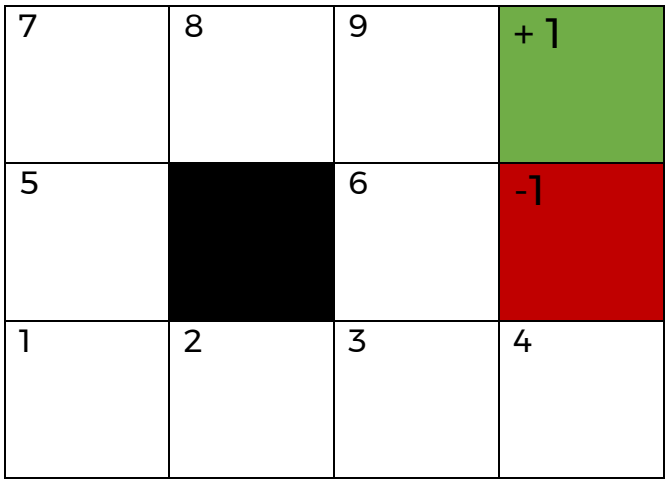
CMSC389F Problem Set 5

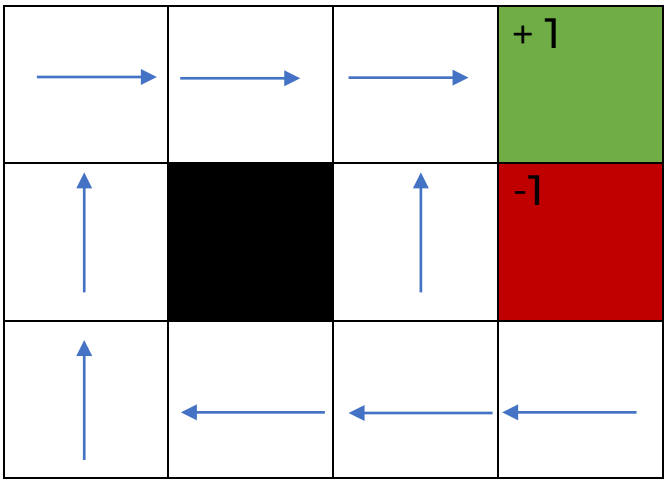
Environment Details:

- Reward: -0.04 for every action. Discount Factor (Gamma) of 1.
- If you take an action, there is an 80% chance of going in the desired direction, 10% chance of going left in relation to the desired direction, 10% chance of going right in relation to the desired direction. (Ex: if you are intending to move right, you have an 80% of going right, 10% chance of going up since up is left relative to moving right, and 10% chance of going down).
- Policy is deterministic: given a state, it returns a single action to take
- The black cell is a wall, it is not a state you can visit
- There are two terminal states: a green state with reward +1 and a red state with -1 reward
- Each state is numbered, shown in the first figure. These numbers do not represent reward.

Gridworld



Policy



State Action Values

0.812	0.868	0.918	+ 1
0.762		0.660	- 1
0.705	0.655	0.611	0.388

Recall that Q(state, action) refers to the action value function, and V(state) refers to the state value function. Please refer to the Lecture 5 Notes on the class webpage for a refresher on the equations and what they mean.

The following questions are for value functions following the policy outlined above. Notation: V(3) signifies the State Value of State 3.

Gridworld Questions:

1. Validate that V(7) is equivalent to 0.812 by showing the calculations using the value functions
2. Validate that V(6) is equivalent to 0.660 by showing the calculations using the value functions
3. Calculate Q(9, right)
4. Calculate Q(9, left)
5. Calculate Q(6, up)
6. Imagine if our policy was not deterministic, but instead stochastic. At state 3, the policy will tell you to go up 50% of the time, and left 50% of the time. What would the state value for state 3, V(3), be?

True or False

_____ Value Functions cannot be defined without first defining a policy

_____ Both state value function and action value function can be defined in terms of each other

_____ You can have a deterministic policy within a stochastic environment (and vice versa)

Intuition for Next Lecture:
While it may not be immediately useful to see how value functions can be useful when we are already given the state values (as in this example), we will see next lecture how we can come up with these values ourselves. This is done through a process called **Value Iteration**.