Final Exam Sample Questions

1. The attack named MobilBye targets which type of sensor?

A. Camera

B. Lidar

C. Radar

ANS: _____

A

2. Can an ASIL-D system be composed of multiple ASIL-B subsystems?

A. Yes

B. No

ANS: _____

A

3. In PID control, which term(s) can be used to eliminate steady-state error?

A. P

B. I

C. D

ANS: _____

B

4. True or False: twiddle() for PID controller tuning can be viewed as a form of Reinforcement Learning.

A. True

B. False

ANS: _____

A

5. Why does AlphaGo not use Tabular RL, such as tabular Q Learning or Sarsa?

ANS: the state space is too large.

6. What is the probability of selecting the greedy action (assuming there is only one) in state s in epsilon-greedy action selection, if the total number of possible actions in state s is N?

ANS: $1 - \epsilon + \frac{\epsilon}{N}$

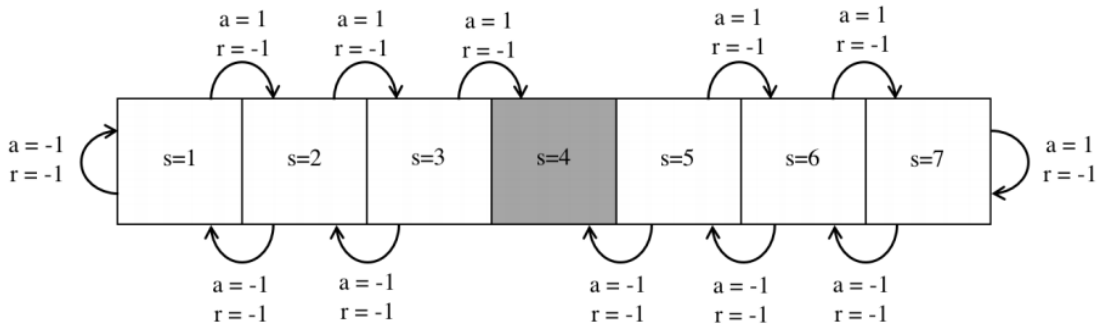
7. What is an Episodic task?

A. A task that has a limited number of actions and then ends.

B. A task with memory

ANS: A

8. Consider the following MDP. Environment is deterministic. In each state, there are two possible actions a $\in \{-1, 1\}$, where −1 corresponds to moving left, and 1 corresponds to moving right. Each movement incurs a reward of r=-1. State s=4 is the goal state: taking any action from s=4 results in reward of r=0 and ends the episode, hence $V(4) \equiv 0, Q(4, a) \equiv 0$ for any action a.



Consider this episode in the form of (s,a,r): (3, −1, −1),(2, 1, −1),(3, 1, −1),(4, 1, 0). Assume $\gamma = 1, \alpha = 0.5$. All value functions are initialized to 0. Derive the following (only show the changed parts):

1. State value functions after TD learning.
2. Action value functions after Sarsa.
3. Action value functions after Q learning.

ANS:

TD update equation: $V(S_t) \leftarrow V(S_t) + \alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t))$

1. $V(3) \leftarrow V(3) + 0.5(-1 + V(2) - 0) = 0 + 0.5(-1 + 0 - 0) = -0.5$
2. $V(2) \leftarrow V(2) + 0.5(-1 + V(3) - 0) = 0 + 0.5(-1 - 0.5 - 0) = -0.75$
3. $V(3) \leftarrow V(3) + 0.5(-1 + V(4) - 0) = -0.5 + 0.5(-1 + 0 + 0.5) = -0.75$

Sarsa update equation: $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t))$

4. $Q(3, -1) \leftarrow Q(3, -1) + 0.5(-1 + Q(2,1) - 0) = 0 + 0.5(-1 + 0 - 0) = -0.5$
5. $Q(2,1) \leftarrow Q(2,1) + 0.5(-1 + Q(3,1) - 0) = 0 + 0.5(-1 + 0 - 0) = -0.5$
6. $Q(3,1) \leftarrow Q(3,1) + 0.5(-1 + Q(4,1) - 0) = 0 + 0.5(-1 + 0 - 0) = -0.5$

Q learning update equation: $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t))$

7. $Q(3, -1) \leftarrow Q(3, -1) + 0.5\left(-1 + \max_{a'} Q(2, a') - 0\right) = 0 + 0.5(-1 + 0 - 0) = -0.5$

8. $Q(2,1) \leftarrow Q(2,1) + 0.5\left(-1 + \max_{a'} Q(3,a') - 0\right) = 0 + 0.5(-1 + \max(-0.5,0) - 0) = -0.5$
9. $Q(3,1) \leftarrow Q(3,1) + 0.5\left(-1 + \max_{a'} Q(4,a') - 0\right) = 0 + 0.5(-1 + 0 - 0) = -0.5$