Assignment 12
Reinforcement Learning
Prof. B. Ravindran

1. (1) In the partial observability example that we saw, there were some observations which uniquely identify the state of the environment. Suppose in a problem, no such observations exist, i.e., there are no observations which allow us to exactly determine the state of the environment. In such problems, is it ever possible to be sure which state you are in given only the sequence of observations?

   (a) no
   (b) yes

2. (1) Suppose that some of the actions that are available to an agent in a partially observable environment have non-deterministic outcomes. How would this impact the agent's ability to determine the state it is in?

   (a) it would make determining the state easier
   (b) it would make determining the state harder
   (c) it would have no effect in determining the state

3. (1) Referring to the partial observability example considered in the lectures and assuming that there is no noise in the sensor outputs, is it true that all observation sequences of length 3 or more result in the elimination of uncertainty regarding the position of the agent in the environment?

   (a) no
   (b) yes

4. (1) To solve a POMDP problem, suppose you decide to maintain belief states. Would the estimation of the belief states benefit from access to the history of observations and/or actions?

   (a) access to past observations, but not actions, would improve belief state estimation
   (b) access to past actions, but not observations, would improve belief state estimation
   (c) access to past observations and actions would improve belief state estimation
   (d) access to past observations or actions will not improve belief state estimation

5. (1) Suppose that you are given a problem in which the agent is able to determine the state it transitions to after each action taken by the agent. However, the state space is described in a manner which requires the agent to consider information in the current, as well as the latest past state in the sequence, to make action decisions. Is the problem so defined an MDP, a POMDP, or neither? Can this problem be solved using RL techniques that we have studied (perhaps with possible modifications to the problem definition)?

    (a) MDP, yes

    (b) POMDP, yes

    (c) neither, no

    (d) neither, yes

6. (1) Recall the example partially observable environment encountered in the lectures. Suppose the goal in this problem is to reach the bottom-right state, with the usual -1 rewards for each transition. Assume that the agent starts two cells below the top left state and the first action selected by the QMDP procedure is to move down. Assuming deterministic action outcomes with no noise in the observations, as well as the use of the QMDP procedure where the underlying MDP has been correctly solved, what can you say about the optimality of the agent's policy in this scenario?

    (a) the agent will follow an optimal policy to reach the goal

    (b) the agent's policy to reach the goal will not be optimal

    (c) the optimality of the agent's policy towards reaching the goal in this scenario cannot be determined with certainty