Group 44:
Ricardo Rei nº 78047
Miguel Carvalho nº 78052

TÉCNICO LISBOA

MSc in Computer Science and Engineering

Learning and Decision Making 2016-2017

Homework 5. Reinforcement Learning

1.

(a)

$\mathbb{X}$ = {00, 01, 02, 03, 04, 10, 11, 12, 13, 14, 20, 21, 22, 23, 24}, where the 00 state is the top left square in figure 1 and 24 is the bottom right square in figure 1.

$\mathbb{A}$ = {Up, Down, Left, Right}

C =

U	D	L	R	
1.0	1.0	1.0	1.0	#00
1.0	1.0	1.0	1.0	#01
1.0	1.0	1.0	1.0	#02
1.0	1.0	1.0	1.0	#03
1.0	**0.0**	1.0	1.0	#04
1.0	1.0	1.0	1.0	#10
1.0	1.0	1.0	1.0	#11
1.0	1.0	1.0	1.0	#12
1.0	1.0	1.0	1.0	#13
1.0	1.0	1.0	1.0	#14
1.0	1.0	1.0	1.0	#20
1.0	1.0	1.0	1.0	#21
1.0	1.0	1.0	1.0	#22
1.0	1.0	1.0	**0.0**	#23
0.0	1.0	1.0	1.0	#24

(b)

$\gamma = 0.95$
$s_0 = 10$
$A_1 = A_2 =$ Right

Transition Information for $A_1 = \{s_0, A_1, 1.0, 11\}$
Transition Information for $A_2 = \{11, A_2, 1.0, 02\}$

(c)

Given the samples from exercise (b) and a Q_0 matrix with all-zeros, we compute Q with the following equation:

$$Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t) + \alpha_t \left[c_t + \gamma \min_{a' \in \mathcal{A}} Q_t(x_{t+1}, a') - Q_t(x_t, a_t) \right]$$

Group 44:
Ricardo Rei nº 78047
Miguel Carvalho nº 78052

which after the first update results in:

$$Q_1 = \begin{vmatrix} 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & \mathbf{0.1} \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \end{vmatrix}$$

and after the second update results in:

$$Q_2 = \begin{vmatrix} 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & \mathbf{0.1} \\ 0. & 0. & 0. & \mathbf{0.1} \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \end{vmatrix}$$