# Assignment 8
## Reinforcement Learning
### Prof. B. Ravindran

1. Is the problem of non-stationary targets an issue when using Monte Carlo returns as targets?

    (a) no

    (b) yes

2. In a parameterised representation of the value function, we use a feature which acts as a counter of some concept in the environment (number of cans the robot has collected, for example). Does such a feature used for representing the state space lead to a violation of the Markov property?

    (a) no

    (b) yes

3. Which of the following will effect generalisation when using the tile coding method?

    (a) modify the number of tiles in each tiling (assuming the range covered along each dimension by the tilings remains unchanged)

    (b) modify the number of tilings

    (c) modify the size of tiles

    (d) modify the shape of tiles

4. For a particular MDP, suppose we use function approximation and using the gradient descent approach, converge to the value function that is the global optimum. Is this value function the same, in general, as the true value function of the MDP?

    (a) no

    (b) yes

5. Which of the following methods would benefit from normalising the magnitudes of the basis functions?

    (a) on-line gradient descent TD($\lambda$)

    (b) linear gradient descent Sarsa($\lambda$)

    (c) LSPI

    (d) none of the above

6. Suppose that individual features, $\phi_i(s, a)$, used in the representation of the action value function are non-linear functions of $s$ and $a$. Is it possible to use the LSTDQ method in such scenarios?

   (a) no

   (b) yes

7. Which among the following statements about the LSTD and LSTDQ methods is/are correct?

   (a) LSTD learns the state value function

   (b) LSTDQ learns the action value function

   (c) both LSTD and LSTDQ can reuse samples

   (d) both LSTD and LSTDQ can be used along with tabular representations of value functions

8. Consider the five state random walk task described in the book. There are five states, $\{s_1, s_2, ..., s_5\}$, in a row with two actions each, left and right. There are two terminal states at each end, with a reward of $+1$ for terminating on the right, after $s_5$ and a reward of 0 for all other transitions, including the one terminating on the left after $s_1$. In designing a linear function approximator, what is the least number of state features required to represent the value of the equi-probable random policy?

   (a) 1

   (b) 2

   (c) 3

   (d) 5