

# Reinforcement Learning—Quiz 8 Solution

April 6, 2020

Name: \_\_\_\_\_

Score: \_\_\_\_\_

1. (25 points) What is planning?

Direct reinforcement learning updates a policy based on interactions with the environment.

Planning (indirect reinforcement learning), however, updates a policy based on a model.

2. (25 points) Dyna-Q

The Dyna-Q algorithm intersperses  $n$  steps of simulated experience for every step of direct RL.

3. (25 points)  $n$ -Step TD

Mark all of the following  $n$ -step Temporal Difference learning algorithms that work off-policy but *don't* require knowing the probabilities of the behavior policy?

**A. Tree-Backup**

B.  $n$ -step Expected Sarsa

C.  $n$ -step Sarsa

4. (25 points)  $n$ -Step TD

Mark all of the following  $n$ -step Temporal Difference learning algorithms that work off-policy but *do* require knowing the probabilities of the behavior policy?

A. Tree-Backup

**B.  $n$ -step Expected Sarsa**

**C.  $n$ -step Sarsa**