AI3000/CS5500: Reinforcement Learning
Assignment 2

Taha Adeel Mohammed CS20BTECH11052

Problem 1: Value Iteration

(a) Prove that the Bellman optimality operator is a contraction under the max-norm. (5 points)

(b) Prove that the iterative policy evalution algorithm converges geometrically. (3 points)

(c) Let M be an infinite horizon MDP and V^* be its optimal value function. Suppose if the value iteration algorithm is terminated after $k+1$ iterations as $\|V_{k+1} - V_k\|_\infty < \epsilon$ for some chosen $\epsilon > 0$, how far is the estimate V_{k+1} from the optimal value function V^*? Provide details of your derivation. (5 points)

Problem 2: Programming Value Iteration

(a) Implement value iteration and policy iteration algorithm. Modularize the implementation to contain separate functions for policy evalution, value iteration and policy improvement. Test the implementation on the 4 x 4 Frozen Lake environment available in Gymnasium (formerly known as Open AI Gym).The stochastic version of the environment is default which can be changed by modifying the is-slippery flag. (8 points)

Reinforcement Learning algorithms are implemented in `rl.py` and the code for the Frozen Lake environment is in `frozen_lake.ipynb`.

(b) Document your findings with number of iterations needed for both algorithms to converge (or nearly converge) to optimal policy on the Frozen Lake envrionment. Further, provide a snapshot of the optimal policy obtained via the two algorithms. (4 points)

(c) Are there any stochastic optimal policies? If so, does any of the algorithms find any stochastic optimal policy? If not, why not? (3 points)

(d) Noisy Environment

(i) Implement the above environment in Python 3.8+. (8 points)

Implemented in `noisy_frozen_lake.ipynb`.

(ii) Use any of the DP algorithms implemented above on this environment and observe the optimal paths for various choices of γ and η. Identify what values of γ and η that could lead the agent to each of the optimal paths listed and explain the reasoning for the answer obtained. (8 points)

(iii) After solving this grid world example, please re-visit your answer to question 2(h) of Assignment 1 (1 point)