RUNI: Reinforcement Learning

MidTerm Project RL - Theory Part

Fall Semester 2022-23 TASHPAG

Submitted by: **Yarden Fogel ID: 011996279 (yarden.fogel@post.runi.ac.il)**
Partner: Sharon Koubi ID: 301315040 (sharon.koubi@post.runi.ac.il)
Date: January 16, 2023

Question 1: Dynamic Programming Policy Evaluation

Using Dynamic Programming, run (theoretically) Policy Evaluation and show what will be the states values at next iteration. Show your computations in details.

Question 1: Solution

Note: We are assuming $\gamma = 1$ since it wasn't given, and thus leaving it out of the below equations and solution, with the exception of the calculation of $v_1^{(1)}$, for the sake of illustration.

The values of the terminal states = 0 and do not change. Therefore, we can infer that:
$v_0^{(1)} = 0$ and $v_6^{(1)} = 0$ The remaining calculations are as follows:

$v_1^{(1)} = 0.5 * (0 + \gamma * v_0^{(0)}) + 0.5 * (0 + \gamma * v_2^{(0)}) = 0.5 * 0 + 0.5 * \frac{2}{6} = \frac{1}{6}$

$v_2^{(1)} = 0.5 * (0 + v_1^{(0)}) + 0.5 * (0 + v_3^{(0)}) = 0.5 * (0 + \frac{1}{6}) + 0.5 * (0 + \frac{3}{6}) = \frac{1}{12} + \frac{3}{12} = \frac{1}{3}$

$v_3^{(1)} = 0.5 * (0 + v_2^{(0)}) + 0.5 * (0 + v_4^{(0)}) = 0.5 * (0 + \frac{2}{6}) + 0.5 * (0 + \frac{4}{6}) = \frac{1}{6} + \frac{2}{6} = \frac{3}{6} = \frac{1}{2}$

$v_4^{(1)} = 0.5 * (0 + v_3^{(0)}) + 0.5 * (0 + v_5^{(0)}) = 0.5 * (0 + \frac{3}{6}) + 0.5 * (0 + \frac{4}{6}) = \frac{3}{12} + \frac{4}{12} = \frac{7}{12}$

$v_5^{(1)} = 0.5 * (0 + v_4^{(0)}) + 0.5 * (0 + v_6^{(0)}) = 0.5 * (0 + \frac{4}{6}) + 0.5 * (1 + 0) = \frac{2}{6} + \frac{3}{6} = \frac{5}{6}$

Question 2 on Next Page

Question 2: Grid World

0	-14	-20	-22
-14	-18	-20	-20
-20	-20	S2	-14
-22	S1	-14	0

In this $4x4$ grid world environment, we were partially given the "Values" for certain states (according to the current policy). Calculate (manually) the Values (according to the current policy) for states: $S1$, $S2$, Show your computations in details.

Question 2: Solution

S1:

$S1 = 0.25 * (-1 + (-22)) + 0.25 * (-1 + (-20)) + 0.25 * (-1 + (-14)) + 0.25 * (-1 + S1)$
$S1 = 0.25 * -23 + 0.25 * -21 + 0.25 * -15 + 0.25 * -1 + 0.25 * S1$
$S1 = -15 + 0.25 * S1$
$0.75 * S1 = -15 \implies S1 = -20$

S2:

$S2 = 0.25 * (-1 + (-20)) + 0.25 * (-1 + (-20)) + 0.25 * (-1 + (-14)) + +0.25 * (-1 + (-14))$
$S2 = 0.25 * -21 + 0.25 * -21 + 0.25 * -15 + 0.25 * -15$
$S2 = -5.25 + (-5.25) + (-3.75) + (-3.75)$
$\implies S2 = -18$

Compiled with LaTeX on January 16, 2023.