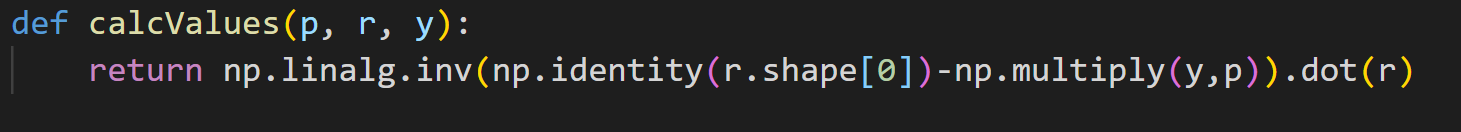
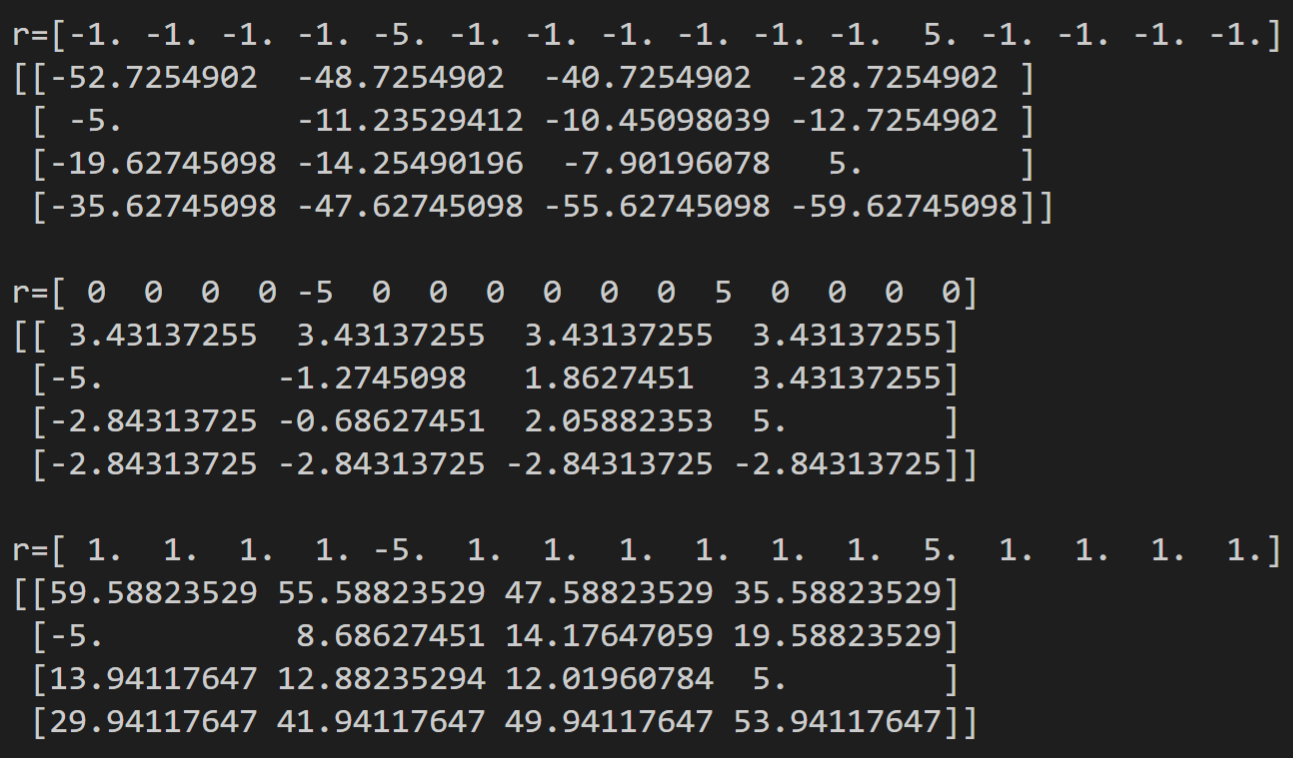
1. A.



Where ‘p’ is the probability matrix, ‘r’ is the reward list, ‘y’ is gamma.

We get the following results:

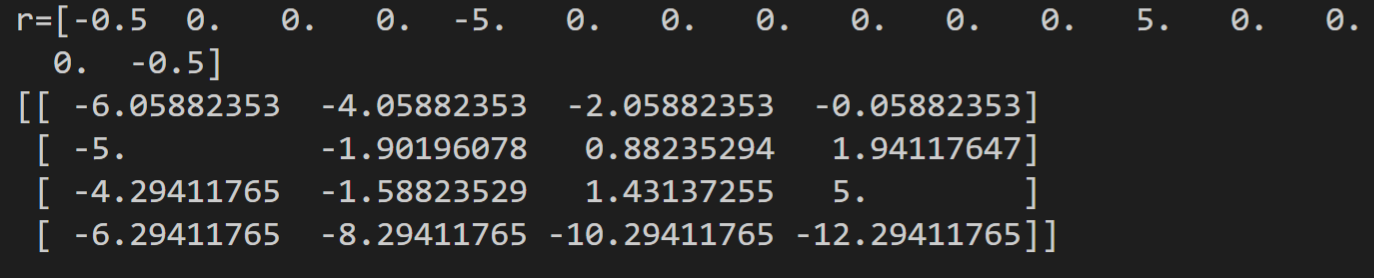


As you can see, will result in a policy that drives the agent to the green square. However, if the agent is in tile #9 will elect to go to tile #5 which will result in a negative outcome. Therefore, if the game starts on the bottom side of the board it will end negatively.

When , because our gamma is set to 1, our agent will become “long-sighted” and recognize that there is actually less reward for finishing the game. Therefore, it will avoid the goal state.

When there is no value gradient that would drive optimal policy improvement from either tiles (1,2,3,4) or (16,15,14,13).

If we assign individual reward values to the tiles we can have greater control over the policy creation. Setting results in a nice value function that drives agent action to the goal state regardless of starting position.



B.