RBE 595 – Reinforcement Learning

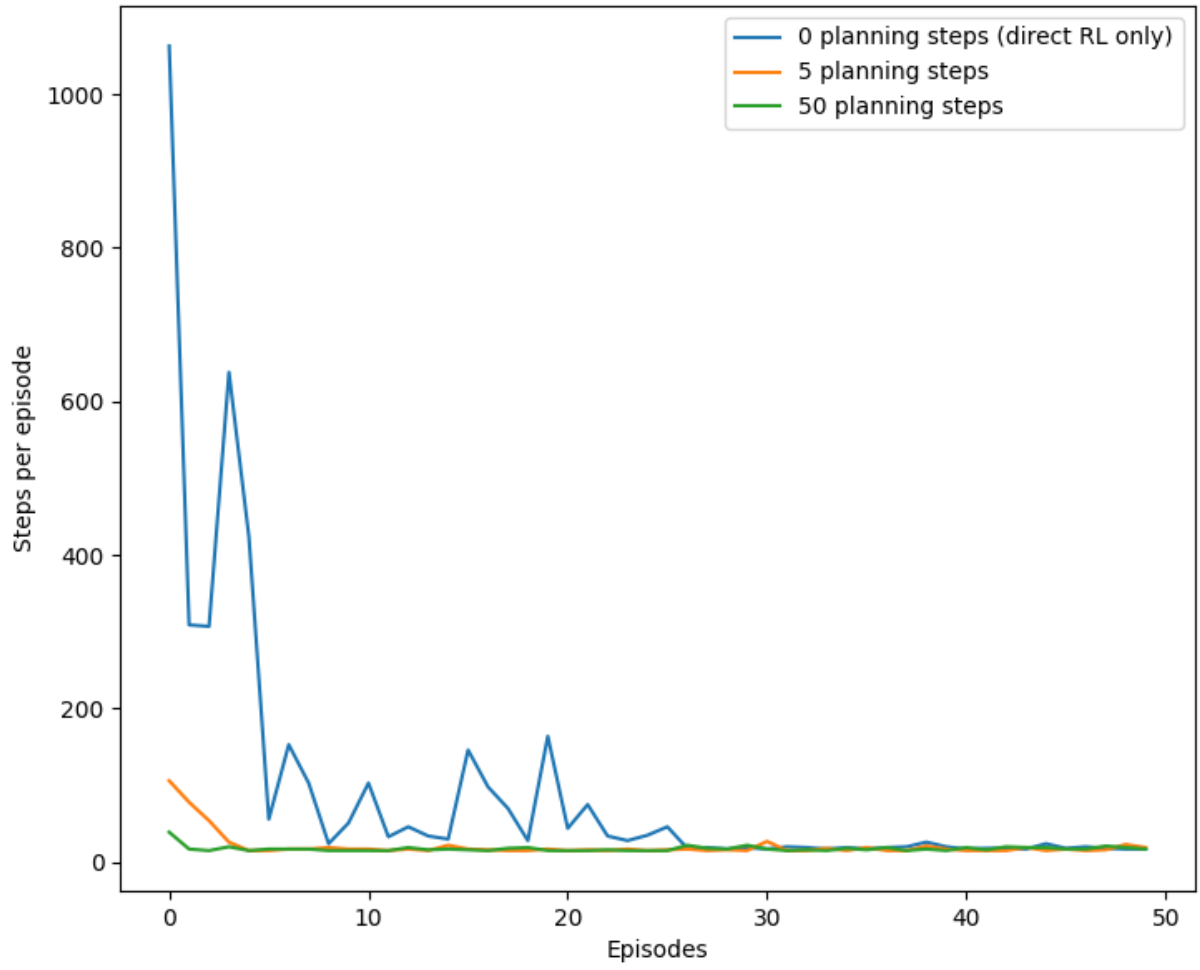
Programming Exercise #5

Model-Based RL

Authors: Taylor Bergeron and Arjan Gupta

RESULTS

We implemented TabularDynaQ, as given by the pseudo-code in the Barto & Sutton textbook. After running with 0, 5, and 50 planning steps, we plotted the results to obtain the following graph.



DISCUSSION

This graph is close to the book's Figure 8.2 and follows the same essential pattern. In particular, we find that when there are no planning steps used, the agent runs through a lot of iteration steps per episodes to reach the goal, which rapidly decreases over time. This is because it is not able to extract all the "juice" out of the episode to use the information to its advantage. With just 5 planning steps, we see a much less rapid change – within 5 episodes, the agent takes very few steps per episode. This is because it is successfully extracting a lot of information from the given episode. For 50 planning steps, we see close to 1 step per episode from the beginning.