

Reinforcement Learning

Exercise 8

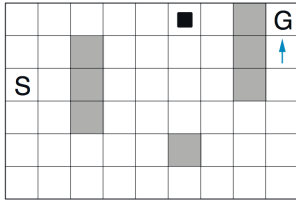
Mathias Niepert, Vinh Tong

June 21, 2024

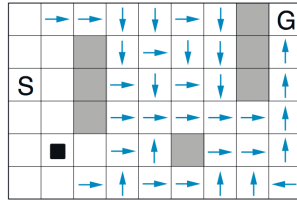
1 n-step and eligibility traces (4P)

a) 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. (2P)

WITHOUT PLANNING ($n=0$)



WITH PLANNING ($n=50$)



b) The λ -return is given by:

$$G_t^\lambda = (1 - \lambda) \sum_{n=1}^{\infty} \lambda^{n-1} G_{t:t+n}$$

with:

$$G_{t:t+n} = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots + \gamma^{n-1} R_{t+n} + \gamma^n V(S_{t+n})$$

Give the λ -return G_t^λ as recursive function in terms of the first reward R_{t+1} and itself one step later G_{t+1}^λ . (2P)

2 n-step Sarsa on the FrozenLake (5P)

The code template can be found on Ilias in `ex08-nstep/ex08-nstep.py`. It has been tested with gym version 0.18.0 (but should also be stable with version 0.18.3).

Implement n -step Sarsa and evaluate it on the 8×8 *Frozen Lake* environment. Evaluate the performance for different choices of n and α . Visualize your results (plot the performance over α for different choices of n , similar to lecture 8 slide 12).