Not required to submit
Please write your own code and compare with the solution.

AV499 - AVD871 - Programming Assignment 6

Solution of a Markov decision process using policy iteration.

Suppose you are given an infinite horizon MDP with the following parameters: state space $\mathcal{S} = \{1, 2, 3\}$, action space $\mathcal{A} = \{1, 2\}$, reward $r(s,a) = s + a^2$, discount factor $\gamma = 0.7$.

The transition probability matrices $p^{(a)}$ are:

$$p^{(1)} = \begin{bmatrix} 0.1 & 0.1 & 0.8 \\ 0.2 & 0.3 & 0.5 \\ 0.8 & 0.1 & 0.1 \end{bmatrix}, \quad p^{(2)} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.6 & 0.2 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$

Solve the above MDP using policy iteration.