Homework Assignment # 3

Due: Tuesday, March 22, 2016, 11:59 p.m.

Total marks: 100

Question 1. [5 MARKS]

Exercise 6.2 in Sutton and Barto 2nd Ed., 2016. [Answer each of the three parts to this question]. Think of the transitions and rewards that could produce this update to V(A). The third part is looking for a numerical answer.

Question 2. [5 MARKS]

Exercise 6.5 in Sutton and Barto 2nd Ed., 2016.

Question 3. [55 MARKS]

Programming question

[Part one: worth 45] Program a windy Gridworld with King's Moves. Re-solve the windy gridworld task assuming eight possible actions, including the diagonal moves, rather than the usual four. As in Example 6.5 we will use $Q_0(s, a) = 0$ for all s, a, and $\epsilon = 0.1$ and $\alpha = 0.5$. You will implement two agents and compare their performance: (1) Sarsa as described in Figure 6.5, and (3) expected Sarsa as described by equation 6.7.

Plot the number of steps to goal over the first 100 episodes (in RL-glue you can use RL_episode() function), averaged over 100 runs of the experiment. This is called a *learning curve*. In RL-glue we can access the number of steps taken in an episode by calling RL_num_steps(). You should submit one plot with two lines—one for each of the two agents—or two separate plots.

This will require you to implement several things:

1. a simulation of the windy gridwold problem, as described in Example 6.5, except with 8 actions (King's moves) (an Environment program)
2. Sarsa, and Expected Sarsa (Agent programs)
3. code to run the experiment for 100 episodes, averaging over 100 runs (Experiment program)

Please submit your plots and ALL your code (including any scripts and data processing).

[Part two: worth 10] Experiment with different α and ϵ values for one of the agents (either Sarsa or Expected Sarsa). Discuss how changing α and ϵ effects the agent's learning performance. You can include graphs to help with your explanation.

Question 4. [5 MARKS]

Exercise 6.9 in Sutton and Barto 2nd Ed., 2016.

Question 5. [5 MARKS]

Exercise 7.2 in Sutton and Barto 2nd Ed., 2016. Online means updating during the episode.

Question 6. [20 MARKS]

Programming question:

Resolve the windy gridworld with King's moves from Question #4, this time using Sarsa(λ). The environment and experiment program will be the same. However, you will implement one new agent: Sarsa(λ) with **replacing traces**. Test λ values of 0.1, 0.5, and 0.9. Plot the number of steps to goal over the first 100 episodes, averaged over 100 runs of the experiment. Your plot should have three lines (learning curves), one for each value of λ.

You may have to experiment with different initializations of the value function Q_0, different exploration rates ϵ, and different learning rates α.

Please submit your plots and ALL your code (including any scripts and data processing).

Question 7. [5 MARKS]

Exercise 7.7 in Sutton and Barto 2nd Edition. Start with the pseudo code in Figure 7.7. How could you modify this algorithm to keep track of the $E(s)$ values greater than some ϵ, and then only update the corresponding values.

Homework policies:

Your assignment will be submitted as a single pdf document and a zip file with code, on canvas. The questions must be typed; for example, in Latex, Microsoft Word, Lyx, etc. or must be written legibly and scanned. Images may be scanned and inserted into the document if it is too complicated to draw them properly.

Policy for late submission assignments: Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rule:

on time: your score 1

1 day late: your score 0.9

2 days late: your score 0.7

3 days late: your score 0.5

4 days late: your score 0.3

5 days late: your score 0.1

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points.

All assignments can be done in collaboration, however, you must write your own answers, write your own programs, and generate your own results (data and graphs). All the sources used for problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Good luck!