methods, Bach Ch6, 2 lectures
- Kernel Methods, Bach Ch 7, 4 lectures
- Neural Networks, Bach Ch 9 and other references, 4 lectures
- Additional topics and final exam review, remaining (2-3) lectures.