

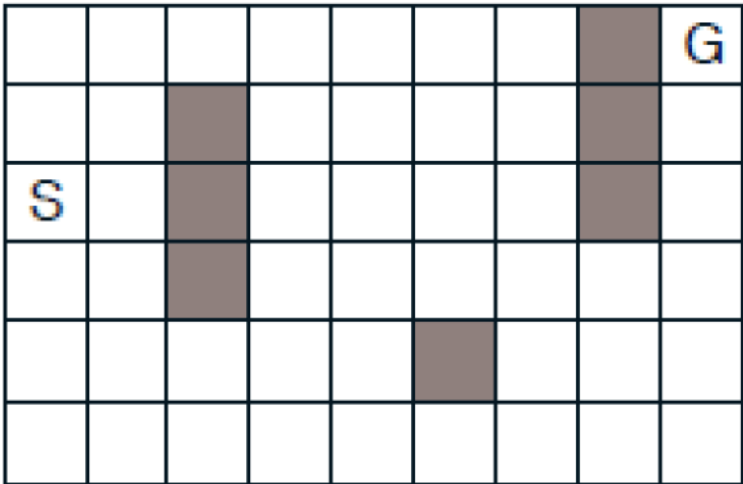
README.md

Dyna-Q

Solving Dyna Maze problem from example 8.1 of Richard S Sutton's RL Book using Dyna-Q algorithm

Input

The following 6X9 maze is used as environment.



actions

Hyper parameters

start = (2,0)

goal = (0,8)

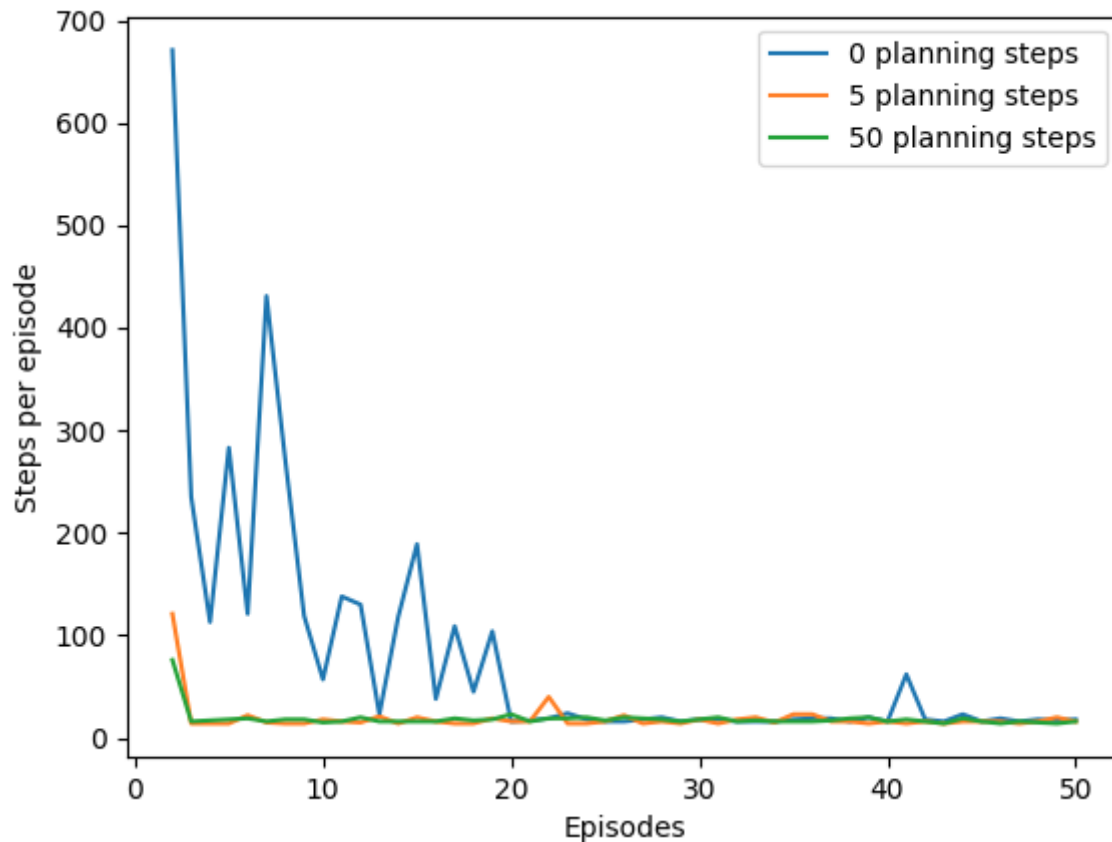
gamma = 0.95

alpha = 0.1

epsilon = 0.1

Output

With different number of planning steps, the number of steps required for each episode is given by the following figure.



The above figure shows learning curves from an experiment in which Dyna-Q agents were applied to the maze task. The initial action values were zero, the step-size parameter was $\alpha = 0.1$, and the exploration parameter was $\epsilon = 0.1$. When selecting greedily among actions, ties were broken randomly. The agents varied in the number of planning steps, n , they performed per real step. For each n , the curves show the number of steps taken by the agent to reach the goal in each episode. After the first episode, performance improved for all values of n , but much more rapidly for larger values. Recall that the $n = 0$ agent is a non-planning agent, using only direct reinforcement learning (one-step tabular Q-learning). This was by far the slowest agent on this problem, despite the fact that the parameter values were optimized for it. The non-planning agent took about 20 episodes to reach the optimal performance, whereas the $n = 5$ agent, and $n=50$ agent took only about 3 episodes.