

University of Virginia
Department of Computer Science

**CS 4501: Introduction to Reinforcement Learning
Fall 2022**

Tuesday 9:30am-9:45am PDT, September 27st, 2022

Name:
ComputingID:

- This is a **closed book** and **closed notes** quiz. No electronic aids or cheat sheets or discussing the questions with anyone else are allowed.
- You are expected to finish this quiz within 15 minutes.
- There are 2 pages, 3 parts of questions, and 20 total points in this quiz.
- The questions are printed on the back of this page!
- Please carefully read the instructions and questions before you answer them.
- If you need any clarification of the quiz questions, please raise your hand and discuss with the instructor within the quiz period.
- Try to keep your answers as concise as possible; our grading is *NOT* by keyword matching.

Total	/20
-------	-----

1 True/False Questions (2×3 pts)

Please choose either True or False for each of the following statements. For the statement you believe it is False, please give your brief explanation of it (you do **NOT** need to explain when you believe it is True). Three point for each question. *Note: the credit can only be granted if your explanation for the false statement is correct.*

1. The goal of a performing reinforcement learning algorithm is to minimize regret.
True
2. In multi-armed bandit problems, an optimal solution is expected to obtain zero regret.
False, and Explain: the regret lower bound of the bandit problems is already $\log(T)$.

2 Multiple Choice Questions (2×4 pts pts)

Please choose ALL the answers that you believe are correct for each question.

1. Why is RL difficult? (a) (b) (d)
(a) it is a combinatorial optimization problem in nature;
(b) we only have bandit feedback;
(c) the policy is stochastic;
(d) the observations are no longer IID.
2. What is/are important consideration(s) in designing a performing bandit algorithm:
(a) (b)
(a) sufficient exploration; (b) converging reward estimation;
(c) being stochastic; (d) understandable by human.

3 Short Answer Question (6 pts)

The question can be answered by one or two sentences; so please make your answer concise and to the point.

1. Assume you are having a job interview at iRobot, which is a leading manufacture in robot vacuum. You are asked to design an RL-based solution for their latest robot vacuum. As the first step, how would you design the reward, states and actions for such an RL system?

This is an open problem and credit will be given as long as your solution makes sense.

- Reward: +1, if the position has not been visited before; otherwise -1. This would encourage the robot to cover as large area as possible and then find the shortest path back to the base station.
- State/state transition: the next position where the robot moves to; if it hits an obstacle, stay at its current position; if it is the base station, terminates.
- Actions: move front, back, left or right