Department of Mathematics & Statistics
Concordia University

STAT 497 (MAST 679/MAST 881), Sec. DS
Topics in Statistics & Probability
'Mathematics of Data Science'
Fall 2024

Instructor: Dr. S. Brugiapaglia, Office: LB 921.09 SGW, Phone: 848-2424, Ext. 4250
Email: simone.brugiapaglia@concordia.ca

Class Schedule: Thursdays, 17:45-20:15, H 609 SGW.
Mid-term break: no class between October 15, 2024, and October 20, 2024.

Office Hours: Wednesdays, 14:30-16:00, LB 921.09 SGW.

Description: This course is an introduction to the mathematical foundations of data science. Topics covered tentatively include:

- **Machine learning theory**
  Rudiments of statistical learning theory, VC dimension, PAC-learnability, support vector machines, reproducing kernel Hilbert spaces, regression and regularization, clustering, dimensionality reduction;
- **Compressive sensing**
  l0 and l1 minimization, restricted isometry property, sample complexity, sparse recovery algorithms;
- **Elements of optimization theory**
  Basic convex optimization, linear programming, duality theory, instances of nonconvex optimization;
- **Deep learning**
  ReLU networks, universal approximation theory, advantages of depth, neural network training, practical existence theory.

Although the course will focus on theoretical aspects, it will also include computational illustrations.

Main Textbook: *Mathematical Pictures at a Data Science Exhibition*, by Simon Foucart. Cambridge University Press, 2022.
E-book: https://concordiauniversity.on.worldcat.org/oclc/1314389857

Other References: Other useful references include (but are not limited to) the following:

A Mathematical Introduction to Compressive Sensing by S. Foucart and H. Rauhut. Springer, 2013.
Available online through the Concordia library website:
E-book: https://link-springer-com.lib-ezproxy.concordia.ca/book/10.1007%2F978-0-8176-4948-7

Foundations of Machine Learning (second edition), by
M. Mohri, A. Rostamizadeh, and A. Talwalkar. MIT Press, 2018.
E-book: https://cs.nyu.edu/~mohri/mlbook/

Foundations of data science by A. Blum, J. Hopcroft, and R. Kannan. Cambridge University Press, 2020.
E-book: https://home.ttic.edu/~avrim/book.pdf

High-dimensional probability: An introduction with applications in data science by R. Vershynin. Vol. 47. Cambridge University Press, 2018.
E-book: https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html

Pre-requisites: Working knowledge of linear algebra, probability theory, and analysis. General familiarity with basic notions about Hilbert and normed spaces, linear operators, and knowledge of measure theory are not essential but could be helpful.

Assignments: Homework will be assigned approximately every other week and will focus on theoretical problems.

Exams: There will be a midterm exam. Moreover, students will be required to complete a final independent project on a topic approved by the instructor and based on recent research papers in the field. The project will be presented in the form of an oral presentation *and* submitted as a written report. The focus of the project can be theoretical or computational, based on the student's background and interests.

Grading Scheme: 20% Homework + 20% Midterm + 60% Final Project (proposal, oral presentation, written report).

PhD students will be required to do additional work (to be determined) compared with their MA/MSc classmates.

PLEASE NOTE: It is the Department's policy that tests missed for any reason, including illness, cannot be made up. If you miss the midterm because of illness (Short-Term Absence form or valid medical note required), the grading scheme will be modified as 30% Homework + 70% Final Project.

If the grading scheme for this course includes graded assignments, a reasonable and representative subset of each assignment may be graded. Students will not be told in advance which subset of the assigned problems will be marked and should therefore attempt all assigned problems.

Communication: Communication between the students and the instructor will take place in class, via Moodle announcements, or emails. Students are responsible for reading and taking note of all electronic communication from the instructor and the University.

Academic Integrity and the Academic Code of Conduct
This course is governed by Concordia University's policies on Academic Integrity and the Academic Code of Conduct as set forth in the Undergraduate Calendar and the Graduate Calendar. Students are expected to familiarize themselves with these policies and conduct themselves accordingly. "Concordia University has several resources available to students to better understand and uphold academic integrity. Concordia's website on academic integrity can be found at the following address, which also includes links to each Faculty and the School of Graduate Studies: concordia.ca/students/academic-integrity." *[Undergraduate Calendar, Sec 17.10.2]*

Behaviour
All individuals participating in courses are expected to be professional and constructive throughout the course, including in their communications.

Concordia students are subject to the Code of Rights and Responsibilities which applies both when students are physically and virtually engaged in any University activity, including classes, seminars, meetings, etc. Students engaged in University activities must respect this Code when engaging with any members of the Concordia community, including faculty, staff, and students, whether such interactions are verbal or in writing, face to face or online/virtual. Failing to comply with the Code may result in charges and sanctions, as outlined in the Code.

Intellectual Property
Content belonging to instructors shared in online courses, including, but not limited to, online lectures, course notes, and video recordings of classes remain the intellectual property of the faculty member. It may not be distributed, published or broadcast, in whole or in part, without the express permission of the faculty member. Students are also forbidden to use their own means of recording any elements of an online class or lecture without express permission of the instructor. Any unauthorized sharing of course content may constitute a breach of the Academic Code of Conduct and/or the Code of Rights

and Responsibilities. As specified in the Policy on Intellectual Property, the University does not claim any ownership of or interest in any student IP. All university members retain copyright over their work.

Extraordinary circumstances

In the event of extraordinary circumstances and pursuant to the Academic Regulations the University may modify the delivery, content, structure, forum, location and/or evaluation scheme. In the event of such extraordinary circumstances, students will be informed of the change.