**DDA 4230: Reinforcement learning** **CUHKSZ**

# Assignment 3

**TA: Sheng Xu**

**Due Date: Nov. 28th, 11:59 pm**
Total points available: 100 pts.

**Note**: *Please note that external references are allowed only if you give appropriate reference. There is no required format of reference. Please elaborate on your answers as well (do not just give a number, etc).*

## Problem 1: Q Learning Computation

Consider a vanilla Pacman game, which is an unknown MDP with three states $[A, B, C]$ and two actions $[Stop, Go]$. In this problem, the discount factor and learning rate are set as $\gamma = 1$, $\alpha = 0.5$. With a random policy taking actions in this unknown MDP, we have collected some samples as follows.

1. Initialize all Q-values to 0. Given the samples as shown in Table 1, calculate the estimation of the Q-values $Q(C, Stop)$ and $Q(C, Go)$ by Q-learning with a detailed updating process. Indicate the samples you used for updating Q-values.

| $s$ | $a$ | $s'$ | $r$ |
|---|---|---|---|
| A | Go | B | 2 |
| C | Stop | A | 0 |
| B | Stop | A | -2 |
| B | Go | C | -6 |
| C | Go | A | 2 |
| A | Go | A | -2 |

Table 1: Samples for Problem 1.1

2. In this part, we will consider a feature-based representation. The features are set as

$$f_1(s, a) = 1$$
$$f_2(s, a) = \begin{cases} 1 & a = Go \\ -1 & a = Stop \end{cases} \tag{1}$$

Then the Q-value functions can be represented as $Q(s, a) = w_1 f_1(s, a) + w_2 f_2(s, a)$. Initialize the weights $\mathbf{w}^0 = [w_1^0, w_2^0]$ as 0-vector, where the superscripts represent the update iterations. Given the samples in Table 2, calculate the updated weights $\mathbf{w}^1 = [w_1^1, w_2^1]$ using the sample in the first row, and $\mathbf{w}^2 = [w_1^2, w_2^2]$ after the second update using the sample in the second row. Show your calculation process.

| $s$ | $a$ | $s'$ | $r$ |
|---|---|---|---|
| A | Go | B | 4 |
| B | Stop | A | 0 |

Table 2: Samples for Problem 1.2

## Problem 2: Q Learning Implementation

Code skeleton is given at `https://colab.research.google.com/drive/1JV0pNhMGVMrm0W2J-3HDxaVPXgznehAK?usp=sharing`. Please copy the file to your folders and complete the Q-Learning algorithm following the instructions in the notebook. Submit your file in .ipynb. Do not revise the original files in the provided link.

## Problem 3: REINFORCE Implementation

Code skeleton is given at `https://colab.research.google.com/drive/1JlP3PO6vc1baeF3FtfCpwQndW8-3zdPP?usp=sharing`. Please copy the file to your folders and complete the REINFORCE algorithm following the instructions in the notebook. Submit your file in .ipynb. Do not revise the original files in the provided link.