University of Alberta

CMPUT 655: Reinforcement Learning 1
LEC A1
ETLC E2-001 2023

Instructor: Marlos C. Machado
TAs: Anna Hakhverdyan, David Szepesvari, Bryan Chan, and Gábor Mihucz
Office: ATH 3-08
E-mail: machado@ualberta.ca
Web Page: https://eclass.srv.ualberta.ca/course/view.php?id=90138

Office hours: Marlos: Thursday 15:00 - 16:45 in ATH 3-08 (Athabasca Hall)
Anna: Monday 11:00 - 13:00 in CSC 3-50
David: Tuesday 13:00 - 15:00 in CSC 3-50
Bryan: Wednesday 14:00 - 16:00 in CAB 3-13
Gabor: Wednesday 9:15 - 11:15 in CAB 3-13
Slack and eClass: asynchronously

TA email address: cmput655@ualberta.ca
Do not personally email the TAs. They will only respond via cmput655@ualberta.ca.

Lecture room & time: ETLC E2-001, Friday 14:00 - 16:50
Attendance isn't mandatory although strongly encouraged.

Slack invitation link:
https://join.slack.com/t/cmput655fall2023/shared_invite/zt-225uyp34m-CpFH7QECLTEeaXe8kmK0gw

---

COURSE CONTENT

Course Description: This course provides an overview of reinforcement learning, which focuses on the study and design of agents that interact with a complex, uncertain world to achieve a goal. We will emphasize agents that can make near-optimal decisions in a timely manner with incomplete information and limited computational resources. The course will cover Markov decision processes, reinforcement learning, value-based methods, policy-gradient methods, planning, function approximation (online supervised learning), and contemporary research topics in the field.

Parts of the course will take place as a flipped classroom, which, as Wikipedia says, is a "learner-centered model where students are exposed to the subject matter outside the classroom so that they can have more meaningful learning opportunities in the classroom". Parts of the course will use the MOOC on Reinforcement Learning, created by faculty members in the department of computing science of the University of Alberta. The initial lecture material and assignments will come from the MOOC. In-class time will be largely spent on discussion, and on the presentation of more advanced topics. The idea is that we want to be able to cover advanced reinforcement learning topics without assuming students have prior knowledge about reinforcement learning. Thus, we will use the MOOC as an introduction to the basic concepts and the classroom for the presentation of more advanced concepts. This model allows for more content to be properly covered and for a more active learning instead of having students passively watching lectures in the classroom.

Course Prerequisites: The course will use Python 3. We will use elementary ideas of probability, calculus, and linear algebra, such as expectations of random variables, conditional expectations, partial derivatives, vectors and matrices. Students should either be familiar with these topics or be ready to pick them up quickly as needed by consulting outside resources.

Course Objectives and Expected Learning Outcomes: By the end of the course, you will have a solid grasp of the main ideas in reinforcement learning, which is the primary approach to statistical decision-making. Any student who understands the material in this course will understand the foundations of much of modern probabilistic artificial intelligence (AI), they will have a broad perspective over the field, and they will be prepared to take other courses revolving around reinforcement learning (for example, CMPUT 609: Reinforcement Learning II, and CMPUT 653: Real-Time Policy Learning), or to apply AI tools and ideas to real-world problems. That person will be able to apply these tools and ideas in novel situations -- e.g., to determine whether the methods apply to this situation, and if so, which will work most effectively. They also will be able to formulate and empirically falsify hypotheses about the behavior of reinforcement learning algorithms. Finally, they will be able to assess claims made by others, with respect to both software products and general frameworks, and also be able to appreciate some new research results.

With a focus on AI as the design of agents learning from experience to predict and control their environment, topics will include:

- Markov decision processes,
- planning by approximate dynamic programming,
- Monte Carlo and temporal difference learning for prediction and control,
- multi-step temporal difference learning,
- Dyna and planning with a learned model,
- prediction and control with linear and nonlinear function approximation, and
- policy gradient methods.

LEARNING RESOURCES

Required Textbook and/or Other Major Course Materials: All course reading material will be available online (textbook and scientific articles). We will also be using videos from the Reinforcement Learning Specialization on Coursera, which was created by faculty members in the department of computing science of the University of Alberta.

We will be using the following textbook extensively:
Sutton and Barto, Reinforcement Learning: An Introduction, MIT Press. 2nd Edition.

The book is available from the bookstore or online as a pdf in the link below:
http://www.incompleteideas.net/book/the-book-2nd.html

On-Line Homework Disclaimer: On-line homework is a component of this course and is provided by a third-party company, Coursera, through a partnership with the University of Alberta. University of Alberta students can get free access to the online content of the Reinforcement Learning Specialisation via the steps outlined here. You are required to set up a free coursera account with your ualberta.ca email address. You will be automatically enrolled in a private version of the course for this class.

Academic Success Centre: The Academic Success Centre provides professional academic support to help students strengthen their academic skills and achieve their academic goals. Individual advising, appointments, and group workshops are available year round in the areas of Accessibility, Communication, Learning, and Writing Resources. Modest fees apply for some services.

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the class on the topic(s) of the week at *11:59:00* (see Course schedule, at the end, for details)
Assessments (graded quizzes / notebooks on Coursera)	9 x 2.3$\overline{6}$% = 21%	Day of the class on the topic(s) of the week at *11:59:00* (see Course schedule, at the end, for details)

Project proposal	15 %	October 25, 2023 *23:59:00*
Midterm exam	20%	November 10, 2023
Final project	35%	December 15, 2023 *23:59:00*

Grades are unofficial until approved by the Department and/or Faculty offering the course.

Statement of Expectations for AI Use: The primary goal of this course is to foster *individual* critical, creative thinking, and problem-solving skills related to reinforcement learning and, more broadly, machine learning. Thus, in order to achieve such learning outcomes, students can submit each practice quiz and graded assignment multiple times, which allows for many learning opportunities. Therefore, the use of advanced AI-tools based on large-language models such as ChatGPT or Bard is strictly prohibited for all quizzes and graded assignments. The only exception is their use for Python-related queries (but the use of such tools to help with the programming assignments themselves is still strictly prohibited). As stated in the university's AI-Squared - Artificial Intelligence and Academic Integrity webpage, "learning is not only about the product; learning is also about the process of acquiring new knowledge or learning ways to think and reason."

Students are also allowed to use advanced AI-tools such as ChatGPT or Bard to proofread their manuscripts, but only after having written a first complete draft of the text to be proofread. Organizing ideas in writing is an essential part of the research process, and shortcuting this process will likely hinder a student's development. One is prohibited from using advanced AI-tools for help with related work. All interactions with an advanced AI-tool are to be submitted as Appendix in the project proposal and final project manuscript. The Appendix does not count toward the pre-specified page limit.

IMPORTANT: Failure to abide by this guideline may be considered an act of cheating and a violation as outlined in the relevant sections of University of Alberta Code of Student Behaviour.

Re-evaluation of the midterm exam: Students will have access to their midterm exam during an exam viewing period. A student who has concerns about how specific questions of their midterm exam were marked can submit a request to the instructor, via email, within two weeks the date they received their marked exam. The request should specify (1) which question is to be re-evaluated, (2) the rationale for such a request, and (3) the proposed marks. Importantly, once a request for re-evaluation is submitted, it is up to the instructor's discretion to adjust the marks. *Students won't be allowed to take their midterm exams with them, nor to take pictures of them*, so in case of concerns the student is advised to take notes during the exam viewing period. The TAs are not authorized to weigh in on the midterm exams, this is something only the instructor can do. *Notice marks can also go down once a question is re-evaluated.*

Past or Representative Evaluative Material: The quizzes and graded assignments are an excellent example of representative material, as well as the exercises in the course's textbook. Additional representative material will be discussed in the classroom and later posted on eClass.

Format of Exams: The midterm exam will be held in a room still to be defined, from 15:30 - 16:50 (1 hour and 20 minutes). *The exams will be closed book and written in real-time.*

Format of Quizzes and Assignments: You are allowed to consult the textbook and materials on eClass while solving the assignments. You are also allowed to discuss the assignments with your classmates. Note, however, that you are not allowed to exchange any written text, code, or to give and/or receive detailed step-by-step instructions on how to solve the proposed problems. *Moreover, as stated above, you are not allowed to use generative AI in your coursework.*

Grade Evaluation: At the end of the term your percentage grade will be converted into a letter grade by following the table below, extracted from the Department Course Policies and the University's Grading System Explained.

Letter	Descriptor	Interpretation
A+ A A-	Excellent	Consistently original thinking that extends the material, demonstrated depth and breadth in the material, ability to integrate material with other subjects, ability to analyze and synthesize material at various levels of abstraction.
B+ B	Good	Like an A, but not consistent over time, or weak in a specific area.
B- C+	Satisfactory	Understand the core material but not its subtleties, can apply it to simple situations on own and to more complex situations with hints, evidence that the material has changed the way of thinking.
C C- D+ D F	Failure	Little evidence of understanding of even the surface issues, poor analysis and synthesis, inability to apply the material.

Grade adjustment: Students are being allowed to not submit 2 practice quizzes and 2 graded assignments (see below). The course project is responsible for 50% of all the marks and it is to be done during the whole term. G*rades will not be rounded at the end, no extra marks will be given, no exceptions.*

Exam Conduct:

- Your student photo I.D. is required at exams to verify your identity.

- Students will not be allowed to begin an examination after it has been in progress for 30 minutes. Students must remain in the exam room until at least 30 minutes has elapsed.
- All cell phones must be turned off and stored in your bags.

POLICIES FOR LATE AND MISSED WORK

Late Policies: Late submissions will not be accepted but students are allowed to miss some submissions, as outlined below.

Missed practice quiz: There are a total of 11 weekly practice quizzes and students are expected to do all of them. But, due to the fact that issues sometimes arise, students have to complete 9 of the 11 quizzes to get the full 9% participation mark. All the quizzes will be due at 11:59:59 of the day mentioned in the schedule. Each practice quiz has equal weight (1%). *Please note students can miss 2 instances of the practice quiz without penalty.*

Missed graded assignment: There are 11 graded assignments, they are usually python notebooks, but sometimes it is a graded quiz. Students should do all of them but, due to the fact that issues sometimes arise, students' final marks will be the sum of their top 9 assignment submissions. Effectively, students only need to complete 9 of the 11 assignments to get the full 21% marks. All the assignments will be due at 11:59:59 of the day mentioned in the schedule. Each graded assignment has equal weight (3%). *Please note students can miss 2 instances of the practice quiz without penalty.*

Missed midterm exam: A student who cannot complete a midterm exam due to incapacitating illness, severe domestic affliction or other compelling reasons can apply for an excused absence. To apply for an excused absence, you must contact the instructor within two working days of missing the assessment or as soon as possible. If an excused absence is granted, then the student will be allowed to take another midterm. A second midterm, to be applied to all students who missed the first midterm exam, will be scheduled by the instructor later in the term, at a time of the instructor's discretion. An excused absence is a privilege and not a right. There is no guarantee that an absence will be excused. Misrepresentation of facts to gain an excused absence is a serious breach of the Code of Student Behaviour.

Late project proposal: The project proposal totals $max(0, (15-d))\%$ of the marks, where d is the number of days the submission is late. The project (and proposal) are designed such that the students work on them during the whole term. Thus, one should plan to have the proposal submitted with enough time in advance such that disrupting events do not impact the student's ability to complete the project. One can submit as many proposals as they want until the deadline,

with the newer one replacing the older one. One should ensure they always have their best version of the proposal submitted.

Missed final project submission: Late submissions will not be accepted. Similar to a conference deadline, the submission deadline cannot be postponed. The project is designed such that the students work on them during the whole term. Thus, one should plan to have the project submitted with enough time in advance such that disrupting events do not impact the student's ability to complete the project. One can submit as many projects as they want until the deadline, with the newer one replacing the older one. One should ensure they always have their best version of the project submitted.

Missed Term Work or Final Exam Due to Non-medical Protected Grounds (e.g., religious beliefs): When a term assessment or final exam presents a conflict based on non-medical protected grounds, students must apply to the Academic Success Centre for accommodations via their Register for Accommodations website. Students can review their eligibility and choose the application process specific for *Accommodations Based on Non-medical Protected Grounds*.

It is imperative that students review the dates of all course assessments upon receipt of the course syllabus, and apply *as soon as possible* to ensure the timely application of the accommodation. Students who apply later in the term may experience unavoidable delays in the processing of the application, which can affect the accommodation.

REMOTE DELIVERY CONSIDERATIONS

Technology for Remote Learning: To successfully participate in remote learning in this course, it is recommended that students have access to a computer with an internet connection that can support the tools and technologies the University uses to deliver content, engage with instructors, TAs, and fellow students, and facilitate assessment and examinations. Please refer to Technology for Remote Learning - For Students for details. If you encounter difficulty meeting the technology recommendations, please email the Dean of Students Office (dosdean@ualberta.ca) directly to explore options and support.

Please contact the instructor by the add/drop deadline September 18 if you do not have access to the minimum technology recommended. The instructor will make arrangements for accommodating students who contact the instructor before this date. Failure to do so may result in a zero in any assessment that depends on the minimum technology.

Student Resources for Remote Learning: Online learning may be new to you. Check out tips for success and find out more about online learning on the Campus Life page, and specifically on the Academic Skills Online & Remote Delivery Resources page.

STUDENT RESPONSIBILITIES

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: You are also allowed to discuss the quizzes, assignments, and the course project with your classmates (beyond the group doing the project). Note, however, that you are not allowed to exchange any written text, code, or to give and/or receive detailed step-by-step instructions on how to solve the proposed problems. We refer students to the University of Alberta's webpage on How to Avoid Inappropriate Collaboration.

Cell Phones: Cell phones are to be turned off during lectures, labs and seminars.

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.

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).

Learning and Working Environment:
The Faculty of Science is committed to ensuring that all students, faculty and staff are able to work and study in an environment that is safe and free from discrimination, harassment, and violence of any kind. It does not tolerate behaviour that undermines that environment. This includes virtual environments and platforms.

If you are experiencing harassment, discrimination, fraud, theft or any other issue and would like to get confidential advice, please contact any of these campus services:

- Office of Safe Disclosure & Human Rights*: A safe, neutral and confidential space to disclose concerns about how the University of Alberta policies, procedures or ethical standards are being applied. They provide strategic advice and referral on matters such as discrimination, harassment, duty to accommodate and wrong-doings. Disclosures can be made in person or online using the Online Reporting Tool*.
- University of Alberta Protective Services: *Peace officers dedicated to ensuring the safety and security of U of A campuses and community. Staff or students can contact UAPS to make a report if they feel unsafe, threatened, or targeted on campus or by another member of the university community.*
- Office of the Student Ombuds*: A confidential and free service that strives to ensure that university processes related to students operate as fairly as possible. They offer information, advice, and support to students, faculty, and staff as they deal with academic, discipline, interpersonal, and financial issues related to student programs.*
- Office of the Dean of Students*: They can assist students in navigating services to ensure they receive appropriate and timely resources. For students who are unsure of the support they may need, are concerned about how to access services on campus, or feel like they may need interim support while you wait to access a service, the Dean of Students office is here to help.*

Feeling Stressed, Anxious, or Upset? It's normal for us to have different mental health experiences throughout the year. Know that there are people who want to help. You can reach out

to your friends and access a variety of supports available on and off campus at the Need Help Now webpage or by calling the 24-hour Distress Line: 780-482-4357 (HELP).

Student Self-Care Guide: This Self-Care Guide, originally designed by the Faculty of Native Studies, has broader application for use during students' learning. It provides some ideas and strategies to consider that can help navigate emotionally challenging or triggering material.

Policy about course outlines can be found in Course Requirements, Evaluations Procedures and Grading of the University Calendar.

Land Acknowledgement: The University of Alberta respectfully acknowledges that we are situated on Treaty 6 territory, traditional lands of First Nations and Métis people.

To learn more about the significance of this land acknowledgement, please read this useful article and associated links to more information.

Disclaimer: Any typographical errors in this Course Outline are subject to change and will be announced in class.

Copyright: Dr. Marlos Cholodovskis Machado, Department of Computing Science, Faculty of Science, University of Alberta (2023).

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 11:59:59)	Readings
1	Fri, Sep 8	Course overview Discussion about what is reinforcement learning Background review: Probability, statistics, linear algebra, and calculus		Chapter 1: Introduction
2	Fri, Sep 15	Fundamentals of RL: An introduction to sequential decision-making Optimality of UCB	Practice quiz (Sequential decision-making) Program. assignment (Bandits & exploration / exploitation)	Chapter 2: Multi-armed Bandits
3	Fri, Sep 22	Fundamentals of RL: Markov decision processes (MDPs) Fundamentals of RL: Value functions & Bellman equations	Practice quiz (MDPs) Practice quiz (Value functions & Bellman equations) Graded quiz (Value functions & Bellman equations)	Chapter 3: Finite Markov Decision Processes
4	Fri, Sep 29	Fundamentals of RL: Dynamic programming Sample-based learning methods: MC methods for Prediction & Control	Practice quiz (Dynamic programming) Program. Assignment (Optimal policies with dynamic programming) Graded quiz (Off-policy Monte Carlo)	Chapter 4: Dynamic Programming Ross's Chapter 2 Chapter 5: Monte-Carlo Methods
5	Fri, Oct 6	Sample-based learning methods: TD learning for prediction	Practice quiz (Advantages of TD) Program. Assignment (Policy	Chapter 6: Temporal-Difference Learning Chapter 7: n-step Bootstrapping

		Sample-based learning methods: TD learning for control	evaluation with TD learning) Practice quiz (Expected Sarsa) Program. assignment (Q-learning & Expected Sarsa)	
	Mon, Oct 9	Thanksgiving		
6	Fri, Oct 13	Sample-based learning methods: Planning, learning, & acting	Practice quiz (Dealing with inaccurate models) Program. assignment (Dyna-Q & Dyna-Q+)	Chapter 8: Planning and Learning with Tabular Methods
7	Fri, Oct 20	Prediction and Control with FA: On-policy prediction with approx. Prediction and Control with FA: Constructing features for prediction	Practice quiz (On-policy prediction with approximation) Program. assignment (Semi-gradient TD(0) with state aggregation) Practice quiz (Constructing features for prediction) Program. assignment (Semi-gradient TD with a neural network) Practice quiz (Control with approximation) Program. assignment (Function approximation & control)	Chapter 9: On-policy Prediction with Approximation Chapter 10: On-policy Control with Approximation
Wed, Oct 25 at 23:59		**Project proposal**		

8	Fri, Oct 27	Prediction and Control with FA: Control with approximation Guest Lecture by Andrew Patterson: Empirical Design in Reinforcement Learning		Chapter 10: On-policy Control with Approximation Patterson et al. (2023): Empirical Design in Reinforcement Learning
9	Fri, Nov 3			Chapter 11: Off-policy Methods with Approximation Chapter 12: Eligibility Traces
	Fri, Nov 10	**Midterm**		
	Mon, Nov 13	Remembrance day holiday in lieu		
	Nov 14 - Nov 17	Reading week		
10	Fri, Nov 23	Prediction and Control with FA: Policy Gradient	Practice quiz (Policy gradient methods) Program. assignment (Average reward softmax Actor-Critic with tile-coding)	Chapter 13: Policy Gradient Methods
11	Fri, Dec 1	Deep Reinforcement Learning Major Successes of Reinforcement Learning		Chapter 16: Applications and Case Studies
12	Fri, Dec 8	Deep Reinforcement Learning Frontiers		Chapter 17: Frontiers
Fri, Dec 15 at 23:59		**Final course project**		