

HOMWORK 2 TEMPLATE

Use this template to record your answers for Homework 2. Add your answers using L^AT_EX and then save your document as a PDF to upload to Gradescope. You are required to use this template to submit your answers. **You should not alter this template in any way** other than to insert your solutions. You must submit all **6** pages of this template to Gradescope. Do not remove the instructions page(s). Altering this template or including your solutions outside of the provided boxes can result in your assignment being graded incorrectly. You may lose points if you do not follow these instructions.

Instructions to upload code have been provided in the handout.

Instructions for Specific Problem Types

On this homework, you must fill in the blank for each problem; please make sure your final answer is fully included in the given space. **Do not change the size of the box provided.** For short answer questions you should **not** include your work in your solution. Only provide an explanation or proof if specifically asked. Otherwise, your assignment may not be graded correctly, and points may be deducted from your assignment.

Fill in the blank: What is the course number?

10-703

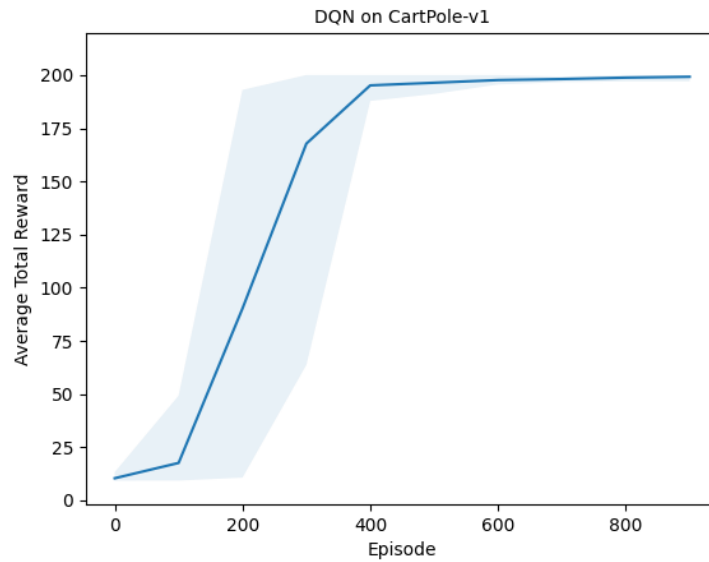
Problem 0: Collaborators

Enter your team's names and Andrew IDs in the boxes below. If you do not do this, you may lose points on your assignment.

Name 1:	<input type="text" value="Ruiyang Zhou"/>	Andrew ID 1:	<input type="text" value="ruiyangz"/>
Name 2:	<input type="text"/>	Andrew ID 2:	<input type="text"/>
Name 3:	<input type="text"/>	Andrew ID 3:	<input type="text"/>

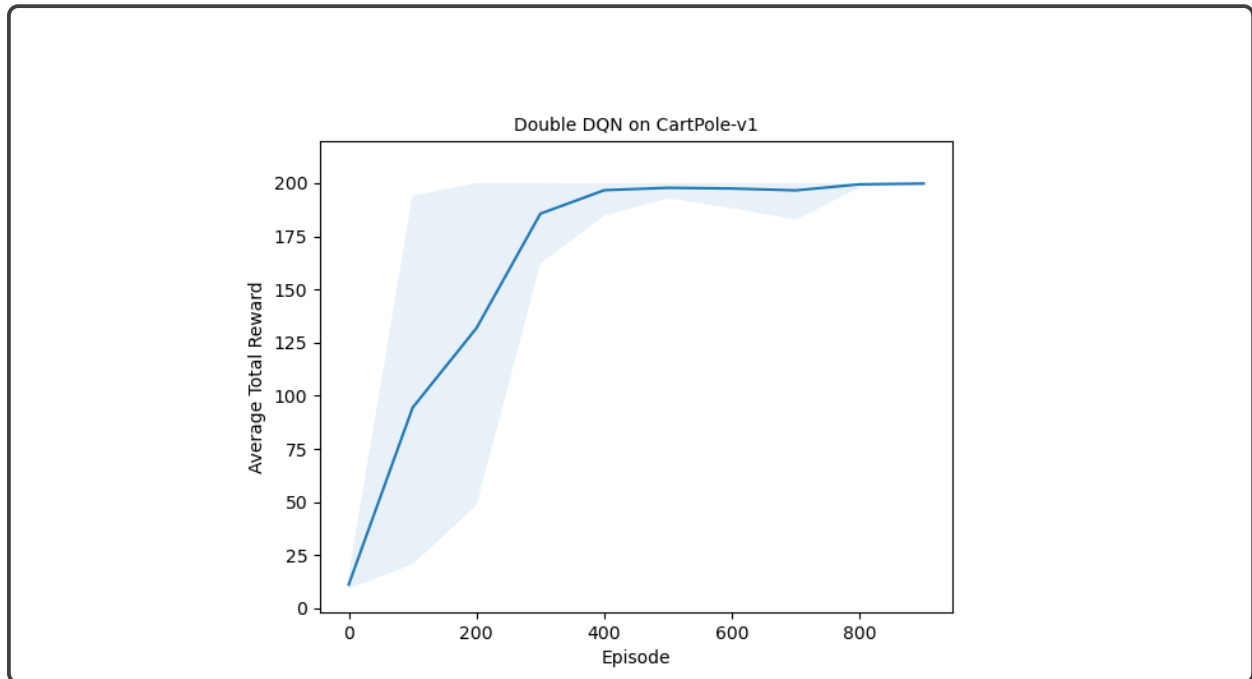
Problem 1: DQN (15 pts)

1.1 DQN plot (15 pts)



Problem 2: Double DQN (21 pts)

2.1: Double DQN plot (10 pts)



2.2 DQN vs. Policy Gradient Algorithms (5 pts)

Double DQN is directly learning from action-value estimates with bootstrapping, which has more sample-efficiency than policy gradient methods on CartPole. It also reduces overestimation bias caused by separating action selection and evaluation.

2.3 Pros and Cons of Policy gradient methods (6 pts)

Policy gradient methods like N-step A2C are more preferable when the action space is continuous or the stochastic policy is beneficial for exploration . The policy gradient method could also handle complex objectives and constraints, by directly parameterized. Meanwhile, DQN and double DQN are more suitable for the environments within a discrete action space and relative low-dimensional observations, like CartPole.

Feedback

Feedback: You can help the course staff improve the course for future semesters by providing feedback. You will receive a point if you provide actionable feedback. What was the most confusing part of this homework, and what would have made it less confusing?

the 'run 20 independent test episode' confused me for a long period of time.

Collaboration: Detail the work division amongst your group below.

Self work and asked some hints from other classmates.

Time Spent: How many hours did you spend working on this assignment? Your answer will not affect your grade. Please average your answer over all the members of your team.

Alone	20 hrs
With teammates	
With other classmates	1 hrs
At office hours	