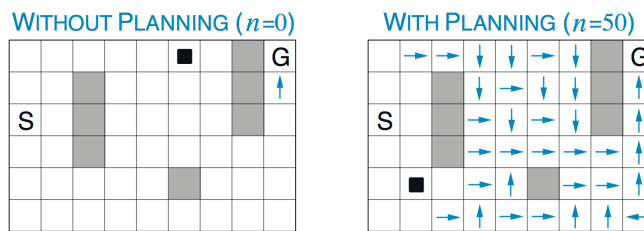


Reinforcement Learning (SS18) - Exercise 6

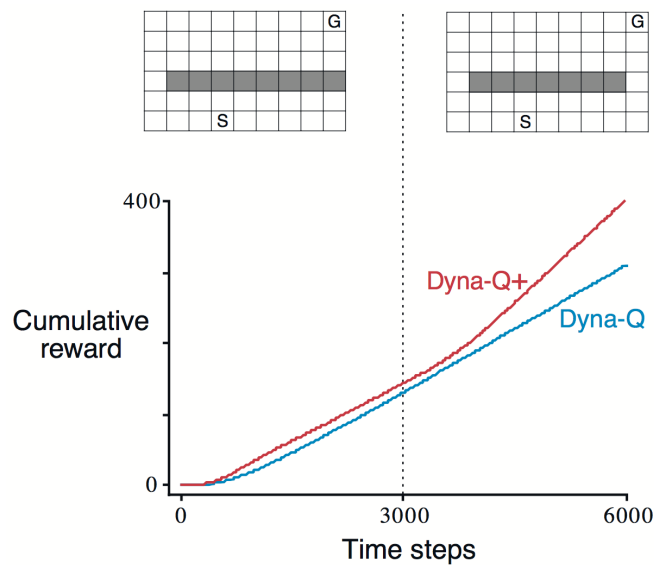
Daniel Hennes

14.06.2018 (due 20.06.2018)

1. Implement n -step Sarsa and evaluate it in the 8×8 *Frozen Lake* environment. Evaluate the performance for different choices of n and α , visualize your results (e.g., using the n -step TD evaluation for the 19-state random walk task as an inspiration).
2. The nonplanning method looks particularly poor in the figure below; a method using n -step bootstrapping would do better. Do you think one of the n -step bootstrapping methods could do as well as the Dyna-Q method? Explain why or why not.

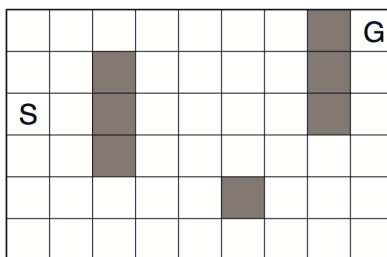


3. Why did the Dyna-Q+ (i.e., with exploration bonus) perform better in the first phase as well as in the second phase of the blocking and shortcut experiments (see figure below)?



4. (extra) Implement a general gridworld environment (example shown in the figure below). If you have chosen OpenAI gym for past exercises, you may want to implement your gridworld using the [gym environment interface](#).
- 4 actions (*west, south, east, north*)
 - deterministic or stochastic transitions
 - single or multiple terminal states
 - cost on each transition (step-cost) / reward to reach goal
 - obstacles
 - ...

Compare multi-step to one-step methods.



5. (extra) Implement and run Dyna-Q for deterministic gridworlds.