



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 12:

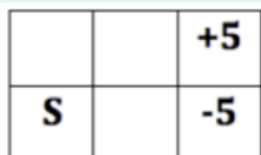


Figure 1: Grid-World

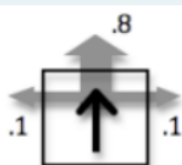


Figure 2: Transition probabilities for an 'UP' action

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.

What is the optimal value function for each state? You can answer this by intuition

Solution:

$$V^*(2,3) = 0$$

$$V^*(2,2) = 5 + 0.8 * 0 = 5$$

$$V^*(1,3) = 5 + 0.8 * 0 = 5$$

$$V^*(1,2) = 0 + 0.8 * 5 = 4$$

$$V^*(2,1) = 0 + 0.8 * 5 = 4$$

$$V^*(1,1) = 0 + 0.8 * 4 = 3.2$$

$$\begin{array}{c} (2,1) \\ V^*(2,1) = 4 \end{array}$$

$$\begin{array}{c} (2,2) \\ V^*(2,2) = 5 \end{array}$$

$$\begin{array}{c} +5 \\ (2,3) \\ V^*(2,3) = 0 \end{array}$$

$$\begin{array}{c} S \\ (1,1) \\ V^*(1,1) = 3.2 \end{array}$$

$$\begin{array}{c} (1,2) \\ V^*(1,2) = 4 \end{array}$$

$$\begin{array}{c} -5 \\ (1,3) \\ V^*(1,3) = 5 \end{array}$$

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