CHEMENG 4H03
Big Data Methods and Modeling in Chemical and Materials Engineering

Course outline, Winter 2024
Last updated: January 6, 2024

Administrative details

Mode of instruction

- Instruction is primarily based on in-person lectures, including in-class discussions and workshop-style exercises. From the second lecture onward, lectures will also have automatic echo360 recordings that will expire after 2 weeks.
- Office hours will be held as video chats in Microsoft Teams.
- Masking is encouraged, and lectures may switch to Microsoft Teams in an emergency.
- All critical communication about deadlines etc. will be posted as "announcements" in Avenue to Learn. Please check the course's Avenue page daily for announcements.

Instructor

Dr. Kamil A. Khan (he/him); (kamilkhan@mcmaster.ca)

Office: JHE 202
Teams ID: khank11

Lecture times, TAs, and office hours

Included in a syllabus addendum in Avenue.

Course and learning objectives

In today's chemical process industries, process data is increasingly abundant, yet it isn't always clear how to infer useful information from this data. Chemical Engineering 4H03 focuses on the application of methods to parse, filter, interpret, learn and extract value from large industrial and consumer data sets. Topics include dimensionality reduction

(PCA, PLS), soft sensors for process monitoring, data clustering, introduction to artificial neural networks (ANNs) and variants, and modeling/integrating these methods using commercial software.

After completing this course, the student should be able to:

- Perform model reduction techniques including principal component analysis (PCA) and partial-least squares (PLS).
- Perform model identification techniques including generation of neural networks, and use the state-of-the-art packages PyTorch and TensorFlow.
- Perform various clustering techniques.
- Weigh statistical significance of an identified model, and determine if overfitting or underfitting occur.
- Visualize large data sets to identify trends and key observations.
- Implement relevant algorithms in MATLAB (and/or Julia).
- Understand and discuss the theoretical justifications underlying covered techniques; these will typically involve aspects of linear algebra and probability/statistics.

Course assessment

Final grade breakdown

	% of final grade
Assignments	20
Midterm exam #1	20
Term project	20
Final exam	40

- In addition, as "stretch goals":
  - an additional $+1$ point will be given for using the programming language Julia in place of MATLAB in at least one assignment question, and
  - a further $+2$ points will be given for using Julia instead of MATLAB in most assignment questions that nominally require MATLAB.
- If any MSAF is approved, then that work's weight will be transferred to the final exam.
- Final course grades may be adjusted at the discretion of the instructor.

Tests and exam

- There will be a midterm test and a final exam; time and date TBD.
- A "McMaster standard calculator" may be used in each test/exam.
- The test/exam are "closed book", with a formula sheet attached (and maintained on Avenue). In addition, one double-sided sheet of prepared notes is permitted in the midterm, and two double-sided sheets are permitted in the final exam.
- The final exam tests cumulative knowledge.

Assignments

There will be 4 or 5 assignments given throughout the semester. Your lowest-scoring assignment (*after* any MSAFs have been considered) will not count toward your final grade. Several assignment problems will involve coding in MATLAB/Julia or Python.

Assignments must be prepared a certain way, and submitted through Avenue. Please refer to the Assignment Policy (posted in Avenue) for details regarding submission, grading, and collaboration. Late assignment submissions will receive the following penalties:

- -20 percentage points if they are one day late,
- -50 percentage points if they are two days late,
- and they will not be accepted or graded after that.

Course project

A course project will be assigned after the midterm recess, with a written report and short oral presentation due in the last week of classes. The project is to be carried out in groups of 3 or 4, and will have both a research component and an implementation component. A list of topic suggestions will be provided, and grading rubrics will be posted for the report and presentation. The instructor and/or TAs will check in on the progress of each group. Each student will submit a confidential project contribution assessment form.

If the project makes use of any external references/sources, then those must be cited correctly in the report, in any reasonable citation style.

Provisional course outline

- Week 1: Introduction; review of statistics and numerical coding
- Week 2: Review of linear algebra; singular-value decomposition (SVD)
- Week 3: Model reduction; Principal Component Analysis (PCA)
- Week 4: Partial Least Squares/Projection of Latent Structures (PLS)
- Week 5: Matrix completion and missing data

- Week 6: Clustering and categorizing I
- Midterm recess
- Week 7: Clustering and categorizing II
- Week 8: Artificial neural networks (ANN) and variants
- Week 9: Implementation in PyTorch/TensorFlow
- Week 10: Big Data techniques for dynamic systems
- Week 11: Survey of advanced topics I
- Week 12: Survey of advanced topics II
- Week 13: Conclusions and project presentations

Materials & fees

- No textbooks are required for this class; lecture slides will be posted to Avenue as `.pdf` files (typically a couple of days before the relevant lectures), and slides annotated in-class will be posted afterwards (typically the day after a lecture).
- The following texts may nevertheless be helpful:
  - S.L. Brunton and J.N. Kutz (2019): *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, available online through McMaster's library;
  - S. Shalev-Shwartz and S. Ben-David (2014): *Understanding Machine Learning: From Theory to Algorithms*, available for legitmate download at the authors' website,
  - and any introductory books on linear algebra or statistics.
- It is recommended (but not required) to use a personal laptop computer for the programming assignments in this course, and to bring it to each lecture (starting in lecture #2). The cluster computers in JHE may also be used for assignments and the project.
- The midterm/final exams will require a McMaster standard scientific calculator.

Approved advisory statements

An excerpt from McMaster's Undergraduate Course Management Policy is attached as the final three pages of this document. The remainder of that policy applies as well.

Additional statements

Students are not permitted to use generative AI in this course. In alignment with McMaster's academic integrity policy, it "shall be an offence knowingly to ... submit academic work for assessment that was purchased or acquired from another source". This includes work created by generative AI tools. Also stated in the policy is the following: "Contract Cheating is the act of "outsourcing of student work to third parties" (Lancaster & Clarke, 2016, p. 639) with or without payment." Using Generative AI tools is a form of contract cheating. Charges of academic dishonesty will be brought forward to the Office of Academic Integrity.

At certain points in the course it may make sense to modify the schedule. The instructor may modify elements of the course and will notify students accordingly (in class, and as an announcement in the course's Avenue page).

COURSE OUTLINE – APPROVED ADVISORY STATEMENTS

ACADEMIC INTEGRITY

You are expected to exhibit honesty and use ethical behaviour in all aspects of the learning process. Academic credentials you earn are rooted in principles of honesty and academic integrity. **It is your responsibility to understand what constitutes academic dishonesty.**

Academic dishonesty is to knowingly act or fail to act in a way that results or could result in unearned academic credit or advantage. This behaviour can result in serious consequences, e.g. the grade of zero on an assignment, loss of credit with a notation on the transcript (notation reads: "Grade of F assigned for academic dishonesty"), and/or suspension or expulsion from the university. For information on the various types of academic dishonesty please refer to the *Academic Integrity Policy*, located at https://secretariat.mcmaster.ca/university-policies-procedures-guidelines/

The following illustrates only three forms of academic dishonesty:

- plagiarism, e.g. the submission of work that is not one's own or for which other credit has been obtained.
- improper collaboration in group work.
- copying or using unauthorized aids in tests and examinations.

AUTHENTICITY / PLAGIARISM DETECTION

Some courses may use a web-based service (Turnitin.com) to reveal authenticity and ownership of student submitted work. For courses using such software, students will be expected to submit their work electronically either directly to Turnitin.com or via an online learning platform (e.g. Avenue to Learn, etc.) using plagiarism detection (a service supported by Turnitin.com) so it can be checked for academic dishonesty.

Students who do not wish their work to be submitted through the plagiarism detection software must inform the Instructor before the assignment is due. No penalty will be assigned to a student who does not submit work to the plagiarism detection software. **All submitted work is subject to normal verification that standards of academic integrity have been upheld** (e.g., on-line search, other software, etc.). For more details about McMaster's use of Turnitin.com please go to www.mcmaster.ca/academicintegrity.

COURSES WITH AN ON-LINE ELEMENT

Some courses may use on-line elements (e.g. e-mail, Avenue to Learn, LearnLink, web pages, capa, Moodle, ThinkingCap, etc.). Students should be aware that, when they access the electronic components of a course using these elements, private information such as first and last names, user names for the McMaster e-mail accounts, and program affiliation may become apparent to all other students in the same course. The available information is dependent on the technology used. Continuation in a course that uses on-line elements will be deemed consent to this disclosure. If you have any questions or concerns about such disclosure please discuss this with the course instructor.

ONLINE PROCTORING

Some courses may use online proctoring software for tests and exams. This software may require students to turn on their video camera, present identification, monitor and record their computer activities, and/or lock/restrict their browser or other applications/software during tests or exams. This software may be required to be installed before the test/exam begins.

CONDUCT EXPECTATIONS

As a McMaster student, you have the right to experience, and the responsibility to demonstrate, respectful and dignified interactions within all of our living, learning and working communities. These expectations are described in the *Code of Student Rights & Responsibilities* (the "Code"). All students share the responsibility of maintaining a positive environment for the academic and personal growth of all McMaster community members, **whether in person or online**.

It is essential that students be mindful of their interactions online, as the Code remains in effect in virtual learning environments. The Code applies to any interactions that adversely affect, disrupt, or interfere with reasonable participation in University activities. Student disruptions or behaviours that interfere with university functions on online platforms (e.g. use of Avenue 2 Learn, WebEx or Zoom for delivery), will be taken very seriously and will be investigated. Outcomes may include restriction or removal of the involved students' access to these platforms.

ACADEMIC ACCOMMODATION OF STUDENTS WITH DISABILITIES

Students with disabilities who require academic accommodation must contact Student Accessibility Services (SAS) at 905-525-9140 ext. 28652 or sas@mcmaster.ca mailto:sas@mcmaster.cato make arrangements with a Program Coordinator. For further information, consult McMaster University's *Academic Accommodation of Students with Disabilities* policy.

REQUESTS FOR RELIEF FOR MISSED ACADEMIC TERM WORK

In the event of an absence for medical or other reasons, students should review and follow the *Policy on Requests for Relief for Missed Academic Term Work*.[3]

ACADEMIC ACCOMMODATION FOR RELIGIOUS, INDIGENOUS OR SPIRITUAL OBSERVANCES (RISO)

Students requiring academic accommodation based on religious, indigenous or spiritual observances should follow the procedures set out in the RISO policy. Students should submit their request to their Faculty Office ***normally within 10 working days*** of the beginning of term in which they anticipate a need for accommodation or to the Registrar's Office prior to their examinations. Students should also contact their instructors as soon as possible to make alternative arrangements for classes, assignments, and tests.

COPYRIGHT AND RECORDING

Students are advised that lectures, demonstrations, performances, and any other course material provided by an instructor include copyright protected works. The Copyright Act and copyright law protect every original literary, dramatic, musical and artistic work, **including lectures** by University instructors

The recording of lectures, tutorials, or other methods of instruction may occur during a course. Recording may be done by either the instructor for the purpose of authorized distribution, or by a student for the purpose of personal study. Students should be aware that their voice and/or image may be recorded by others during the class. Please speak with the instructor if this is a concern for you.

[3] Updated to the *Policy on Requests for Relief for Missed Academic Term Work* effective May 1, 2021

EXTREME CIRCUMSTANCES

The University reserves the right to change the dates and deadlines for any or all courses in extreme circumstances (e.g., severe weather, labour disruptions, etc.). Changes will be communicated through regular McMaster communication channels, such as McMaster Daily News, Avenue to Learn and/or McMaster email.