COMP532 Assignment 1 –
Reinforcement Learning

You need to solve each of the following problems. Problem 1 concerns an example/exercise from the book of Sutton and Barto. You must also include a brief report describing and discussing your solutions to the problems. This work can be carried out in pairs of two persons.

- This assignment is worth 10% of the total mark for COMP532
- 80% of the marks will be awarded for correctness of results.
- 20% of the marks will be awarded for the quality of the accompanying report

Submission Instructions

- Send all solutions as 1 PDF document containing your answers, results, and discussion of results. Attach the source code for the programming problems as separate files.
- Submit your solution by email to shan.luo@liverpool.ac.uk, clearly stating in the subject line: “COMP532 Task 1 Solution”
- **The deadline for this assignment 09/03/2018, 5:00pm**

- Penalties fo tal policy as set out in the student han

dbook.pdf
and the Uni
https://www.liverpool.ac.uk/media/liva ce-on-assessment/code_of_practice

Problem 1 (12 marks)

Re-implement (e.g. in Matlab) the results presented in Figure 2.2 of the Sutton & Barto book comparing a greedy method with two ε-greedy methods ($\varepsilon = 0.01$ and $\varepsilon = 0.1$), on the 10-armed testbed, and present your code and results. Include a discussion of the exploration - exploitation dilemma in relation to your findings.

Problem 2 (8 marks)

Consider an MDP with states $S = \{4,3,2,1,0\}$, where 4 is the starting state. In states $k \geq 1$ you can walk (W) and $T(k, W, k-1) = 1$. In states $k \geq 2$ you can also jump (J) and $T(k, J, k-2) = 3/4$ and $T(k, J, k) = 1/4$. State 0 is a terminal state. The reward $R(s, a, s') = (s - s')^2$ for all (s, a, s'). Use a discount of $\gamma = 1/2$. Compute both $V^*(2)$ and $Q^*(3, J)$. Clearly show how you computed these values.

Problem 3 (5 marks)

a) What does the Q-learning update rule look like in the case of a stateless or 1-state problem? Clarify your answer. (2 marks)
b) Discuss the main challenges that arise when moving from single- to multi-agent learning, in terms of the learning target and convergence. (3 marks)

Problem 4 (15 marks)

Re-implement (e.g. in Matlab) the results presented in Figure 6.4 of the Sutton & Barto book comparing SARSA and Q-learning in ... sing different values for the exploration parameter ε fo... our discussion clearly describe the main difference betwee... gs.

Note: the book is not completely clear on this example. Use ... for both algorithms. The "smoothing" that is mentioned in the caption of Figure 6. ... ng over 10 runs, and 2) plotting a moving average over the last 10 episodes.