

MSc in Computer Science and Engineering

**Learning and Decision Making 2016-2017**

Homework 5. Reinforcement Learning

**(a)**

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.

A = {Up, Down, Left, Right}

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

C = | 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)**

**γ** = 0.95

= 10

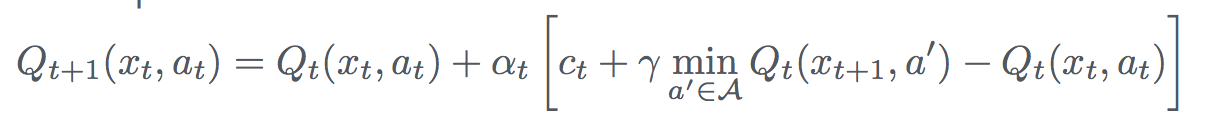
= = Right

Transition Information for = {, , 1.0, 11}

Transition Information for = {11, , 1.0, 02}

**(c)**

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



which after the first update results in:

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. **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. |

| 0. 0. 0. 0. |

and after the second update results in:

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. 0. |

| 0. 0. 0. **0.1**|

| 0. 0. 0. **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. |