CS747 Assignment 2

-Ninad Chaphekar, 200010016.

Task1:

Here, I simply implement the Value Iteration, Howards Policy Iteration and Linear Program solver. (Direct analogous to the formula)

For Linear Program, we use pulp a standard library, and maximize our objective function to get our optimal value function.

Default algorithm to solve is Value Iteration, which has a threshold of 10^ (-10) to stop the iteration.

If policy is provided, it will just evaluate the policy and not compute the optimal policy/value function.

I observe that for small number of states, Linear Program seems to be the fastest but for big number of states, Value Iteration performs the best. Howard's Policy Iteration is relatively slow in both cases.

Task2:

I have implemented this using 2*balls*runs + 2 states. For each player A and B, there are of course balls*runs states possible, and the last two are for "won" and "lost" state, acting as terminal states as well.

Note: I have assumed that both A and B can perform the same actions, just that outcomes of B are much more limited. This has been done to maintain consistency for the MDP and its file.

To create the MDP file, we need the transition probabilities and rewards. The agent is given reward 1 only when it reaches the terminal state "won" and rest receives zero reward. Looping over possible states, we check which transition to another state is possible via a particular outcome, governed by the parameters given to us. Further, strike changed has been implemented whenever odd number of runs are scored given its not the last ball of the over, and whenever even runs are scored given it's the last ball of the over. Thus, a strike change involves a sudden jump in the series of changes in states, and is incorporated accordingly. Instances where multiple outcomes lead to the same next state have also been considered and their probabilities are added. Since this resembles an episodic task, it has an mdptype is episodic with a discount factor of 1. Multiple if/else conditions are there to involve these conditions, depending who is on strike, how many balls are left, what the next outcome/ next state can be, etc.