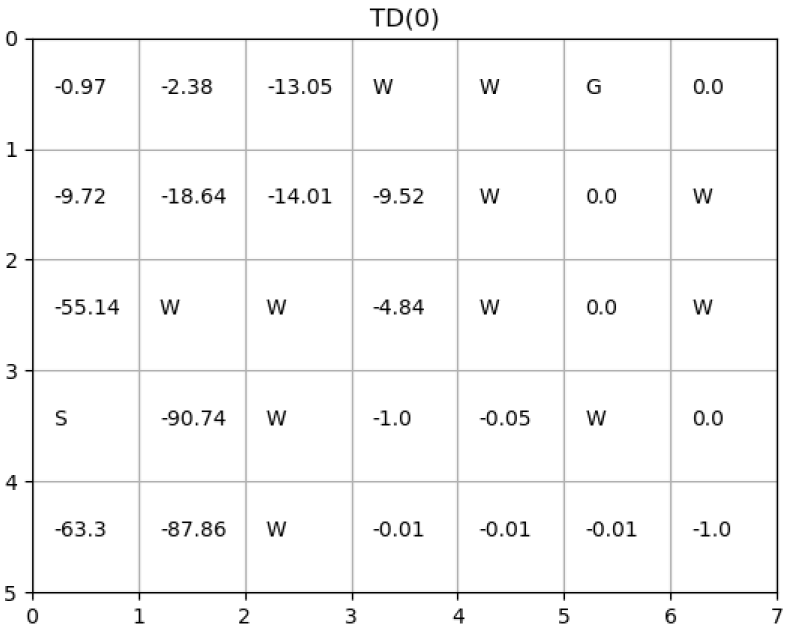
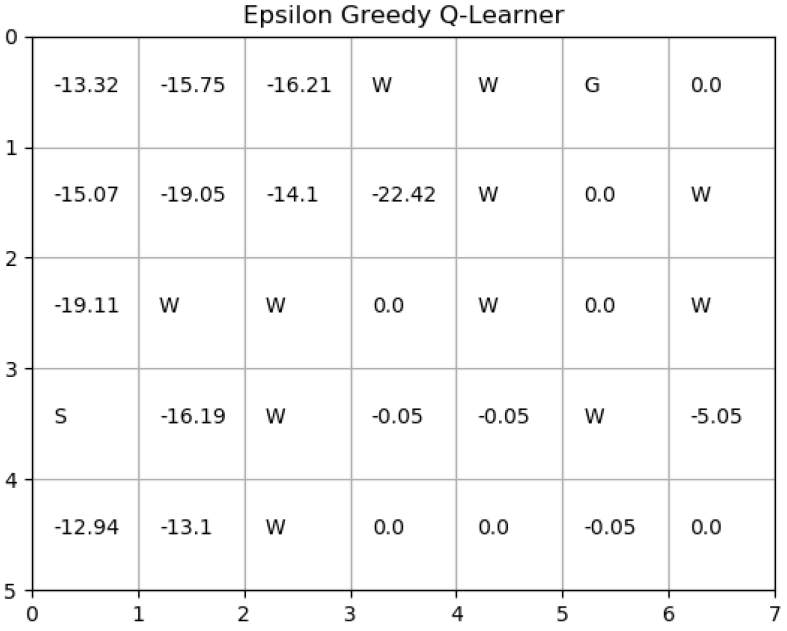
4.1 Solution

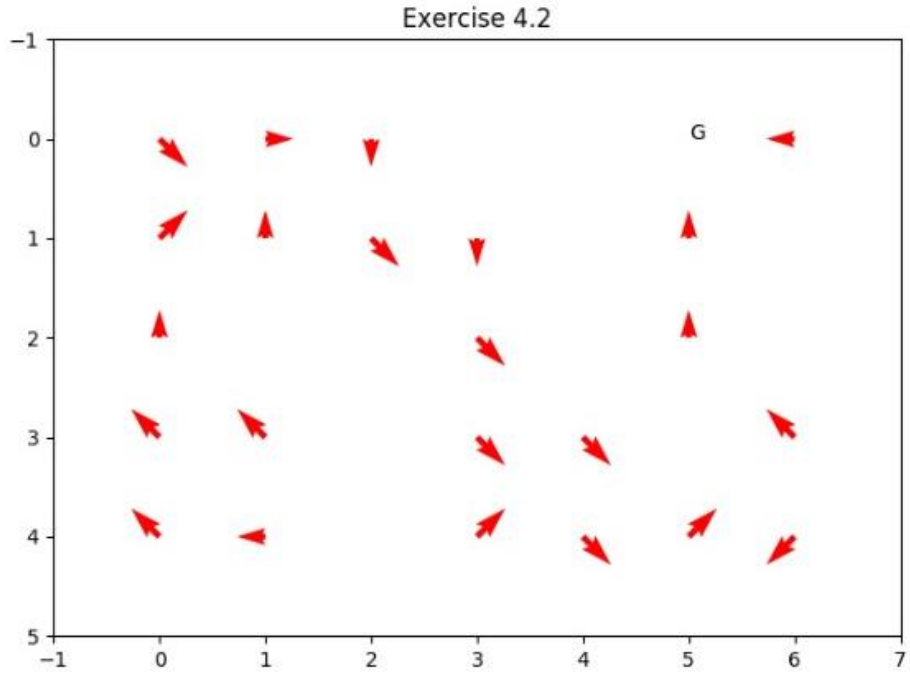


4.2 Solution

Value grid:



The optimal policy:



We got this policy which lookes meaningful and precise after a number of tries.

Note: As the policy selection is non-deterministic in both the cases, the value grid or the policy changes with every try and the solution (in case of 4.2) is not always optimal or meaningful. Also, the convergence of the algorithm depends on the parameters alpha and gamma. We have used the values for which we got the fastest convergence in the presented solution.