Homework Assignment # 1

Due: Tuesday, February 9, 2016, 11:59 p.m.

Total marks: 100

1 Questions from Chapter 2

Question 1. [10 MARKS]

Exercise 2.2 in Sutton and Barto 2nd Ed., 2016. That is, starting from equation 2.5 (page 38, Sutton & Barto V2, 2016) re-derive equation 2.6 so that α is replaced by α_i. Show your steps; if you do not know what "show your steps" means please ask the instructor.

Question 2. [5 MARKS]

Exercise 2.4 in Sutton and Barto 2nd Ed., 2016. Start off by explaining the big spike in early learning.

Question 3. [35 MARKS]

Programming: Design and conduct an experiment to demonstrate the difficulties that sample-average methods have for non-stationary problems. Use a modified version of the 10-armed testbed in which all $q^\star(a)$ start out equal and then take independent random walks. That is for each $q^\star(a)$, every 100 steps randomly increment (50% chance) or decrement (50% chance) $q^\star(a)$ by some small amount. The choice of how big the increment/decrement is yours, but choose it so that we can see a clear difference between the two learning algorithms described below. Prepare a plot like the top one in Figure 2.2 (average reward) for an action-value method using sample averages, incrementally computed by $\alpha = \frac{1}{n}$, and another action-value method using a constant step-size parameter, $\alpha = 0.1$. Use $\epsilon = 0.1$ and, if necessary, runs longer than 1000 plays. Don't forget to sample each arm stochastically (i.e, $R_t \sim \mathcal{N}(q^\star(A_t), 1)$) on each time step).

This will require you to implement three things:

1. a non-stationary 10-armed bandit environment
2. two learning agents
3. code to run the experiment (average over runs, etc.)

You do not have to use RL-glue (code can be found on canvas) for this question, however, all future programming assignment questions will require using RL-glue.

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

2 Questions from Chapter 3

Question 4. [10 MARKS]

Exercise 3.1 in Sutton and Barto 2nd Ed., 2016. Reinforcement learning tasks.

Question 5. [5 MARKS]

Exercise 3.5 in Sutton and Barto 2nd Ed., 2016. Careful design of reward functions.

Question 6. [5 marks]

Exercise 3.6 in Sutton and Barto 2nd Ed., 2016. Understanding the Markov property.

Question 7. [10 marks]

Exercise 3.7 in Sutton and Barto 2nd Ed., 2016. Recursive relationship between action-value of a state and future states.

Question 8. [5 marks]

Exercise 3.8 in Sutton and Barto 2nd Ed., 2016. Checking the Bellman equation.

Question 9. [5 marks]

Exercise 3.11 in Sutton and Barto 2nd Ed., 2016. Relationship between state values and action values. Show your steps.

Question 10. [10 marks]

Exercise 3.16 in Sutton and Barto 2nd Ed., 2016. Think of what the policy does and the rewards it gets.

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!