CS 5180 – HW2

# Q1 – Formulating an MDP

1. The state space consists of all empty spaces in the four rooms environment. It is an 11 x 11 grid so it has 121 states, but only 104 states are reachable, the others are obstacles. Thus, the state space is

A black background with white text

Description automatically generated

The action space is the 4 set of actions the agent can take

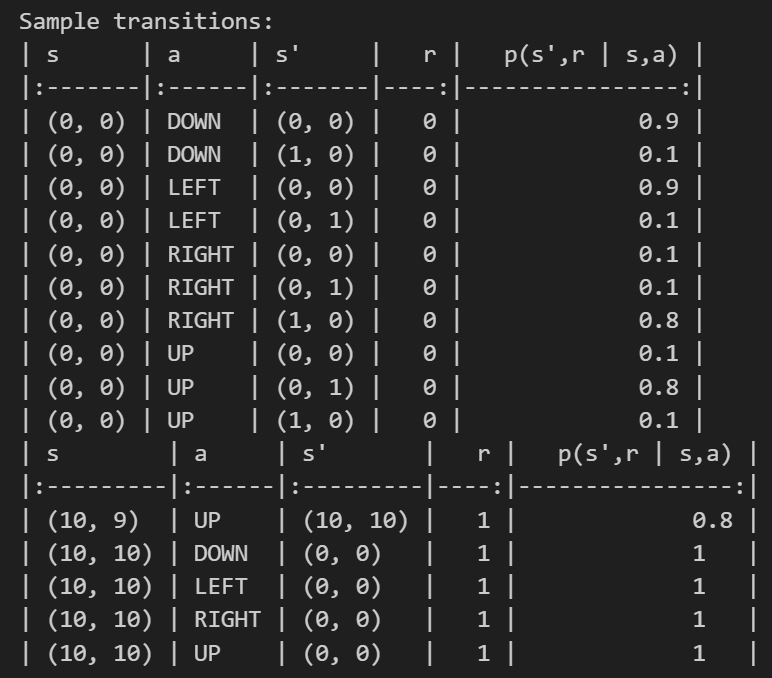
A black background with white letters

Description automatically generated

1. There are 104 valid states, each of those states can have 4 possible actions. So 104 x 4 =416 entries. However, the transitions are not deterministic, for each action selected, there are 3 possibilities, so the total goes up to 416 x 3 = **1248 entries possible.**

The actual table will possibly have less entries than that because the goal state is not considered and there will be some identical entries on the walls.

1. Please see the csv attached for full output. The first 10 and the last 5 entries are attached below.



# Q2 – Discounted Return

1. The discounted return is given by

A math equations on a black background

Description automatically generated

Given:

A math equation on a black background

Description automatically generated

Using the formula

G5 =

A black background with white text

Description automatically generated

G4=

A black background with white text

Description automatically generated

G3=

A black and white math equation

Description automatically generated

G2=



G1=

A black background with white numbers and symbols

Description automatically generated

G0=

A black background with white numbers and symbols

Description automatically generated

1. The same formula is used for the infinite sequence

Given:

A black background with white text

Description automatically generated

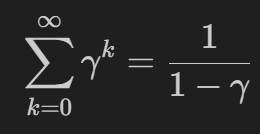
G1=

A black background with white text

Description automatically generated

Since R2 onwards is 8, we can take the 8 outside the summation

Using the geometric series formula:

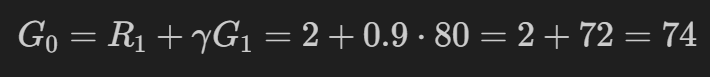


Therefore,

A number and equation on a black background

Description automatically generated

And G0=



# Q3 – The RL Objective

Based on the equation 3.7

A white letter on a black background

Description automatically generated

The return, irrespective of what path the agent takes would be 1. RT = 1 is the only term in that equation. This is the sparse rewards case, thus the agent does not really have the information it needs to get the optimal path. If we either introduce a -1 reward for states that are not the goal states, or implement a discounted return for all states that are not the goal state. The agent will have incentive and information to find the optimal path for escaping from the maze. Thus, the issue in this case is simply that there is no reward data to optimize the path. If we introduce something like that, the optimal path will be found.

# Q4 – Modifying the reward function

The signs of the reward are not important. Because the rewards play a role in decision making, as long as undesirable states have a smaller reward than desirable states, the agent will learn to maximize its return by reaching the desirable (or goal) states. Thus the intervals between the rewards is important.

Consider the bellman equation for the value function

A black background with white text

Description automatically generated

Adding constant c to the rewards

A black background with white text

Description automatically generated

Solving the second term using infinite geometric series

A black background with white text

Description automatically generated

Notice that the second term is a constant. Thus, adding a constant offset to the rewards just shifts the value function by a constant vc.

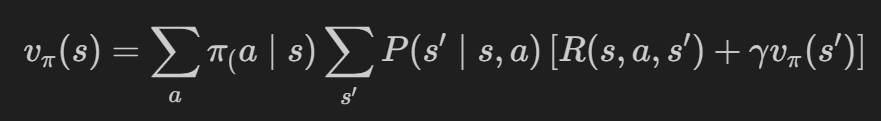
The value of vc is:

A black background with white text

Description automatically generated

# Q5 – Bellman equation

1. Using the following Bellman equation

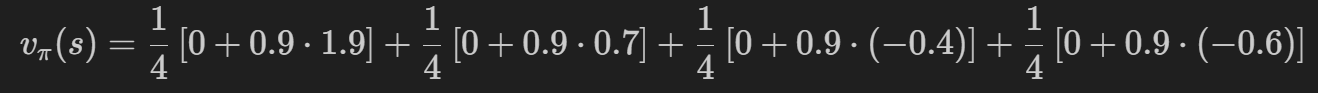


Given that

A black background with white text

Description automatically generated

Substituting in the equation

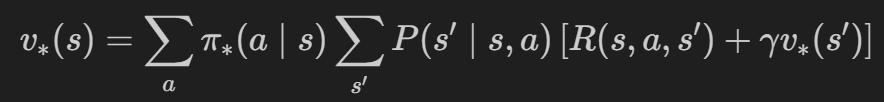


A number on a black background

Description automatically generated

Thus, the result is verified for approximation to one decimal point.

1. For the optimal policy,



According to the figure, the state we are looking at has equiprobable chance to go in two directions, LEFT and UP.

Thus,

A black background with white text

Description automatically generated

Substituting in the equation

A black background with white numbers

Description automatically generated

A black background with white numbers

Description automatically generated

Thus, the result is verified for approximation to one decimal point.

# Q6 – Solving for the value function

1. A piece of paper with writing on it

   Description automatically generated

Attaching the final equations using latex, incase my handwriting is not legible

A black background with white text

Description automatically generated

A black background with white text

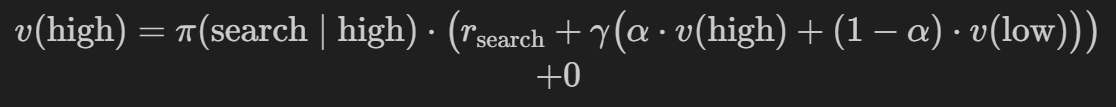
Description automatically generated

1. Given that

A black background with white text

Description automatically generated

For high state, policy always searches. Therefore,



A black background with white text

Description automatically generated

Similarly, for low state, The policy only waits or recharges. Therefore equation becomes

A black background with white text

Description automatically generated

A black background with white text

Description automatically generated

Solving the two equations simultaneously we get:

A black background with white text

Description automatically generated

Substituting in step 1 of both bellman equations, The equations are satisfied.

For the v(high)

A black background with white numbers

Description automatically generated

For the v(low)

A screenshot of a black background

Description automatically generated