

CAS 775, Theoretical Foundations of Unsupervised Learning

Course Outline, Fall 2021

Hassan Ashtiani, McMaster University

1 Lectures

- Hassan Ashtiani, zokaeiam@mcmaster.ca, ITB 246
- TuFr 2:30-4pm (virtual)
- Office Hours: TuFr 4-4:30pm (virtual - right after the class)

The virtual lectures are held on Zoom. Use this *link*.

Alternatively the Meeting ID is : 963 3801 5659 and the Passcode is : 359127

Please do not share the Zoom information with others. The lectures will be recorded. The whiteboard notes will be shared with the students after each lecture but it is recommended that the students write their own notes.

2 Structure and Workload

This is a theory course with a focus on the modern aspects of distribution learning. In the first half of the course we lay down the foundations of distribution learning. There is an assignment due around the midpoint of the semester. The course will continue with other related topics and will finish with student presentations.

3 Final Grade Breakdown

- Assignment: 40%
- Presentation: 50%
- Participation in Presentations: 10%

4 Textbook

The course does not follow a specific textbook as some of the discussed topics are based on recent papers in the field. However, for the introductory parts, the following textbook is useful:

- Combinatorial Methods in Density Estimation by Luc Devroye and Gabor Lugosi

5 MS Teams and Avenue

We will use MS Teams for the announcements and Q&A. If you are not automatically enrolled in the team please let the instructor know.

We will use Avenue2Learn for uploading the files (such as recorded lectures) and also as a drop box for the assignment.

6 The Assignment

- There will be one assignment, and it will be based on the materials covered roughly in the first half of the class.
- The assignment consists mathematical questions. Your solutions should be typeset in Latex and uploaded as a pdf file in the Avenue drop box by the deadline. Unless in extra-ordinary situations late submissions are not accepted.

7 The Presentations

Each student chooses a paper from a list of theoretical papers compiled by the instructor. Each paper can be assigned to at most one student. The students can also suggest a relevant paper; in that case, the instructor has to approve the paper.

The students will then try to understand the paper and present it. The papers are often highly theoretical and challenging to comprehend; it may help to use other resources, such as reading related material, watching video lectures, etc. The presentations are typically around 15 minutes. Each presentation is followed by a short Q&A session. Participation in (some of) the Q&As is required. More details will be given later.

8 About the Contents

Given samples from an (unknown) data generating distribution, how/when can we estimate/approximate the distribution? This fundamental question has been studied by statisticians for more than a century. Yet, modern challenges such as handling high dimensional, noisy, and private data has revealed that our understanding of this question is quite limited. The computer science and machine learning communities have therefore stepped in, and have made significant progress in addressing these issues.

In this class, we will cover some classic and some more recent results in distribution learning. The main model that we use is the Probably Approximately Correct (PAC) learning. In particular, we will study the sample complexity of learning various classes of distributions. We will also discuss the design of computationally efficient estimators.

More concretely, some of the topics discussed in the class are Scheffe estimator, Scheffe tournament, Yatracos's minimum distance estimator, distribution compression schemes, robust distribution learning, and private distribution learning. The last few weeks of the class will be dedicated to student presentations.

9 Course Objectives and Outcomes

- Familiarizing with the fundamentals of distribution learning
- Familiarizing with the recent advances in distribution learning and the relevant open problems
- Developing mathematical maturity to be able to follow formal arguments and understand theoretical papers in the area
- Becoming acquainted with the standard models of learning and to some extent developing the ability to formulate learning problems
- Practicing problem solving and writing formal proofs
- Practicing presentation skills

10 Disabilities

Students with disabilities who require academic accommodation must contact Student Accessibility Services (SAS) to make arrangements with a Program Coordinator. Student Accessibility Services can be contacted by phone 905-525-9140 ext. 28652 or e-mail sas@mcmaster.ca. For further information, consult McMaster University's Academic Accommodation of Students with Disabilities policy.

11 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.

12 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. A2L, 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,

13 Courses with an online element

Some courses may use on-line elements (e.g. e-mail, Avenue to Learn (A2L), LearnLink, web pages, capa, Moodle, ThinkingCap, Piazza, MS Teams, 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.

14 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.

15 Conduct Expectation

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.

16 Academic accomodations for religious, indigenou or spiritual observations (RISO)

Students requiring academic accommodation based on religious, indigenou or spiritual observances should follow the procedures set out in the RISO policy. Students requiring a RISO accommodation 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.

17 Copyright and recordings

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.

18 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, A2L and/or McMaster email.

Moreover, the students should check out MS Teams for announcements regarding the course.