This syllabus is a guideline for the course and not a contract. As such, its terms may be altered when doing so is, in the opinion of the instructor(s), in the best interests of the class.

University of Alberta
CMPUT 466/566 Machine Learning
Winter 2024

Course Information

Instructor: Bailey Kacsmar
Office: ATH 3-17
E-mail: kacsmar@ualberta.ca
Instructor Office Hours: 2-3pm Tuesday
Lectures: *Tuesday/Thursday 11:00am-12:20pm, CCIS 1-140.*
TAs: Shivam Garg, Shuai Liu, Alex Ayoub, Dongheng Li, Bryan Pui, and Yin Chan
Teaching Assistant's Office Hours: 5-6pm Wednesdays, CCIS L1-160
Lab Component: CCIS L1-160, 5-6:50pm, Wednesdays, not mandatory/not graded, will be used for TA office hours/assistance, background material, tutorials, Q&A, etc.

Course Communications: Important course information will generally be posted to eclass, but may also be sent to your ualberta.ca email address. It is your responsibility to keep up with all course-related information posted to e-class. For personal matters, such as an illness, please email the instructors directly. We will only reply back to email from your ualberta.ca email address, following privacy rules. For all other course matters, please use the eclass forum.

Textbooks and Readings: There is no required textbook. Additional readings may be assigned; readings marked as mandatory contain required material for the course. You must read these mandatory readings. Below are some supplemental materials you may find useful.

- More in-depth reference: Pattern Recognition and Machine Learning - by C. M. Bishop, Springer 2006.
- Less technical reference: An Introduction to Statistical Learning: with Applications in R - by James et al.
- In-depth reference, covering a broader range of topics and with good exercises (free online): Bayesian Reasoning and Machine Learning - by Barber
- Theory-oriented reference: The Elements of Statistical Learning - by T. Hastie, R. Tibshirani, and J. Friedman, 2009

Course Description:

Learning is essential for many real-world tasks, including adaptive control, recognition, diagnosis, forecasting, and data-mining. This course covers foundational methods, particularly for supervised learning, introducing many classical regression and classification algorithms. More advanced modelling techniques will also be introduced, including kernels and neural networks. The course will also provide the formal foundations for understanding when learning is possible and practical.

Prerequisites: CMPUT 204 or 275; MATH 125; CMPUT 267 or MATH 214; or consent of the instructor. Credit cannot be obtained for both CMPUT 367 and CMPUT 466.

Course Objectives and Expected Learning Outcomes:

By the end of this course students should be able to:

- Understand how to formalize prediction problems Understand the basics of deriving learning algorithms
- Be able to determine when learning is likely to succeed, when it is cost-effective and when to use which algorithm
- Understand the basics of evaluating algorithms and their theoretical properties

Outline

CMPUT 466 Grading Scheme

- 40% Assignments (four throughout the term, each worth ten percent)
- 20% Midterms (1 weighted 20% of the grade, the other weighted at 0% or 20%, replacing worst two assignment grades, cannot be used to replace both A3 and A4)
- 30% Final exam
- 5% Final mini-project write-up (optional; if not done, the 5% goes to final exam)
- 5% Weekly participation/thought exercises (14 weeks, so 14 total)

CMPUT 566 Grading Scheme

- 85% Weighted value of the above 466 components (final, midterm, assign., participation, with 5% Project proposal replacing mini-project write-up)
- 15% Project write-up
- ———Detailed breakdown with the 85% weighting———
  - 34% Assignments
  - 17% Midterms
  - 25.5% Final Exam
  - 4.25% Project Proposal
  - 4.25% Weekly participation/thought exercises
  - 15% Project write-up

Assignments Each assignment consists of a written portion and a programming portion. The written portion can include proofs, conceptual questions, etc. The programming portion can include implementation, evaluation, experimental execution, etc. Assignments are to be submitted to eClass by the designated deadline. Assignments are to be completed individually. Students can expect the midterm and final examination to include questions that reflect the ones seen on the assignments. All assignments are due on their designated date in the calendar by 4pm Edmonton time.

Participation/Thought Exercises Each week there will be a prompt during one of the classes. This prompt will be in the slide deck, so if you miss that class you can still respond to the prompt. The response is to be submitted via eclass in the designated place. In class, these prompts will be discussed among the attending students. If you discuss the prompt with someone in class, indicate who you discussed with in your submission. These will be graded on a trinary ”best effort” metric. If you do not submit before the start of the next class (no extensions on these), you get zero, if you submit on time, something at all relevant to the prompt, you will get one, and if you submit something good, definitively engaging with the prompt you get two.

Midterms The midterms question distributions (can include multiple-choice, true and false, short answer, proofs, etc.). Both midterms are comprehensive (covers all material from the course up to that point). They are closed book, no notes or cheat sheets permitted. No calculators allowed (and they also are not necessary). There are two midterms. Your best performing midterm is your midterm grade (the 20% in the syllabus). The other midterm, if the grade is better than your worst two assignments, can replace your grade for those assignments. Note you cannot replace both A3 and A4 this way. This replacement option is because the assignments may tend towards being difficult, as they are intended to help you learn the material such that you could do well on the examinations.

Final Mini Project Write-up (466 ONLY) Any students wishing to do the project in 466 **MUST** submit a notification by the proposal deadline indicated in the course schedule (January 30, 2024) by 4pm Edmonton time. *For the 466 proposal,* students must send via eclass the topic/title and a short description of what they plan to do. The following are the basic requirements for the project.

- Formulate a task into a machine learning problem. Cannot be a task from any of the assignments or the datasets used for the assignments.
- Implementation of the task. Include appropriate train-test-validation (as we will learn about during the course)
- Comparison. Include baselines (where appropriate/possible). Compare three algorithms. For instance, three that are reasonable/applicable for the task, but employ a different approach to solving the task. E.g., using a different configuration of hyper-parameters is not a different approach.
- It is permissible to use machine learning packages for the project. However, the package/code base cannot be for the selected task (e.g., cannot just use a library where you can just run it and it solves the task via ab ”sh run.sh” etc.

The following are the requirements for the final submission. Not fulfilling these requirements will result in deductions.

- Submission must include a pdf write up and the code to reproduce the reported results. The code can be a zip file or a public repository (it should not require a login or an invitation to access it).
- There must be a README in the code submission that briefly explains the repository and how to use it to replicate the results. It is recommended that you run your code once on a fresh machine setup to ensure you have correctly documented your repository and that the version you submit will run and produce your results.
- Write-up must include:
  - An introduction, describing the task, background, and motivation
  - Problem setup/formulation, datasets, samples, configuration, etc
  - Methodology, how you approached the problem and what baselines you measured and compared to
  - Evaluation, how did you evaluate success, why was it a good metric to use, why is it reasonable?
  - Results, what is the output, what were the differences, how do you interpret the results, what are the implications?

  - Note: Code snippets, program outputs, etc. are not part of your write-up. Figures, tables, algorithms can me contained int he write-up

- Graded components/formatting overview

  - There is no minimum page length. In general, a few pages should be sufficient for your report. More than five pages double column (e.g., ACM SIG format) is likely to be an over-length report and not only will overly long reports not yield a higher grade it will likely be indicative of poor organization and presentation of the material; resulting in a lower grade. The report must be written in English organized with the results and relevant tables and figures.

Final Project (566 ONLY) Students registered in CMPUT 566 must do a project and project write-up. The *566 proposal* must be submitted by the deadline January 30, 4pm Edmonton time. Proposals for 566 must include topic/title and problem description. Your proposal should be one page in length and include at least 5 references. You should include two possible datasets and a (some) possible approach(s) you aim to take to address the problem you have formulated. At this stage you will not have done any evaluation or started the project (no experiments nor results).

Project write-up Your paper should be a summary of past and current work on your topic, as well as an overview of known open problems and potential future directions in the area. You should provide a concise summary of work, emphasizing major accomplishments, rather than a detailed accounting of individual pieces of project/implementation activity. Your literature review/related work portion will be one third of the evaluation of your report. One third will be on project formulation and presentation. One third will be on evaluation and results interpretation.

Final Exam The final exam will be similar to the midterms in terms of question distributions (can include multiple-choice, true and false, short answer, proofs, etc.). The final exam is comprehensive (covers all material from the course). It is closed book, no notes or cheat sheets permitted. No calculators allowed (and they also are not necessary).

EVALUATION IMPORTANT - Subject to Random Test All graded components of the course are subject to an oral test at the discretion of the instructor. A random subset of submissions (and any selection of ones based on the instructor/TAs determination) will be evaluated this way. The performance of the student on the oral test will effect the students grade on that submission. If the student is reasonably able to explain their submission, the oral test will *not* impact their original grade. If the student cannot reasonably explain their submission (are not sufficiently familiar with it), the oral test mark will result in a decrease in their original grade. While students may be referring to internet materials and solutions or discussing assignments with friends; they must do the assignment by themselves and acknowledge any sources used. Further, they must be able to defend/explain their submissions (e.g., how they arrived at the solution, why that solution works etc.). Be aware that any suspected cheating will be reported to the Faculty of Science.

Week	Date	Topic	Assign.	566 Proj.	Exams
1	Jan 9	Introduction			
1	Jan 11	Math primer			
2	Jan 16	Decision Trees			
2	Jan 18	Eval. and performance	A1 out		
3	Jan 23	MLE and optimization			
3	Jan 25	Naive Bayes			
4	Jan 30	Perceptron		Proposals due	
4	Feb 1	Logistic Regression	A1 due		
5	Feb 6	Neural Networks			
5	Feb 8	Neural Networks	A2 out		
6	Feb 13	Midterm 1			up to Feb 8 (inclusive)
6	Feb 15	CNNs			
7	Feb 20	No class			
7	Feb 22	READING WEEK			
8	Feb 27	Adv. attacks, GANs			
8	Feb 29	Interpretability	A2 due		
9	Mar 5	Ethical Implications	A3 out		
9	Mar 7	SVMs			
10	Mar 12	SVMs/Unsupervised Learning			
10	Mar 14	Unsupervised Learning			
11	Mar 19	EM	A3 due		
11	Mar 21	Midterm 2	A4 out		up to Mar 19 (inclusive)
12	Mar 26	Language Models			
12	Mar 28	Language models			
13	Apr 2	Reinforcement Learning			
13	Apr 4	Topics in ML	A4 due		
14	Apr 9	Topics in ML			
14	Apr 11	Review Class - bring q's		Reports due	
Final Exams are scheduled and announced by registrar					

Course Policy Information

Missed or Late Assessments: Please start working on the material in advance of the deadlines. To motivate you to do so, we may require you to submit milestones for some or all of them. Late submissions for assignments and project reports will be accepted only up to 72 hours after the original due date. They will not be accepted at all after this 72 hours has ended. All other graded components (participation/thought exercises and proposals) must be done on time. There is no penalty for accepted late submissions within the 72 hour window. Course personnel will not normally give assistance after the original due dates.

Remarking Policy: If you have an assignment that you would like to have reappraised, please email the instructor to submit your request. Include a clear justification for your claims. Note that reevaluation can result in the grade going up or down. Not providing a clear justification is reason for the instructor to decline the reappraisal. The appeals deadline is **one week** after the respective graded item is first made available. Appeals will not be considered after this date; so be sure to review your assessments in advance of this. If your appeal is concerned with a simple calculation error, please email me or speak with me during my office hours.

University Policy Information

Academic Integrity and Student Conduct:

The University of Alberta is committed to the highest standards of academic integrity and honesty, as well as maintaining a learning environment that fosters the safety, security, and the inherent dignity of each member of the community, ensuring students conduct themselves accordingly. Students are expected to be familiar with the standards of academic honesty and appropriate student conduct, and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behaviour and the Student Conduct Policy, and avoid any behaviour that could potentially result in suspicions of academic misconduct (e.g., cheating, plagiarism, misrepresentation of facts) and non-academic misconduct (e.g., discrimination, harassment, physical assault). Academic and non-academic misconduct are taken very seriously and can result in suspension or expulsion from the University.

All students are expected to consult the Academic Integrity website for clarification on the various academic offences. All forms of academic dishonesty are unacceptable at the University. Any suspected academic offence in this course will be reported to the College of Natural and Applied Sciences. Suspected cases of non-academic misconduct will be reported to the Dean of Students. The College, the Faculty of Science, and the Dean of Students are committed to student rights and responsibilities, and adhere to due process and administrative fairness, as outlined in the Code of Student Behaviour and the Student Conduct Policy. Anyone who is found in violation is likely to receive a sanction. Typical sanctions for academic misconduct include conduct probation, a mark reduction or a mark of 0 on an assessment, a grade reduction or a grade of F in a course, a remark on the transcript, and a recommendation for suspension or expulsion. Sanctions for non-academic misconduct include conduct conditions, fines, suspension of essential or non-essential University services and resources, and suspension or expulsion from the University.

Appropriate Collaboration:

Students are not permitted to copy solutions on homework assignments. Here are some tips to avoid copying on assignments:

1. Do not write down something that you cannot explain to your instructor.

2. When you are helping other students, avoid showing them your work directly. Instead, explain your solution verbally. Students whose work is copied also receive academic sanctions.

3. If you find yourself reading another student's solution, do not write anything down. Once you understand how to solve the problem, remove the other person's work from your sight and then write up the solution to the question yourself. Looking back and forth between someone else's paper and your own paper is almost certainly copying and will result in academic sanctions for both you and your fellow student.

4. If the instructor or TA writes down part of a solution in order to help explain it to you or the class, you cannot copy it and hand it in for credit. Treat it the same way you would treat another student's work with respect to copying, that is, remove the explanation from your sight and then write up the solution yourself.

5. There is often more than one way to solve a problem. Choose the method that makes the most sense to you rather than the method that other students happen to use. If none of the ideas in your solution are your own, there is a good chance it will be flagged as copying.

Students Eligible for Accessibility-Related Accommodations:

In accordance with the University of Alberta's Discrimination, Harassment, and Duty to Accommodate policy, accommodation support is available to eligible students who encounter limitations or restrictions to their ability to perform the daily activities necessary to pursue studies at a post-secondary level due to medical conditions and/or non-medical protected grounds. Accommodations are coordinated through the Academic Success Centre, and students can learn more about eligibility on the Register for Accommodations website.

It is recommended that students apply as early as possible in order to ensure sufficient time to complete accommodation registration and coordination. Students are advised to review and adhere to published deadlines for accommodation approval and for specific accommodation requests (e.g., exam registration submission deadlines). Students who request accommodations less than a month in advance of the academic term for which they require accommodations may experience unavoidable delays or consequences in their academic programs, and may need to consider alternative academic schedules.

Academic Success Center: The Academic Success Centre (ASC) provides services to support University of Alberta students in the areas of accommodations, learning, and writing. The ASC coordinates reasonable accommodations to eligible students who encounter medical or non-medical restrictions to their ability to perform the daily activities necessary to pursue studies at a post-secondary level. To that end, they work with students to coordinate disability-related accommodation needs for participation in University programs. For more information, and to register for services, visit the Academic Accommodations webpage.

The ASC also provides peer-based and professional academic support in the areas of learning and writing. They offer individual appointments, group workshops, and online courses to students in all University of Alberta programs, and at all levels of achievement and study.

At Writing Services, undergraduate students can work with a peer tutor to get feedback on a draft of their paper. Graduate students can book an appointment with a graduate writing advisor to get feedback on their documents. For more information, please visit the Writing Services webpage.

Recording and/or Distribution of Course Materials:

Audio or video recording, digital or otherwise, of lectures, labs, seminars or any other teaching environment by students is allowed only with the prior written consent of the instructor or as a part of an approved accommodation plan. Student or instructor content, digital or otherwise, created and/or used within the context of the course is to be used solely for personal study, and is not to be used or distributed for any other purpose without prior written consent from the content author(s).