Programming Assignment 2

Hari Krishna Nama (170050077)

Value Iteration:

- Initiated Values with Zeroes
- Chose 10^{-12} as convergence condition for the change of values

Howard Policy Iteration:

- Initiated Values and policies with Zeroes
- Allowed 10^{-4} tolerance for the precision of values calculated
- Used Pulp to solve linear equations for policy value evaluation

Linear Programming:

- Used Pulp to solve linear inequations with objective function mentioned in slides

MDP Formulation:

- Each element of the grId is a state in the MDP, i.e., the number of states is equal to the number of the columns * number of the rows
- The number of actions is 4 and encoded as (North, 0), (South, 1), (East, 2), (West, 3)
- No transitions to any walls, Probability to all other possible transitions is 1
- Reward is 1,00,00,000 if it's a end state else -1
- MDP type is episodic since the maze problem terminates at some point
- Chose 0.9 as the discount factor

Linear Programming is faster than Value Iteration which is faster than Howard Policy Iteration when the number of the states in MDP is high