University of Virginia
Department of Computer Science

CS 4501: Introduction to Reinforcement Learning
Fall 2022

Monday 10:30am-10:45am EST, December 12th, 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. We choose to optimize mean square error in value function approximation because it is guaranteed to lead to improved policy.

   False*, and Explain*: We choose to optimize mean square error only because of its mathematical simplicity; instead, Bellman error is the ideal objective for function approximation, but it is computationally infeasible to optimize.

2. Experience replay helps us reduce bias in approximation-based RL algorithms.

   False*, and Explain*: It helps us to reduce variance in parameter estimation, just as batch update in SGD.

2 Multiple Choice Questions (2×4 pts pts)

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

1. What are the key design choices for AlphaGo? (a) (b) (c) (d)
   (a) Supervised policy learning; (b) Self-play; (c) Monte Carlo tree search; (d) CNN for state encoding.

2. What constitute the so-called deadly triad? (a) (c) (d)
   (a) function approximation; (b) non-stationarity; (c) off-policy learning; (d) bootstrapping.

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. Recall in Quiz 1, you were asked to design an RL-based solution for a robot vacuum. Now how are you going to modernize your robot vacuum with deep RL techniques?

   We can keep your previous state and reward design, but use neural networks to enhance the state representation, Q-function and/or your policy. Basically, neural networks provide you a flexible way to encode the problem.
   For example, we can use CNN to encode imagery sensor input as part of the state, and use an RNN to encode the traveled trajectory to enhance state representation. For the value function and policy, we can use an MLP to map state action pairs into value or probablity estimates.