Assignment 3

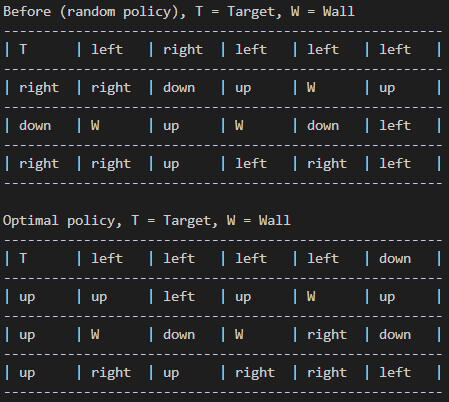
61075029H Jun Yu SHEN 沈峻宇

1. MC-Epsilon greedy and MC-Exploring
   1. MC-Exploring Start  
      Max iteration = 20000

Gamma = 0.9

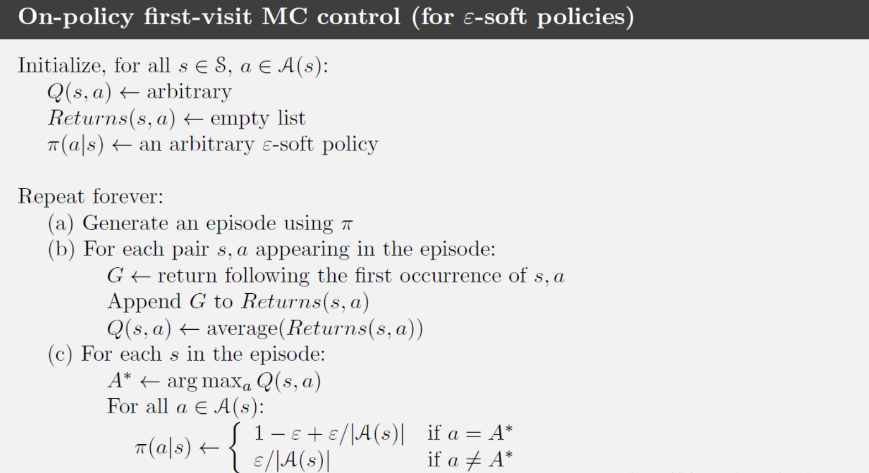
Horizon = 3

The results obtained using the above parameters are as follows



* 1. MC-Epsilon greedy without Exploring Starts(On-policy)

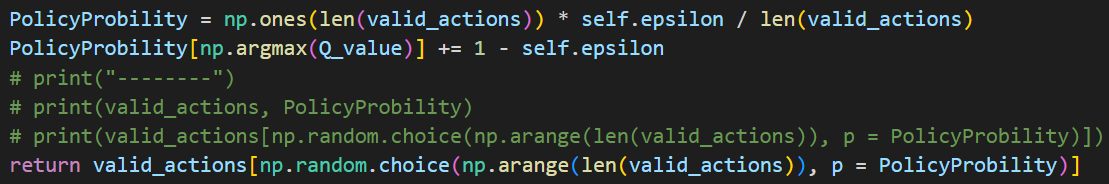
The only way to avoid the assumption that exploring starts is to ensure that all actions can be selected.



With probability ε, the current action with the largest action value estimate is selected, while with probability 1-ε, an action is randomly selected from all actions at random.

If there are multiple actions to choose from, you can use the following formula to calculate the probability and then select.

The code is shown in the figure below and is selected according to the odds calculated by epsilon:



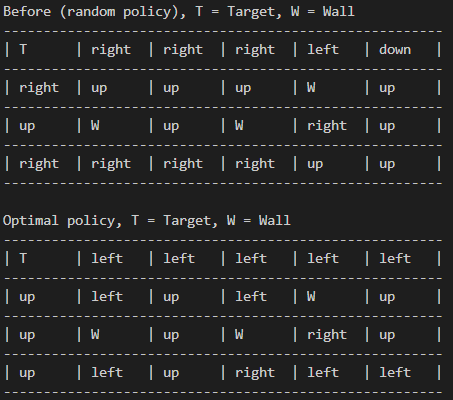
Max iteration = 20000

Gamma = 0.9

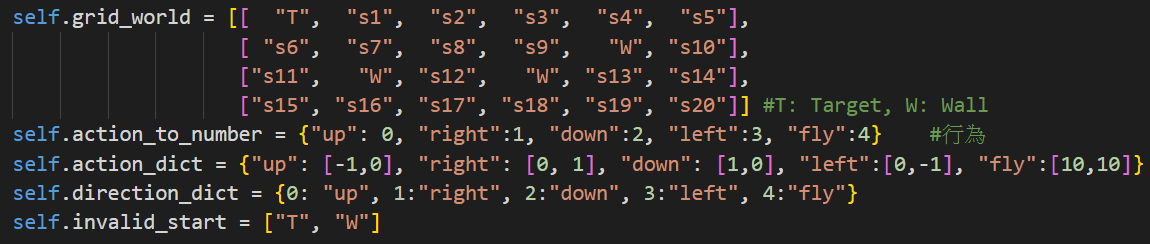
Horizon = 3

Epsilon = 0.2

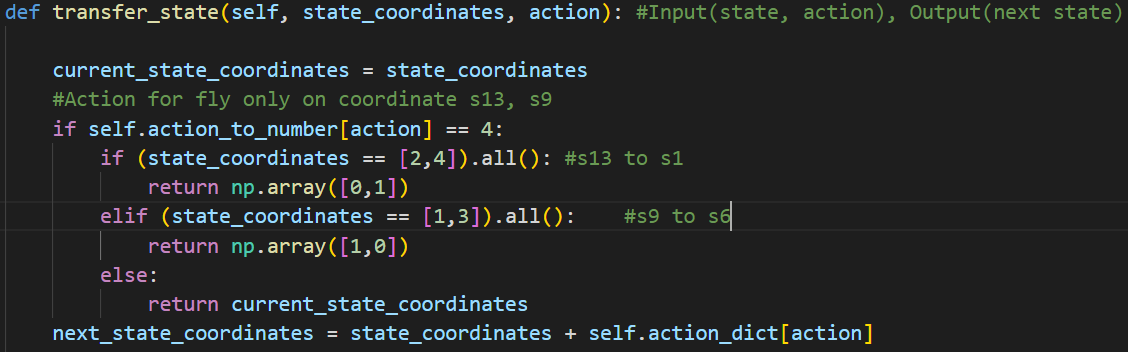
The results obtained using the above parameters are as follows:



By comparing the two results, we can see that in the state of small “Horizon”, **“without ExploringStarts”** is better because it can explore and exploration with epsilon.

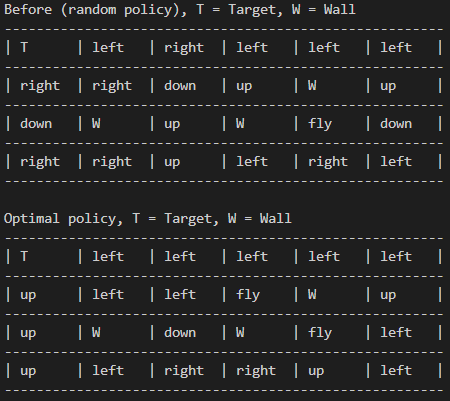
1. Fly~~
   1. Add new action fly

Movements for two special coordinates



* 1. Monte Carlo Algorithm addfly

Same parameters as the first question.



* 1. Monte Carlo Algorithm addfly without Exploring Starts

Same parameters as the first question.

