MATH 462 Fall 2022, First day handout

Honours Mathematics for Machine Learning

Professor: Adam Oberman

Class: Monday, Wedensday, 10:05 am-11:25 am BURN 1214

Office Hours: BURN 1106, MW 11:30-12:00, and by appointment. Additional office hours TBA.

Audience:

Math and Stats Majors/Honours students, Computer Science students.

Course Description:

A mathematically rigorous approach to Machine Learning (ML). This course will cover the mathematical models which go into current machine learning models, as well as deep learning architectures, and application areas. It will provide the necessary background for understanding deep learning models and reading contemporary research papers. The sequel, Math 562, will go more deeply into mathematical aspects, such as statistical learning theory, and regularization.

Prerequisites

- Math 223/Math 248 or equivalent (linear algebra majors/honours)
- Math 222 (Calculus 3)
- Math 358 Honours Vector Calculus/Math 314 Advanced Calculus

Math 324/Math 357 (Statistics) is not needed for this course, (probability course is needed).

Course Webpages

- Course material: https://adam-oberman.github.io/
- Assigment submission: https://mycourses2.mcgill.ca/

Grading

- 5-6 HW assignments : 20%
- Attendance and participation : 10%. In person attendance of lectures is a requirement for this course.
- Midterm Exam : 30%
- Final exam : 40%

Key Dates

https://www.mcgill.ca/importantdates/key-dates#Fall_2022

- Classes begin Weds August 31,
- Last class: Wednesday Nov 30th.
- Midterm Exam:
  - Wednesday Nov 9th, in class.
- Exam period : Dec 7 - 21st.
- No Class:
  - Monday Sept 5th (labour day)
  - Monday Oct 10, Weds Oct 12 (Fall reading break)
  - Monday Dec 5th (class cancelled, study instead)
- Extra class: Thursday Oct 13 (**makeup class**)

Related Courses

- Math 308 Fundamentals of Statistical Learning. (offered Winter 2023)
  - Some overlap, but more probabilistic models. OK to take both.
- COMP 451 Fundamentals of Machine Learning
  - Overlap with the introduction to this course, but COMP 451 focuses on probabilistic machine learning and implementation. Normally mutually exclusive.
- COMP 551 Applied Machine Learning
  - A complementary course, focused on implementation.
- Math 562, Winter 2022
  - Some overlap with Math 462 in the first few weeks, but then covers different, and more advanced, topics.
- COMP 424 Artificial Intelligence

  - Complementary, no overlap. Covers AI topics (e.g. topics from Sections II and III of AIMA book, tabular RL) but does not cover machine learning.
- COMP 579 Reinforcement Learning (RL)
  - Advanced course in RL, no overlap.

References

In most cases, PDF of relevant chapters are easily obtained.

- Mathematics for Machine Learning by Diesenroth [MML]
  - Review of prerequisites, with a ML focus.
- Understanding Machine Learning: From Theory to Algorithms [SS] by Shalev-Shwartz and Ben David This book is very good for presenting machine learning problems.
  - Several Chapters
- Artificial Intelligence: A Modern Approach, 4th ed. by Stuart Russell and Peter Norvig (US edition). e-textbook from the publisher for approx $50. Exercises: https://aimacode.github.io/aima-exercises/
  - Chapter 19 for supervised learning
- Foundations of Machine Learning [M] by Mohri, Rostamizadeh, Talwalkar.
  - Alternative to SS
- High dimensional statistics, a non-asymptotic viewpoint [WW] by Martin J Wainwright
  - Ch 12 for kernels
- A primer on PDEs by Sandro Salsa, [Salsa]
  - Chapter 7 for Hilbert Space Theory

Topic Schedule

Planned outline of the topics, schedule may change, refer to webpage for updated schedule.

Machine Learning (12-13 lectures)

Midterm covers this material:

- Supervised Learning [AI Ch 19] (4-5 lectures)
  - Regression (Matrix Calculus),
  - Classification,
  - Decision Trees,

  - Learning Theory, VC dimension
  - Model Selection and Optimization, nonparametric models.
- Unsupervised Learning [SS 22, 23] (3 lectures)
  - Clustering: K-means and variational algorithms [SS 22]
  - Clustering Impossibility Theorem [SS 22.5]
  - Linear Algebra Review [e.g. Mohri Appendix, MML Ch 2]
  - PCA (SS 23.1)
- Probabilistic Learning [AI Ch 21] (1 lecture)
  - Density Estimation, Generative vs. Discriminative [AI 21.2].
    - Computational limitations.
- Deep Learning and Computer Vision (3-4 lectures)
  - Vector Calculus Review ML: gradients, chain rule [MML Ch 5], Implicit Differentiation
  - CNN, RNN, autograd, GANs [AI Ch 22]
  - Losses and Softmax Calculus [teacher's notes]
  - Computer vision [AI Ch 27]
    - Image features [AI 27.3]: edges, segmentation
    - Image classification [AI 27.4]: convolutions

Functional Analysis for ML (6-8 lectures)

- Hilbert Space (HS) [Salsa 7.2]
- Riesz Rep Theorem, Linear operators in HS [Salsa 7.3]
  - When is the evaluation operator is continuous?
- Fourier series, Dirichlet and Fejer Kernels, [e.g. Strauss 5.1, 5.2] Convolution Theorem
- PSD kernels and Reproducing Kernel HS [Wainwright Ch 12]
- Representation Theorem
- Mercer's Theorem

Convex optimization for ML and stability-based learning bounds (6 lectures)

- Convexity, convex functions, strong convexity [My notes / Boyd]
- Inequalites and calculus, Jensen's ineq.
  - Sidebar: KL-divergence and Entropy [WW 3.1.1]
- Gradient descent and convergence rates [My notes / SS 14]

- Abstract and mini-batch SGD (statement of rates) [My notes / SS 14]
- Convex Learning problems [SS 12]
- Regularization and Stability, Learning bounds [SS 13]
    - Abstract Stability, other stability examples
- Learning bounds via stability for: regression, PCA, classification, kernel methods.