CS 5180: Exercise 0 Solutions

Harin Kumar Nallaguntla
Study Partner: Gopi Sainath Mamindlapalli

Q3. The manual policy, driven by human expertise and problem-solving, is inherently superior to the random policy in the context of this stochastic environment. Human-guided decision-making allows for the deliberate selection of actions that are likely to lead to favorable outcomes, optimizing the agent's path to the goal. In contrast, the random policy's reliance on arbitrary actions chosen without consideration of the state or prior experiences leads to inconsistent and often suboptimal results. Moreover, the manual policy has the potential for learning and adaptation over time, refining its strategy based on feedback and experience, while the random policy remains static, lacking the capacity to improve its performance. This fundamental difference in approach and capability underscores the manual policy's advantage in reliably achieving the desired goal compared to the random policy.

Moreover, the manual policy capitalizes on prior knowledge, goal-oriented decision-making, rapid problem-solving, simulation understanding, and the ability to develop general strategies based on an operator's understanding of the environment. These qualities enable the manual policy to expedite goal attainment, prioritize relevant actions, adapt to challenges, exploit insights into the simulator, and perform well across various scenarios. In contrast, the random policy relies solely on chance, lacking the cognitive advantages that make the manual policy more effective, thereby leading to a performance gap between the two policies.

Q4. Better Policy: Wavefront Planner

Strategy: The Wavefront Planner is a pathfinding algorithm that plans the optimal path from the current state to the goal state by systematically exploring the environment. It starts at the goal state and works backward, assigning increasing values (wavefront numbers) to states in a manner that ensures the agent always moves closer to the goal. The agent selects actions that lead to states with lower wavefront numbers, effectively navigating towards the goal.

Performance: This policy is better than the random policy because it employs an intelligent, systematic approach to reach the goal. It makes decisions based on a global understanding of the environment, minimizing the number of steps required to reach the goal. However, its performance may still be limited by the complexity of the environment and the accuracy of the wavefront planning algorithm.

Worse Policy: Always Go Up

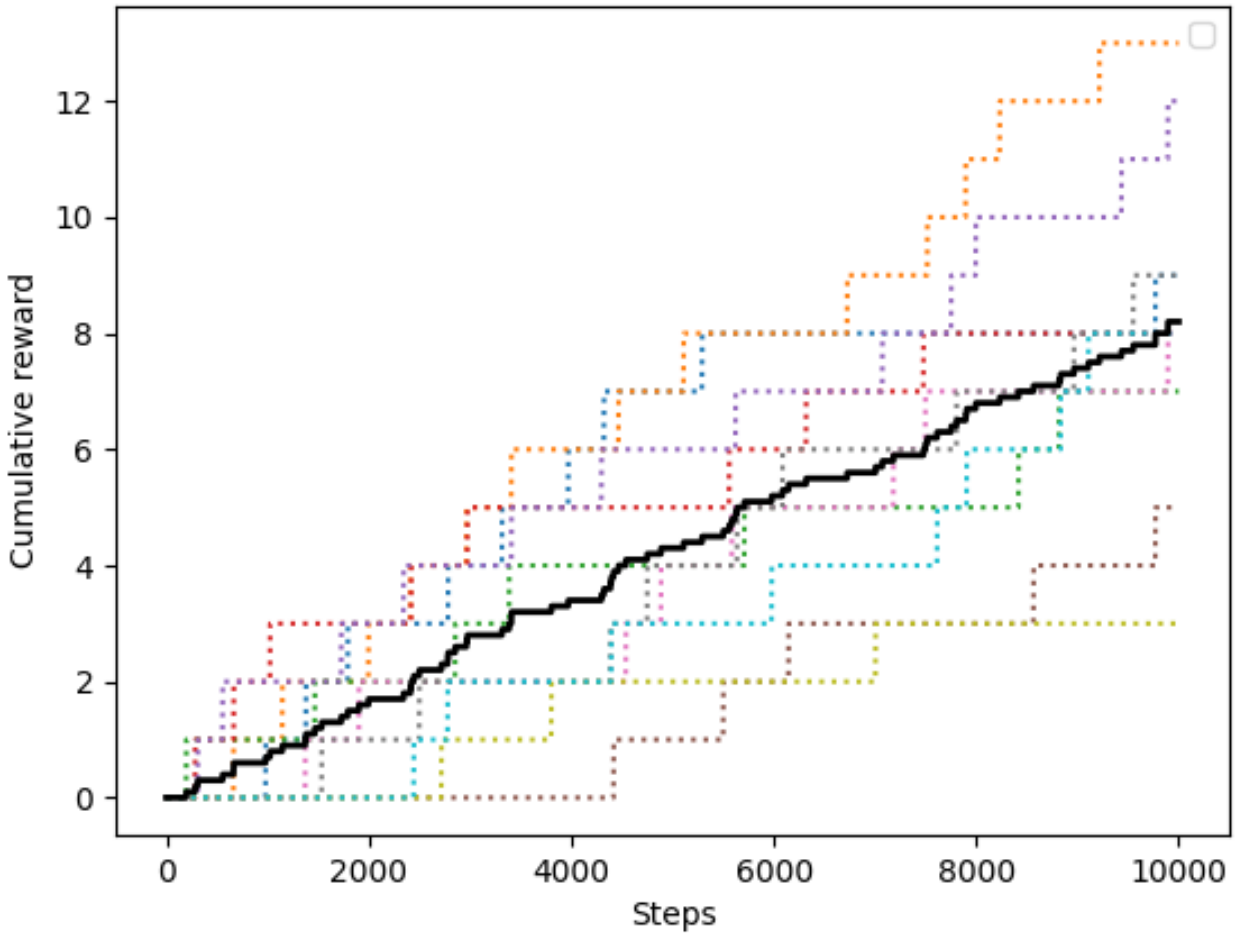
Strategy: The "Always Go Up" policy is a highly simplistic and suboptimal strategy that always selects the action to move upward, regardless of the current state or the position of

the goal. In this strategy, the agent consistently chooses an action that takes it further from the goal, making it unlikely to reach the goal effectively.

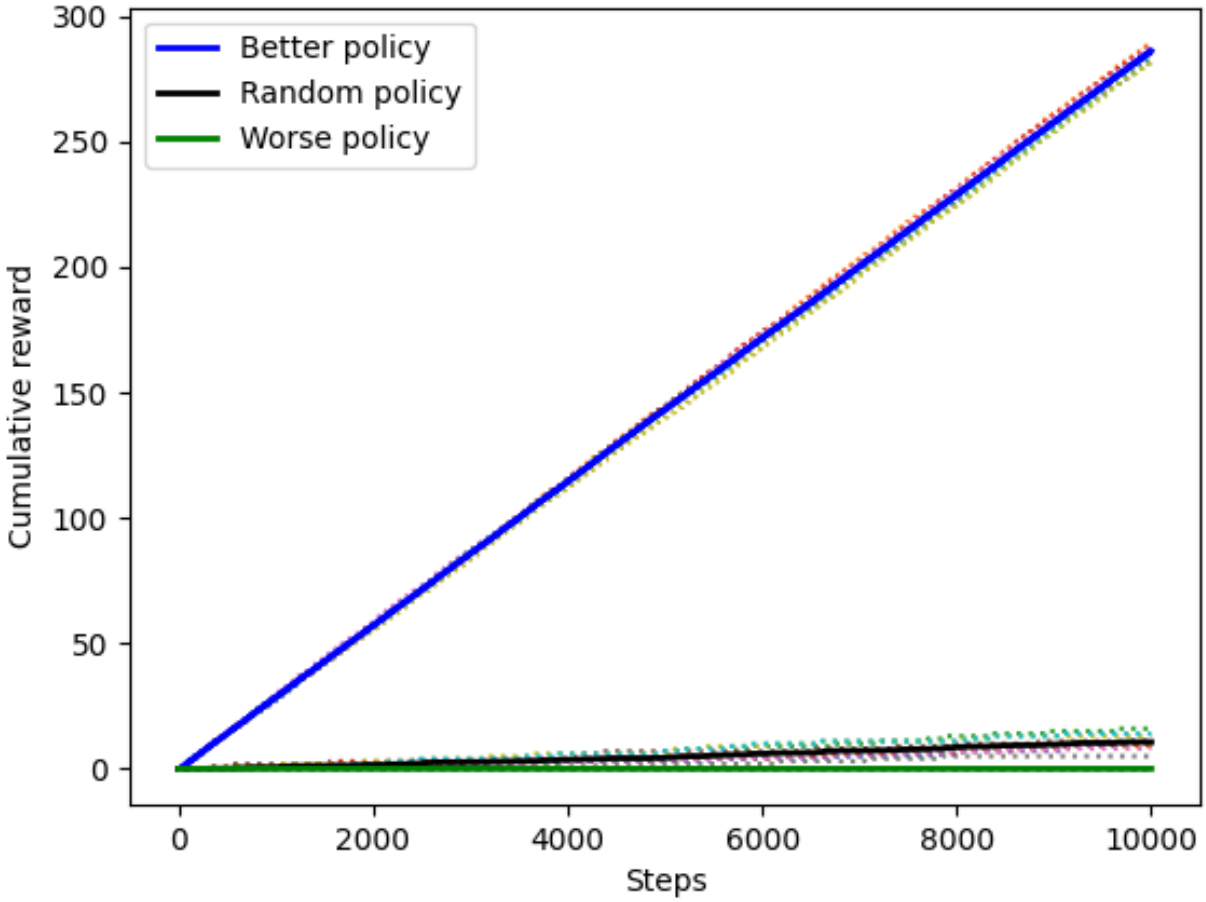
Performance: This policy is worse than the random policy because it completely lacks adaptability and strategic decision-making. By always moving upwards, it essentially ignores the state of the environment and the goal location. Consequently, it is highly likely to get stuck in local loops or move away from the goal, resulting in poor overall performance.

In summary, the "Wavefront Planner" policy outperforms the random policy due to its systematic pathfinding approach that leverages global environment knowledge, allowing it to make informed decisions to reach the goal efficiently. Conversely, the "Always Go Up" policy is worse than the random policy because it follows a rigid and counterproductive strategy, leading to inefficient navigation and a higher probability of failure to reach the goal. These policies highlight the critical role of intelligent decision-making and adaptability in reinforcement learning tasks.

Plots:



Q3. Plot of Cumulative reward vs Steps for Random policy for steps=10^4, trials=10



Q4. Plot of Cumulative reward vs Steps for Better policy, Random policy and Worse policy for steps=10^4, trials=10