Name: Derek Kwon

Due Date: March 3rd 2024, 11:59 PM

Class: DATA 320 Reinforcement Learning

In a deterministic environment, the agent does not need to decide randomly as it has all the data it needs to predict the best outcome. A deterministic environment will always produce the same output for a given set of inputs. This results in the agent finding the shortest path to the end every time without fail, for both policy iteration and value iteration. The deterministic environment had 4 iterations for policy iteration and 112 iterations for value iteration.

The policy for the deterministic environment is: [1 2 1 0 1 0 1 0 2 1 1 0 0 2 2 0]

A deterministic policy will always output the same action given a particular state, which allows the agent to solve the maze and reach the end instantly without any errors.

In a stochastic environment, however, the agent must randomly choose its next action from its current state. This leads to the agent taking much longer to reach the end and appearing to be more confused since it may end up going backwards and going back to the start. Sometimes, the agent may not reach the end at all because it fell into a hole or exceeded 100 steps. It takes many more iterations of policy and value iteration for an agent in a stochastic environment to get the same result as an agent in a deterministic environment. The stochastic environment had 6 iterations for policy iteration and 368 iterations for value iteration.

The policy for the stochastic environment is: [0 3 0 3 0 0 0 0 3 1 0 0 0 2 1 0]

A stochastic policy will choose an action at random based on the probability distribution of the possible actions an agent can take at a given state. In the stochastic policy, there are more 0s and the presence of 3s.