Computer Science 688: Machine Learning

George Mason University

Fall 2023

Instructor: Sanmay Das
email: sanmay at gmu dot edu
Class times: Wednesdays 7:20-10:00, Art and Design Building 2003
Office: ENGR 4422
Office hours: Mondays from 3-4 PM and by appointment.
Textbook: *Patterns, Predictions, and Actions: Foundations of Machine Learning* by Moritz Hardt and Benjamin Recht.
Communication and Class Link: Piazza: `http://piazza.com/gmu/fall2023/cs688`
Teaching Assistant: Gaurab Pokharel (gpokhare at gmu dot edu)
Office hours: TBA
Additional books that may be useful:
Learning from Data by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin
Pattern Recognition and Machine Learning by Christopher M. Bishop

1 Course Description

This course covers the theory and principles underlying different machine learning paradigms. The emphasis is on statistical theory and methodology. Topics include: Model selection and generalization; Overfitting and under fitting; Bayesian theory and Decision theory; Maximum Likelihood estimation, MAP; Regularization; Bias-variance tradeoff; Curse of dimensionality; Dimensionality reduction; Linear Models for classification; Probabilistic Generative Models; Probabilistic Discriminative Models; Neural Networks (Backpropagation); Deep Learning (CNNs); Kernel methods; Support Vector Machines; Ensemble Methods; Unsupervised Learning (Clustering, EM, Mixture Modeling); Reinforcement learning.

1.1 Overview

Machine learning studies computer algorithms for learning to do things. For example, we might be interested in learning to complete a task, or to make accurate predictions, or to navigate in an unexplored environment. The learning that is being done is always based on some sort of observations or data, such as examples (the most common case in this course), direct experience, or instruction. So in general, machine learning is about learning to do better in the future based on what was experienced in the past. The emphasis of machine learning is on automatic methods. In other words, the goal is to devise learning algorithms that do the learning automatically without human intervention or assistance.

The machine learning paradigm can be viewed as "programming by example." Often we have a specific task in mind, such as recognizing handwritten digits on an envelope to perform

automated mail dispatching. But rather than program the computer with rules to solve the task directly, in machine learning, we seek methods by which the computer will come up with its own program based on examples that we provide. The course covers key algorithms and theory at the core of machine learning.

1.2 Prerequisites

Formally, you must have received a grade of B- or better (or XS) in either CS 580 or CS 584. Programming experience in at least one modern programming language will be needed. Students should be comfortable with probability and statistics, calculus, and linear algebra. This is a conceptual, mathematical, and theoretical class. You will need to write proofs and engage with the mathematical foundations of ML. The emphasis is **not** on programming projects, although there will be some.

1.3 Format

Class sessions will be lectures, but they may also involve in-class activities, quizzes, and two tests. Quizzes will not be announced in advance, there may be anywhere from 3-5 of them, and they will all be taken online, so please bring a computer to class. In addition to the textbook, other material may also be discussed, in which case pointers to appropriate reading will be provided. Grading will be based on homework assignments, the quizzes, the tests, and class participation.

2 Policies

2.1 Assessment and Course Grade

Your overall course score will be determined (on a curve) using the following weights. There is **no absolute correspondence of scores to grades. Do not assume you can calculate your total score and then map that to your grade.**

1. Homework assignments: 50%
2. Quizzes: 15%
3. Tests: 30% (September 27 and November 15)
4. Class participation: 5%

Homeworks will be submitted on Gradescope – we will make instructions available. Late assignments will not be accepted.

2.2 Late Days

Assignments will typically be due **by 10PM on the due date**. Each student will be given a budget of five *late days* that they can use to turn in homeworks late. A late-day can be used without explanation to extend a homework submission deadline by 24 hours, but **no more than two late-days can be used on any one homework**.

Any part of a late day that you use counts as a full late day. For example, if you do not submit your homework until 11PM if it is due at 10PM, that counts as a full late day. If you submit 26 hours late, you will have used two late days. You are responsible for keeping track of your

usage of late days. **Use your late days wisely, if at all.** This late-day policy is intended to cover unanticipated things like minor sickness, exams in other classes, etc. so that you do not have to ask for extensions. Once you have used up your budget of late days you will not be allowed to turn in homeworks late for any reason other than true medical or family emergencies.

2.3 Make-Ups and Incompletes

We will not provide make-up quizzes or tests or incomplete grades. **Excused** absences (excused by the instructor prior to the class meeting) will result in your grade being "filled in" based on performance on the ones you are present for.

2.4 Academic Integrity and GMU Honor Code

Please make sure you are cognizant of the GMU Honor Code: `https://oai.gmu.edu/mason-honor-code/full-honor-code-document/` and the CS Department Honor Code: `https://cs.gmu.edu/resources/honor-code/`. You should follow the GMU honor code and the CS department honor code, with the following caveats and modifications.

Collaboration with other students: Collaboration in thinking through problems can be highly beneficial, but it is essential to come to a solid understanding of the material yourself. In this class, you are allowed to collaborate on assignments to the following extent. You are welcome to discuss problems with each other and to take your own notes during these discussions. However, you must write up solutions on your own. You may not share or look at any written material (code, answers to problems) that will be part of your or another student's submission, or that has been used in a prior edition of this or a related class. You must write, on the assignment, the names of students you discussed each problem with, and any external sources you used in a significant manner in solving the problem. Lack of citation of a source is a serious violation of this policy. You may not give or receive help from other students in the class on quizzes or tests.

Use of generative AI models: Generative AI models (including, but not exclusive to Bard, ChatGPT, or Claude) may be used in this course as an assistant in homework assignments.

Any use must follow the fundamental principles of the Honor Code and include the following statement with assignment submission: **The ideas in this submission are original and were generated by (my name). `<Generative-AI model (specify)>` was used as an editorial/coding assistant, however, I take full responsibility for the originality and accuracy of the content.**

Here are a few warnings: Large Language Models (LLMs) **hallucinate** and they do so frequently, while maintaining a tone of confidence and authority in the language they generate that would usually only be used by a human who is very confident in their answer. Recent studies have found that in technical domains they are typically wrong *at least 40% of the time*. LLMs are not human and you should be careful not to anthropomorphize them.

More generally, sharing your own original ideas with generative AI models can lead to loss of control and ownership of those ideas and coding. It can also derail your learning objectives by sacrificing the opportunity to acquire the knowledge, skills, and critical thinking taught in this course. If you rely on them, you risk being unable to perform to expectations in quizzes, tests, and situations where they are not available. Utlimately, this could endanger your employability. This course is an opportunity to learn. Treat it as such.

Consultation: If you have any questions about the level of collaboration permitted, or any other aspect of this policy, please speak with the instructor or TA about it before handing in the assignment! Any deviation from this policy will be considered a violation of the GMU Honor Code.

2.5 Accommodations and resources for disabilities

If you have a documented learning disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Office of Disability Services (SUB I, Rm. 222; 993-2474; `http://www.gmu.edu/student/drc` to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

2.6 Campus Closure or Emergency Class Cancelation/Adjustment Policy

If the campus closes, or if a class meeting needs to be canceled or adjusted due to weather or other concern, students should check Piazza for updates on how to continue learning and for information about any changes to events or assignments.

2.7 Non-Discrimination and Inclusiveness

GMU, the CS Department, and the CS Faculty are committed to providing an educational environment free from any discrimination on the basis of race, color, religion, national origin, sex, disability, veteran status, sexual orientation, gender identity, gender expression, age, marital status, pregnancy status, or genetic information. If you feel there has been a violation of the University's policies on discrimination, please contact GMU's Office of Diversity, Equity, and Inclusion (`https://diversity.gmu.edu`). I strive to use people's preferred names and pronouns.

2.8 Sexual Harassment, Sexual Misconduct, and Interpersonal Violence

It is important to remember that all faculty members are designated "Responsible Employees", and are **required to report all disclosures** of sexual assault, interpersonal violence, and stalking to Mason's Title IX Coordinator per university policy 1412. I will not be able to keep such disclosures confidential, by law. If you wish to speak with someone confidentially, please contact the Student Support and Advocacy Center (`https://ssac.gmu.edu`) or Counseling and Psychological Services (`https://caps.gmu.edu`). You may also seek assistance from Mason's Title IX Coordinator directly (`https://diversity.gmu.edu/title-ix/who-can-i-call`).