EE932 Assignment-2 Solution

---

eMasters in Communication Systems, IITK
EE932: Introduction to Reinforcement Learning
Instructor: Prof. Subrahmanya Swamy Peruru
Student Name: Venkateswar Reddy Melachervu
Roll No: 23156022

Question 13:

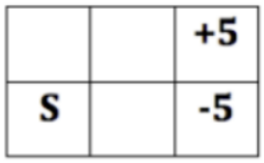

Figure 1: Grid-World

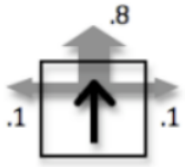

Figure 2: Tranistion probabilities for an 'UP' action

Consider the grid shown in Figure 1. The states are grid squares, identified by their row and column numbers. The agent always starts in the bottom left state (1,1), marked with the letter S. (Note that the bottom row is denoted by number 1, the top row by number 2). There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. (The reward for a state is received as the agent moves into the state.) The transition function is such that the intended agent movement (UP, Down, Left, or Right) happens with probability 0.8. With probability 0.1 each, the agent ends up in one of the states perpendicular to the intended direction. Please refer Figure 2.

Verify that the optimal value function that you intuitively guessed is indeed correct by using the Bellman optimality equations

Solution:

$$V^*(s) = \max_a \left\{ R_s^a + \sum_{s'} P_{ss'}^a V^*(s') \right\}$$

$$V^*(s') = \sum_a Q^*(s,a)$$

$V^*(2,3) = max\{0\} = 0$

Per Bellman's optimality = $V^*(2,2) = \underset{a}{max}\{5 + \{0.8 * V^*(2,3), 0.1 * V^*(1,2), 0.1 * V^*(2,2)\}\} = 5$

Intuitivelu arrived at: $V^*(2,2) = 5 + 0.8 * 0 = 5$

Per Bellman's optimality = $V^*(1,3) = \underset{a}{max}\{5 + \{0.8 * V^*(2,3), 0.1 * V^*(1,2), 0.1 * V^*(1,3)\}\} = 5$

Inutitively arrived at :$V^*(1,3) = 5 + 0.8 * 0 = 5$

Per Bellman's optimality = $V^*(1,2) = \underset{a}{max}\{0 + \{0.8 * V^*(2,2), 0.1 * V^*(1,3), 0.1 * V^*(1,1)\}\} = 4$

Intuitively arrived at: $V^*(1,2) = 0 + 0.8 * 5 = 4$

Per Bellman's optimality = $V^*(2,1) = \underset{a}{max}\{0 + \{0.8 * V^*(2,2), 0.1 * V^*(1,1), 0.1 * V^*(2,1)\}\} = 4$

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu

Intuitively arrived at : $V^*(2,1) = 0 + 0.8 * 5 = 4$

Per Bellman's optimality = $V^*(1,1) = \max_a \{0 + \{0.8 * V^*(2,1), 0.1 * V^*(1,1), 0.1 * V^*(1,2)\}\} = 3.2$
Intutively arrived at : $V^*(1,1) = 0 + 0.8 * 4 = 3.2$

(2,1) $V^*(2,1) = 4$	(2,2) $V^*(2,2) = 5$	+5 (2,3) $V^*(2,3) = 0$
S (1,1) $V^*(1,1) = 3.2$	(1,2) $V^*(1,2) = 4$	-5 (1,3) $V^*(1,3) = 5$

---------- End of the Document ----------

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu