Assignment 11 (Sol.)

Reinforcement Learning

Prof. B. Ravindran

1. Using the MAXQ approach leads to solutions which are

   (a) hierarchically optimal

   (b) recursively optimal

   (c) flat optimal

   Sol. (b)
   Since the MAXQ policy of the core MDP is the set of policies of individual sub-tasks, with individual sub-task policies aiming to solve the sub-tasks optimally, you can expect to obtain recursively optimal solutions using the MAXQ approach.

2. We saw that each sub-task has an associated pseudo-reward function. Are the rewards of the core MDP available to the agent while it is learning policies of individual sub-tasks or is the agent restricted to the corresponding sub-task's pseudo rewards?

   (a) only pseudo rewards are available

   (b) both pseudo rewards and core MDP rewards are available

   Sol. (b)
   As we observed in the example taxi problem, rewards of the core MDP are available while learning the policies of the sub-tasks.

3. In the MAXQ framework, is termination in a sub-task deterministic or stochastic as in the options framework?

   (a) deterministic

   (b) stochastic

   Sol. (a)
   We saw in the sub-task definition that for each sub-task, all states of the core MDP are partitioned into a set of active states and a set of terminal states, where sub-task termination is immediate (and deterministic) whenever a terminal state is entered.

4. Each sub-task M_i is an SMDP because

   (a) the state space of the sub-task is a subset of the state space of the core MDP

   (b) each sub-task has its own policy

(c) actions in a sub-task can be temporally extended

(d) the rewards received in sub-tasks depend not only on the state but also on the sub-task in which an action was executed

Sol. (c)
In the definition, we saw that the actions in a sub-task comprise both, primitive actions as well as other sub-tasks. The invocation of a sub-task results in a sequence of actions being executed (similar to an option). Thus, each sub-task is an SMDP.

5. The expected reward function $\bar{R}(s, a)$ of the SMDP corresponding to sub-task M_i is equivalent to the projected value function $V^{\pi_i}(a, s)$. True or false?

   (a) false

   (b) true

   Sol. (a)
   Recall that $\bar{R}(s, a) = V^{\pi}(a, s)$ not $V^{\pi_i}(a, s)$.

6. In the MAXQ approach to solving a problem, suppose that sub-task M_i invokes sub-task M_j. Do the pseudo rewards of M_j have any effect on sub-task M_i?

   (a) no

   (b) yes

   Sol. (b)
   The pseudo rewards of one sub-task are not directly considered when solving a different sub-task regardless of their connectivity. However, since sub-task M_i invokes sub-task M_j, and hence depends upon the policy of M_j, the rewards of M_j do effect sub-task M_i, as the pseudo rewards of sub-task M_j would be a factor determining the policy of M_j.