XCS234: Reinforcement Learning

Syllabus and Course Information

Welcome

Welcome to XCS234: Reinforcement Learning! This professional course is based on graduate-level material from Stanford's on-campus course CS234, adapted for a professional certificate format. In this course you will:

- Learn from Stanford graduate lecture videos (Winter 2021) that have been edited and segmented by topic for easier navigation, reference, and review.
- Complete guided homework assignments implementing content covered in the course lectures.
- Receive support from Stanford-affiliated Course Facilitators.
- Connect with a cohort of peers from diverse locations and professional backgrounds.

Course Launch

All lecture videos will be available on the first day of the course **(July 18th) at 12:00 pm Pacific Time**. Course assignments will be made live on the days indicated in the calendar below, without exception. Maintaining the assignment schedule enables Course Facilitators to be most effective in providing support and answering questions on the subject matter throughout the course.

Getting Started

This course will use different tools to distribute content, run assignments, and deliver support. They are:

- **SCPD Learning Management System** – accessed via the mystanfordconnection site which you used to apply to and enroll in this course.
- **GitHub** – to distribute programming assignment code.
- **Azure Lab Services** – to provide cloud computing resources for assignments.
- **Slack** – for additional course support and class discussions.

Joining Slack

In addition to individual support from Course Facilitators (more details and guidelines in Course Facilitators, Support, and Guidelines section below), the cohort will have a Slack workspace to ask additional questions and discuss course topics. An email invitation to the Slack workspace will be sent to your email address on file with SCPD on **July 15th**.

If you have previously joined an SCPD/Stanford Slack Workspace for a previous course in the AI Professional Program, Slack does not send a notification when our staff 're-invites' you to this workspace. Instead, you are automatically re-activated, and on **July 15th** should proceed directly to http://xcs234-scpd.slack.com/→ I have a guest account → Log in using your credentials.

Joining GitHub

You will receive an email invitation to a GitHub team called "XCS234-Summer-2022" at your address on file with SCPD. If you'd prefer that you receive a GitHub invitation at a different address, let us know at xcs234-staff@stanford.edu

GitHub invites can end up in the spam folder or directly in your GitHub account (especially if you have previously joined another SCPD GitHub). Let us know if you can't find your invite at xcs234-staff@stanford.edu

You will need to accept the invitation and be logged into your GitHub account in order to view course assignments and code.

The team's repository will be blank to begin with – code files will be added gradually as each assignment is released (see schedule below).

Getting Started (On July 18)

Accessing Your Course

1. On **July 18th after 12pm Pacific Time**, log in to the [mystanfordconnection](#) account you used when applying for the XCS234 course.
2. XCS234: Reinforcement Learning will be visible as a live course. Click the link titled "Course Videos and Assignments" to enter our learning management system.
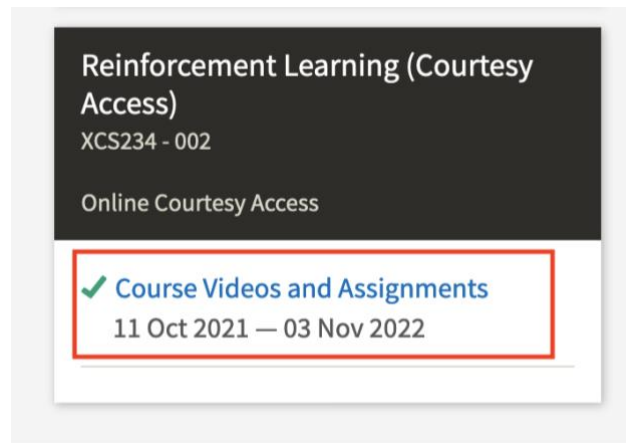
Course Calendar and Potential Pacing Guide

Below is a *potential* pacing guide if you are interested in watching all videos within a 10-week window. **However, you are free to view the videos at any pace you'd like.**

Week	Suggested Videos	Assignments	
Weeks 1 & 2 (Jul 18 – Jul 31)	Intro to RL Markov Decision Process Planning Model-Free Policy Evaluation Model-Based Policy Evaluation Policy Evaluation Algorithm Comparison Model-Free Control	A1 Released A2 Released	Monday, July 18 Monday, July 18
Weeks 3 & 4 (Aug 1 – Aug 14)	Intro to Value Function Approximation RL with Linear Value Function Approximation Deep RL	A1 Due A3 Released	Sunday, August 7 Friday, August 5
Weeks 5 & 6 (Aug 15 – Aug 28)	Intro to Policy Search Methods Gradient-Free Methods Policy Gradient Methods Fast RL: Intro to Multi-Armed Bandits	A2 Due A4 Released	Sunday, August 21 Friday, August 19
Week 7 & 8 (Aug 29 – Sep 11)	Fast RL: Evaluating RL Algorithms (MABs, Bayesian MABs, MDPs) Generalization and Strategic Exploration Batch/Offline RL	A3 Due	Sunday, September 4
Weeks 9 & 10 (Sep 12 – Sep 25)	Model-Based RL Monte Carlo Tree Search	A4 Due	Sunday, September 25

Assignments and Grading

Assignments will be made available via the SCPD course platform on the dates noted above in the course calendar. Below is a brief summary of what each assignment will entail:

Assignment 1

In this assignment, you will be exploring tabular solution methods. In particular, you will apply your understanding of MDPs to solve for the optimal paths an agent can take in a mobile game called "Flappy Karel". By the end of the assignment, you will have implemented both value iteration and policy iteration to solve for optimal value functions and policies in OpenAI's Frozen Lake gym environment as well as having explored topics such as the performance difference lemma and the Bellman operator.

Assignment 2

In this assignment we will be exploring deep reinforcement learning. In particular, we will explore the application of deep Q-learning in training an agent to outperform the average human in an Atari game known as Pong! The ultimate goal of this assignment is to demonstrate the effectiveness of combining the reinforcement learning concepts we have learnt thus far with the effectiveness of neural networks as function approximators. In addition, we will convey the importance of some of the techniques used in practice to stabilize training and achieve better performance.

Assignment 3

In this assignment, we will implement a policy gradient method (REINFORCE) on a number of rl environments commonly used for benchmarking (cartpole, pendulum, cheetah). This will include dealing with both continuous and discrete action spaces. We will also explore variance reduction methods to allow us to improve upon the performance of the algorithm we will be implementing (REINFORCE). By the end of this assignment, you should have an understanding of how to implement the policy gradient algorithm REINFORCE. You should also be able to apply and discuss variance reduction techniques for policy gradient methods.

Assignment 4

In this assignment, you will get the opportunity to apply what we have learned in class about multi-armed bandit problems to a real-world application. We will be predicting the appropriate dosage of a drug called Warfarin to prescribe to a patient given a set of features describing the patient. Following this practical application, we will also investigate a Bayesian regret bound for Thompson Sampling.

Honor Code

Students will be asked to review and maintain the standards set forth by the Stanford Honor Code when completing quizzes and assignments in this course. You can review the section labeled Violations of the Honor Code for representative examples relevant to this course.

Students are strongly encouraged to form study groups, discuss, and work on homework problems in groups and help each other; However, each student must write down the solutions independently and cannot refer to written notes from the joint session. **In other words, you must understand the solution well enough in order to reconstruct it independently**. Further, because we occasionally reuse assignment questions from previous years, you are expected not to copy, refer to, or look at the solutions in preparing your answers. It is an honor code violation to intentionally refer to a previous year's solutions

After completing this course, you are welcome to share your experience and credential with others; however, **it is considered a violation of the honor code to share assignment solutions including on public platforms such as GitHub**. Faculty in the computer science department have strongly encouraged us to refrain from posting solutions for assignments, thus we ask that you **DO NOT** share the exact code.

Grading

All assignments will be submitted via the Gradescope tool embedded within the SCPD learning management system.

- **Multiple Choice Quizzes, Coding Assignments and "Piazza Polls/Checks for Understanding"** can be submitted unlimited times up until the due date.
- **Written Questions** will be manually graded by Course Facilitators no later than one week after an assignment's 'on-time' due date. Problem sets turned in late may be graded slightly later.

Note on Final Projects

In the adaption from the CS234 graduate course to the XCS234 professional course, the final project has been removed.

Late Assignments and One-time Penalty Waiver

We understand that personal or professional events may cause you to miss a deadline on an assignment. All assignments can be turned in **up to five days late and are assessed a penalty of -1 point per day**. After five days, the submission link will close, and entries will no longer be accepted.

Each student can use a **one-time penalty waiver to remove a late penalty.** Note, that the waiver does **NOT** serve as an extension, and **cannot** be split into smaller parts (e.g., you cannot use two days on Assignment 3 and three days on Assignment 4.). In order to use your penalty waiver, contact your Course Facilitator and SCPD staff.

Passing the Course and Earning the Certificate

In order to earn the Certificate of Achievement associated with this course, you must complete problem sets with a total cumulative score of 70% or higher. Once you have successfully completed the course and the post-class survey, a digital badge for course completion will be emailed to you, which is the official certified Stanford document generated and held by Stanford Online. If you are interested in calculating your progress along the way, it may be helpful to know:

- There is a total of 300 base points in the course (meaning 210 to achieve 70%)
- There are 14 extra credit points – you will receive these points if you participate in Piazza Polls/Checks for Understanding

Deliverable	Points
A1 (Coding)	21
A1 (Written)	45
A1 (Quiz)	8
A2 (Coding)	33
A2 (Written)	25
A2 (Quiz)	18
A3 (Coding)	53
A3 (Written)	17
A3 (Quiz)	6
A4 (Coding)	57
A4 (Written)	15
A4 (Quiz)	2
Piazza Polls/Checks for Understanding (Extra Credit)	(14)

Total Available	**300** (+ 14 points of extra credit)
Minimum Passing Total	**210 (70 %)**

**Note, that along with moving towards more flexible and learner-friendly digital badges, SCPD is no longer issuing hard copies of course certificates to promote *Stanford green sustainable initiative*.

Videos and Slides

As noted, this course utilizes content originally delivered in the CS234 graduate course. A few things you will notice about this adaptation process:

- At times you will hear instructors make reference to the final project. As noted above these have been removed for the current version of XCS234 and you need not worry about the reference.
- Instructors may make reference to "Week 1", "Week 2", "Week n" of the course – in general, these references still hold due to the fact that the schedule/layout of the course has changed very little from the graduate course.
- Some sections of video have been removed, and accordingly those slides have been removed from the decks that accompany and play alongside each video. However, decks have not been re-numbered to ensure that they are still easy to match with the video segments. You will therefore see some slide decks that begin on a slide number other than 1 (often, as graduate course logistics videos have been cut) or slide decks that skip a few numbers due to removal of a clip (rarely, as much of the core lectures remain).
- In a few specific cases you may see names and/or faces blurred. In general, this is usually due to guidelines regarding student privacy.

Course Facilitators, Support, and Guidelines

You have a wide range of support available to you throughout the course. You will be assigned and receive contact information for an individual Course Facilitator (CF) who will act as your primary point of contact. Below is a summary of the available resources and course support:

Office Hours

Your CF will be in touch with availability and scheduling logistics for video conference office hours. Office hours may be conducted using the Zoom conference service or via Slack video (more information below on the course Slack workspace). Your CF will provide further information on how they will schedule and run office hour sessions.

Email

Your CF will also be available to answer questions via email – a Stanford contact address will be provided when you are first connected to your CF.

Slack Workspace – Usage and Guidelines

In addition to the individual and small group support provided by CFs, Slack will be a place where questions may be posed to the entire community (this is the fastest way to get an answer!). **In order to keep the Slack workspace readable, searchable, and useful to all, please follow the following guidelines:**

Reply in Threads to Keep Conversation Organized – When you are replying to a post or joining a conversation, respond by starting or joining a threaded conversation, rather than responding in the full flow of the standard timeline. See below for an example of how to respond in a threaded conversation:
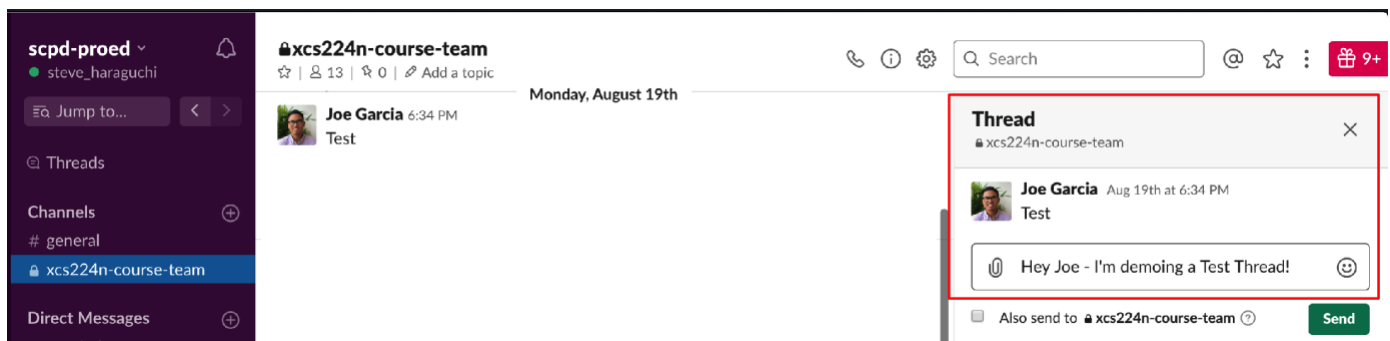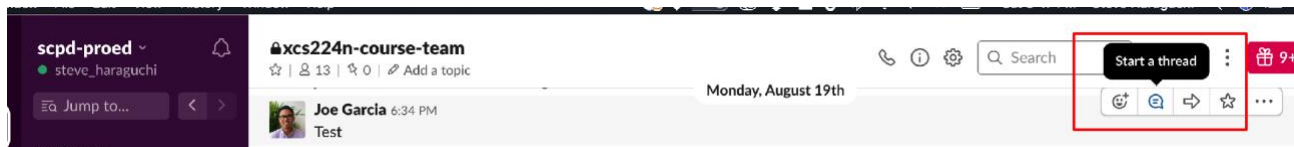
<mark>Use Multi-line Messages</mark> – Even if messages are threaded, you will soon see that Slack becomes unmanageable unless people use **single, multi-line messages instead of multiple, single-line messages**. Especially for mobile Slack users, it gets out of control!

Rather than the following:
"Hey all I have a question" [RETURN] <-- Creates new message
"I am a little confused about the quiz" [RETURN] <-- Creates new message
"I'm getting F for Question 40, but it seems like T is better" [RETURN] <-- Creates new message
----
Instead, try this!
"Hey all I have a question" [SHIFT+RETURN] <-- Creates new line in SAME message
"I am a little confused about the quiz" [SHIFT+RETURN] <-- Creates new line in SAME message
"I'm getting F for Question 40, but it seems like T is better" [SHIFT+RETURN] <-- Creates new line in SAME message
{RETURN} <-- Posts message

Note on Code Assignments and Debugging

While the course team is here to help and support your experience, it is ultimately your responsibility to write, test, and de-bug your own coding assignments. CFs may view and provide guidance on your work, however they will not send you exact answers on what to insert into your assignments. Additionally, before reaching out to a CF or Slack for help, it's expected that you have taken the reasonable step of performing an analysis yourself. This policy is meant to ensure that you leave the course having mastered the material and enable CFs to focus attention on questions where their guidance is most impactful.

Note on Networking

One of the benefits of this course is to be able to network with other course participants and create study groups. We encourage this kind of interaction and want to make sure that it is a positive experience. It is imperative that no course participant is made to feel uncomfortable or their ability to learn or otherwise benefit from the course is impeded by the actions of another participant. Please use good judgment. Keep interactions professional and focused on coursework or career networking. Avoid using offensive language and respect your colleagues' preferences regarding direct messaging. Please respect and uphold the rights and dignity of others regardless of race, color, national or ethnic origin, sex, age, disability, religion, sexual orientation, gender identity, or socio-economic status. Our team is always available either here in Slack or via email, so please feel free to reach out to us if you have any questions, concerns, or if any situation arises.

You can review SCPD's terms of service here, including rules for online conduct.

Drop/Transfer Policy

You may drop this course for a full refund up until July 18th, 2022 – the day the course starts. Once the course has begun, if you request to drop the course by Friday at 5:00 pm PST on the third week of the cohort (**August 5th, 2022**) you will be reimbursed 100% of your tuition minus a drop fee of $100. Beyond the third week of the course, tuition refunds are not granted. Before **August 5th, 2022** you may also request a transfer to a future cohort of XCS234 or a different course in the AI Professional Program, also for a $100 transfer fee. To request a drop or transfer, send an email to xcs234-staff@stanford.edu.

Questions

For course-specific questions or concerns (content, assignments, CF support), please contact your designated Course Facilitator.

For other course-related questions, email xcs234-staff@stanford.edu