Value iteration for the infinite horizon Markov decision process (MDP)

An infinite horizon Markov decision process is specified by a state space S, action space A, transition probability matrices $p(a)$, reward function $r(s,a)$, and the criterion of maximising the total expected discounted reward, with discount factor γ. One method to solve the MDP is to use value iteration. In value iteration, we repeatedly apply the following recursion

$$V^n(s) = \max_a \{ r(s,a) + \gamma \cdot \sum_{s'} P_{ss'}^{(a)} V^0(s') \}$$

with $V^0(s)$ set to be $V^n(s)$ at the start of every step.

Implement the above recursion and solve the following MDP for the infinite horizon case. What is the condition that you use to stop the recursion and declare a solution?

$S = \{1, 2, 3\}$, $A = \{1, 2\}$, $r(s,a) = s + a^2$, $\gamma = 0.7$

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

Scanned by CamScanner

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.

For both of these tasks, you can use the provided code (which is in Matlab) to write your own code. Please note that the provided code is only a guideline and may not be a complete solution.