

Reinforcement Learning

Exercise 8 - Solution

Jonathan Schnitzler - st166934

Eric Choquet - st160996

June 24, 2024

1 n-step TD compared to planing

The difference of the n-step temporal difference to the Dyna-Q planning, is that only the reward of the path which was taken can be accounted for. Therefore, unlike the image it is not possible after one episode to have a policy for all tiles, but instead only for the taken path. In contrast, Dyna-Q planning revisits arbitrary a virtual tile, which allows a richer interpretation.