CS480/680: Introduction to Machine Learning

Spring 2021

Web: http://www.gautamkamath.com/courses/CS480-sp2021.html

Piazza: http://www.piazza.com/uwaterloo.ca/spring2021/cs480680/home

Syllabus

Instructor

Gautam Kamath
Assistant Professor
School of Computer Science
University of Waterloo

Web: www.gautamkamath.com
Email: gckamath@uwaterloo.ca
Office hours: TBD, or by email appointment.

Course Description

This course introduces students to the design of algorithms that enable machines to "learn," as well as some of their significant impacts on society. In contrast to the classic paradigm where machines are programmed by specifying a set of instructions that dictate what exactly a machine should do, a new paradigm is developed whereby machines are presented with examples from which they learn what to do. This is especially useful in complex tasks such as natural language processing, information retrieval, data mining, computer vision and robotics where it is not practical for a programmer to enumerate all possible situations in order to specify suitable instructions for all situations. Instead, a machine is fed with a large collection of examples from which it automatically learns suitable rules to follow.

The course will introduce the basics of machine learning and data analysis, and demonstrate how machine learning techniques have impacted life and society. The course starts with some classical tasks and algorithms in machine learning, then transitions to modern advances and successful applications in real life, and finally ends with discussions on the societal impact that machine learning has generated. As machine learning becomes increasingly popular in practice and starts to penetrate into critical areas such as law, health care, scarce resource allocation, autonomous driving, elections, etc., various ethical and safety issues have been discovered. A significant part of the course will be devoted to raising students' awareness on such ethical issues, including how to make machine learning algorithms fair and unbiased, how to protect individual privacy without compromising learning, how machine learning algorithms might be attacked, how to measure and improve the security of machine learning algorithms, and how to make machine learning algorithms explainable and interpretable.

Tentative Course Objectives

At the end of the course, students should have the ability to:

- Recognize and formalize a task as a machine learning problem;
- Identify suitable algorithms to tackle different machine learning problems;
- Understand and implement a plethora of foundational machine learning algorithms;
- Apply machine learning algorithms to real datasets;
- Become aware of potential ethical and safety issues of machine learning on society.

(Tentative) Course Overview

- Introduction
- Perceptron
- Linear Regression
- Optimization Basics
- Statistics Basics
- K-nearest Neighbours
- Logistic Regression
- Support Vector Machines
- Reproducing Kernels
- Decision Trees
- Bagging, Boosting and Random Forest
- Multi-layer Perceptron
- Deep Neural Networks
- Convolutional Neural Networks
- Recurrent Neural Networks
- Graph Neural Networks
- Mixture of Gaussians
- Graphical Models
- Generative Adversarial Networks
- Flow Models
- Adversarial Machine Learning
- Attention
- Learning to Learn
- Explainable Machine Learning
- Causality
- Privacy

Prerequisites

Good knowledge of linear algebra (vector space, eigenvalue, matrix multiplication, etc.) and basic probability (random variable, distribution, expectation, conditional probability, Bayes rule, etc.). Exposure to numerical computing and optimization is a plus but not required. Basic programming language such as Python.

Textbooks

There is no required textbook. We will pose lecture notes or slides before class. You are encouraged to check out the following classical books.

- *Dive into Deep Learning (2019).* Aston Zhang, Zack C. Lipton, Mu Li and Alex J. Smola.
- *Deep Learning (2016).* Ian Goodfellow, Yoshua Bengio and Aaron Courville.
- *Understanding Machine Learning: From Theory to Algorithms (2014).* Shai Shalev-Shwartz and Shai Ben-David.
- *Elements of Statistical Learning (2nd edition, 2009).* Trevor Hastie, Robert Tibshirani and Jerome Friedman.

Grading

There will be 5 homework assignments, each worth 20% of your final grade. We will calculate the top 4 assignments into your grade hence in total 4 x 20% = 80%. Assignments will be posted on the course webpage and announced on course piazza page. Expect to have 1 assignment every other week (roughly). Programming questions will be in Python.

Completed assignments will be submitted through LEARN and/or CrowdMark. Submit early and often!

As usual, it is OK to seek for help, but you must write your solutions independently and individually, and you should always acknowledge any help you get (book, friend, internet, etc.).

Mark appeals should be requested within two weeks of receiving the mark. The appeal could go either ways, so request only if you truly believe something is wrong.

Late Policy

Two 48 hour extensions per student are provided. They may be each be used on one of the five homework assignments (at most one may be used per homework assignment). Email the instructor and the TA for the assignment at least 24 hours before the deadline to let us know that you're using it, and why. There are no extensions for the final project or any other deadline. Beyond this, we do not accept any late homework submissions, unless you have a legitimate reason with formal proof (e.g. hospitalization, family urgency, etc.). Traveling, busy with other stuff, or simply forgetting to submit, are not considered legitimate.

Project (20% of the final grade)

Students are expected to conduct a research project: For CS480 students we are going to organize a Kaggle competition and you will submit and compete your results there.

For CS680 students, your project could be a survey of a subfield of machine learning, or an empirical comparison of several related algorithms on an interesting dataset, or an application of machine learning algorithms to a different field, or designing a novel algorithm to address a need in machine learning, or theoretically analyzing the performance of a machine learning algorithm (new or old). CS680 students are allowed to work on this project in pairs, submitting a single project. The same grade will be assigned to all members of a group. Some possible projects will be suggested as we progress in the course, but you are highly encouraged to choose your own project (that interests you the most).

You project should

- relate to machine learning (obviously)
- allow you to learn something new (and hopefully significant)
- be interesting and nontrivial, preferably publishable in a top machine learning conference

The project proposal will be due on June 14, 2021. Please concisely describe what your project is about, what are the related works, what is your execution plan, what do you expect to learn/contribute, and how are you going to evaluate your results. I expect the proposal to be less than **4 pages** (excluding references). This will be worth 5% of your final grade (25% of the project grade).

An (ungraded) update (to instructor and your assigned TA) is due on July 12, 2021. This should describe what you've done so far, and what is left to be done. You are expected to have measurable "work done" by this point.

The project report will be due August 10, 2021. Please summarize all your findings (empirical, algorithmic, theoretical) in a scientific report. I expect there is an introduction section, a background section, a main result section, and a conclusion section. Depending on your project, you may include an experimental section and/or discussion section. Please always give proper citations to prior work or results. Be precise and concise. I expect the report to be less than **8 pages** (excluding references). This will be worth 15% of your final grade (75% of the project grade).

Your project report will be evaluated by its clarity, significance, rigor, presentation, and completeness.

Academic Integrity

In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. Check the university website for more information.

Grievance

A student who believes that a decision affecting some aspect of hisher university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Student Petitions and Grievances, Section 4. When in doubt please be certain to contact the department's administrative assistant who will provide further assistance.

Discipline

A student is expected to know what constitutes academic integrity to avoid committing an academic offence, and to take responsibility for hisher actions. A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course instructor, academic advisor, or the undergraduate Associate Dean. For information on categories of offences and types of penalties, students should refer to Policy 71, Student Discipline. For typical penalties check Guidelines for the Assessment of Penalties.

Appeals

A decision made or penalty imposed under Policy 70 (Student Petitions and Grievances) (other than a petition) or Policy 71 (Student Discipline) may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to Policy 72 (Student Appeals).

Students with Disabilities

AccessAbility Services collaborates with all academic departments to arrange appropriate accommodations for students with temporary or permanent disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations, please register with the AccessAbility Services at the beginning of each academic term.

Intellectual Property

Students should be aware that this course contains the intellectual property of their instructor, TA, and/or the University of Waterloo. Intellectual property includes items such as:

- Lecture content, spoken and written (and any audio/video recording thereof);
- Lecture handouts, presentations, and other materials prepared for the course (e.g., PowerPoint slides);
- Questions or solution sets from various types of assessments (e.g., assignments, quizzes, tests, final exams); and
- Work protected by copyright (e.g., any work authored by the instructor or TA or used by the instructor or TA with permission of the copyright owner).

Course materials and the intellectual property contained therein, are used to enhance a student's educational experience. However, sharing this intellectual property without the intellectual property owner's permission is a violation of intellectual property rights. For this reason, it is necessary to ask the instructor, TA and/or the University of Waterloo for permission before uploading and sharing the intellectual property of others online (e.g., to an online repository).

Permission from an instructor, TA or the University is also necessary before sharing the intellectual property of others from completed courses with students taking the same/similar courses in subsequent terms/years. In many cases, instructors might be happy to allow distribution of certain materials. However, doing so without expressed permission is considered a violation of intellectual property rights.

Please alert the instructor if you become aware of intellectual property belonging to others (past or present) circulating, either through the student body or online. The intellectual property rights owner deserves to know (and may have already given their consent).