CSC 411/2515

Machine Learning and Data Mining

Course Information

Lectures	Mon 11-1pm (S1), Wed 11-1pm (S2), Thu 4-6pm (S3), Fri 11-1pm (S4)
Instructors	Ethan Fetaya (S1, S2), Emad Andrews (S3), James Lucas (S4)
Office Hours	Tue 11am-12pm, Wed 8am-10am (PT290A), Thu 6-8pm, Fri 9-10:45am (BA3219)
TA Email	csc411-20179-ta@cs.toronto.edu
Tutorials	Mon 3-4pm (S1), Wed 3-4pm (S2), Thu 6-7pm (S3), Fri 3-4pm(S4)
Class Webpage	www.cs.toronto.edu/~jlucas/teaching/csc411

Course Overview

Machine learning aims to give computers the ability to learn from data without being directly programmed. A machine learning system is taught how to behave based on presented examples of correct behaviour or through trial-and-error experience. We train these systems using learning algorithms which define how the systems behaviour should be modified during this process.

Machine learning spans across a range of applications and academic disciplines. Within machine learning we make use of optimization, probability theory and statistics, linear algebra and more. Machine learning has seen applications in computer vision, email filtering, anomaly detection, speech translation, and many other fields.

This course will present a broad overview of machine learning techniques applicable to a variety of domains. We will explore these methods and discuss when each is most appropriate. As part of this process we will also consider the difficulties faced when implementing a machine learning method and how to recognize a suitable task for such techniques.

Prerequisites

You will be expected to be familiar with basic probability theory and statistics (STA247/255/257). For example, you should be familiar with common probability distributions (Gaussian, Poisson, etc.). You should also be familiar with college-level linear algebra and calculus (MAT135/137/157). Experience with programming will be valuable for the assignments (CSC263/265); especially knowledge of Python.

Reading

We will be awarding a small percentage of total marks (5%) for reading classic research papers associated with the topics covered during lectures. This is intended to encourage independent research - an extremely valuable tool in a field which moves as quickly as machine learning. These points will be awarded through an honour system.

External reading will not be necessary to cover all course material. However, we will also suggest reading to supplement lecture material.

Course Requirements and Grading

The grading for this course will be divided up as follows:

Assignments 45%
Mid-Term Exam 20%
Final Exam 30%
Reading 5%

There will be three assignments, each worth 15%.

A minimum of 30% will need to be scored on the final exam to ensure a passing grade for this course.

For graduate students a course project will be graded in lieu of the final exam.

Assignments

There will be three assignments - each worth 15% of the final course grade. The assignments will generally involve a small amount of mathematical derivation coupled with implementations of machine learning algorithms. You will be expected to implement the algorithm and evaluate your implementation on some provided data. You will then provide a summary of your work and analyse your results briefly. These implementations may be completed in any language but Python is recommended.

The assignments must be completed independently – collaboration is not permitted. Each student will be responsible for their own work. Discussion of the assignment should be limited to clarification only. Violation of this policy is grounds for a semester grade of F, in accordance with university regulations.

The assignment schedule is included in the schedule section of this syllabus. Assignments will be due at 10pm on the designated date. Assignments handed in late will incur a 10% penalty per day up to 3 days at which point submission will be blocked. Extensions will be granted only in special situations requiring either a Student Medical Certificate or a written request approved by the instructor at least one week before the due date.

Exams

There will be a closed-book mid-term which will focus on all course material covered up to that point in lectures (unless otherwise specified). In the mid-term, you will not be tested on any of the reading assignments.

The final exam will focus primarily on material covered after the mid-term.

Attendance

Students are expected to attend all classes and tutorials. This is particularly important as the tutorials will cover material which is not presented during lectures.

Communication

The bulk of communication should take place on Piazza – **this will be the easiest way to get the attention of the TAs and instructors.** If necessary any other queries can be directed to the course instructor email list: csc411-20179-instrs@cs.toronto.edu .

The course TAs can be reached at: csc411-20179-ta@cs.toronto.edu .

Schedule

The following is a tentative schedule.

Week of	Topics to be covered
Sept. 7 - Sept. 13	Introduction; Linear regression
Sept. 14 - Sept. 20	Linear classification; Logistic regression
Sept. 21 - Sept. 27	Nearest neighbor; Decision trees
Sept. 28 - Oct. 4	Multi-class classification; Probabilistic Classifiers I
Oct. 5 - Oct. 11	Probabilistic Classifiers II; Neural Networks I
Oct. 12 - Oct. 18	Neural Networks II;PCA
Oct. 19 - Oct. 25	t-SNE; Clustering
Oct. 26 - Nov. 1	Mixture of Gaussians; EM
Nov. 2 - Nov. 8	SVM; Kernels (Nov. 6-10 Reading week)

Nov. 9 - Nov. 15	SVM; Kernels (Nov. 6-10 Reading week)
Nov. 16 - Nov. 22	Ensembles
Nov. 23 - Nov. 29	RL
Nov. 30 - Dec. 7	Learning theory
Dec. 9 - 20	Final Exam Period