Assignment 11
Reinforcement Learning
Prof. B. Ravindran

1. Using the MAXQ approach leads to solutions which are

   (a) hierarchically optimal
   (b) recursively optimal
   (c) flat optimal

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

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

   (a) deterministic
   (b) stochastic

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

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

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