

Exercise 6

Task 1 - Planning and Learning

- a) In Dyna-Q, in every step all $Q(s,a)$ are updated, which leads to a better performance.
- b) The cumulative reward for the Dyna-Q+ agent is higher since he will get a special bonus reward for testing all accessible state transaction (especially those who have not been considered in a long time).

Task 2 - n-step sarsa on the FrozenLake