CMSC 422 - Introduction to Machine Learning

Syllabus

Instructor

Mohammad Nayeem Teli, nayeem@umd.edu.
Office: IRB 2224, Office phone: 301-405-0405.

Welcome

Welcome to CMSC 422. Machine Learning studies representations and algorithms that allow machines to improve their performance on a task from experience. This is a broad overview of existing methods for machine learning and an introduction to adaptive systems in general. Emphasis is given to practical aspects of machine learning and data mining.

Textbook

Our primary source of readings will be A Course in Machine Learning, a collection of notes by Hal Daumé III, which provides a gentle and thorough introduction to the field of machine learning.

Topics

This course provides a broad introduction to machine learning and statistical inference. We will attempt to cover the following topics:

- **Supervised Learning**
  - Decision trees and inductive bias
  - Geometry and nearest neighbors
  - Perceptron
  - Beyond binary classification
  - Linear models and gradient descent
  - Support Vector Machines
  - Naive Bayes models and probabilistic modeling
  - Neural networks
  - Kernels
  - Ensemble Learning
- **Unsupervised Learning**
  - Clustering
  - Principal Component Analysis (PCA)
- **Advanced Topics (if time permits)**

- Transformers
- Diffusion Networks
- Fairness
- Interpretability of Models

Note that this is a tentative list and we may add or remove some topics as it fits.

Prerequisites

Minimum grade of C- in CMSC320, CMSC330, and CMSC351; and 1 course with a minimum grade of C- from (MATH240, MATH461); and permission of CMNS-Computer Science department.

CMSC 422 is a mathematical course. Linear algebra and probability background are required. You must be able to take derivatives by hand (preferably of multivariate functions). You must know what the chain rule of probability is, and Bayes' rule. More background is not necessary but is helpful: for instance, dot products and their relationship to projections onto subspaces, and what a Gaussian is. We provide some reading material to help you refresh your memory, but if you haven't at least seen these things before, you will need to invest a significant amount of time to catch up on math background.

We will make extensive use of the Python programming language. It is assumed that you know or will quickly learn how the program in Python. You should understand basic computer science concepts (like recursion), basic data structures (trees, graphs), and basic algorithms (search, sorting, etc.).

Course Discussions

Ask and answer questions, participate in discussions and surveys, contact the instructors, and everything else on Piazza.

Course Requirements and grading

- Assignments / Homeworks
  - Assignments will provide you an opportunity to think about the material presented in lectures. You are allowed and encouraged to work in groups but you must prepare and submit your solutions independently.
  - Assignments should be submitted through Canvas. Late assignments are not allowed.
- Exams
  - We will have one midterm and one final exam. Exams will cover material in the previous lectures and will be closed book. However, you're allowed to use one sheet of paper (double sided) of size A4 or 8.5×11 in, either printed or handwritten.
- Grading
  - Home works / Quizzes 20%
  - Programming projects 30%
  - Midterm exam 25% (Friday, June 14 from 9:30-10:45 AM in IRB 0324)
  - Final exam 25% (Friday, July 5 from 9:30-10:45 AM in IRB 0324)

Online Posting of Homework Solutions Not Allowed

- Do not post your homework solutions online (e.g., GitHub, PasteBin) where they can be seen by others. Making your solutions accessible to others can lead to academic integrity violations.
- Even if the course is over, do not make your homework solutions available to others.
- Notice we constantly monitor online sources.

Academic Honesty

Note that academic dishonesty includes not only cheating, fabrication, and plagiarism, but also includes helping other students commit acts of academic dishonesty by allowing them to obtain copies of your work. You are allowed to use the Web for reference purposes, but you may not copy straight from any website or any other source. In short, all submitted work must be your own.

Cases of academic dishonesty will be pursued to the fullest extent possible as stipulated by the Office of Student Conduct. Without exception every case of suspected academic dishonesty will be referred to the Office. If the student is found to be responsible of academic dishonesty, the typical sanction results in a special grade "XF", indicating that the course was failed due to academic dishonesty. More serious instances can result in expulsion from the university. If you have any doubt as to whether an act of yours might constitute academic dishonesty, please contact your TA or the course coordinator.

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit https://www.studentconduct.umd.edu/.

Examples of Academic Integrity Violations

The following are examples of academic integrity violations:

- Hardcoding of results in a project assignment. Hardcoding refers to attempting to make a solution appear as if it works correctly (e.g., printing expected results).
- Using any material available on the internet/web or any other source.
- Hiring any online service to complete an assignment for you.
- **You may not post the implementation of your assignments, materials related to the class (e.g., project description), or any other material associated with this course. Even if the class is over and you have graduated, you may NOT post any material.**
- Sharing your homework solutions with any student.
- Providing ideas/suggestions on how to solve/implement an assignment.
- Using online forums to ask for help regarding our assignments.

AI Tools

The use of AI tools like Chat GPT and Llama etc. are permitted in this class, however, you are expected to adhere to the following guidelines to avoid any academic integrity violations:

- Cite any line(s) of code that you use directly from an AI tool.
- Do no use the output of a piece of code directly from an AI tool, you are expected to run the line(s) of code yourself to generate an output and report that.
- You are still expected to explain the output of any code and comment it to explain what the line(s) of code are doing whether written by you or generated by an AI tool.
- Also report whether you are using line(s) of code verbatim from an AI tool or you have modified the code, clearly identifying the modifications.
- Do not post the contents of this class including assignments (prose or code) verbatim into any AI tool or any external website.

Class Announcements

You are responsible for reading the class announcements that are posted on ELMS and Piazza. Please check them often (at least once a day). Important information about the course (e.g., deadlines, assignment updates, etc.) will be posted on ELMS and you're responsible for keeping track of them and meeting them. Late assignments are not accepted.

Excused Absence and Academic Accommodations

1. Any student who needs to be excused for an absence from a single class session , due to a medically necessitated absence shall:
   - **Make a reasonable attempt to inform the instructor of his/her illness prior to the class**. If you are going to miss an in-class assignment then we expect to hear from you (either email or telephone message) before the class session begins.
   - Upon returning to the class, present their instructor with a self-signed note attesting to the date of their illness. The note must contain an acknowledgment by the student that the information provided is true and correct. Providing false information to University officials is prohibited under Part 9(h) of the Code of Student Conduct (V-1.00(B) University of Maryland Code of Student Conduct) and may result in disciplinary action.
   - **This self-documentation may not be used for the Major Scheduled Grading Events as defined below and it may only be used for one class meeting during the semester.**
2. Any student who needs to be excused for more than one absence, or for a "Major Scheduled Grading Event", must provide written documentation of the illness from the Health Center or from an outside health care provider. This documentation must verify dates of treatment and indicate the timeframe that the student was unable to meet academic responsibilities. The documentation should be given to the instructor, not the TA. **We will not accept a "self-signed" note for "major scheduled grading events", as defined below. The note must be signed by a health care professional.**

   The Major Scheduled Grading Events for this course include:
   - Midterm
   - Final Exam
   - Homeworks / Aassignments

It is also the student's responsibility to inform the instructor of any intended absences from exams for religious observances **in advance**. Notice should be provided as soon as possible but no later than one week prior to the exam.

Accessibility and Disability Support

Any student eligible for and requesting reasonable academic accommodations due to a disability is requested to provide, to the instructor in office hours, a letter of accommodation from the Office of Disability Support Services (DSS) within the first two weeks of the semester.

Course Evaluations

The Department of Computer Science takes the student course evaluations very seriously. Evaluations will usually be open during the last few weeks of the course. Students can go to https://courseevalum.umd.edu/ to complete their evaluations.

Copyright

Acknowledgements

Thanks to Dr. Soheil Feizi for sharing his course materials.