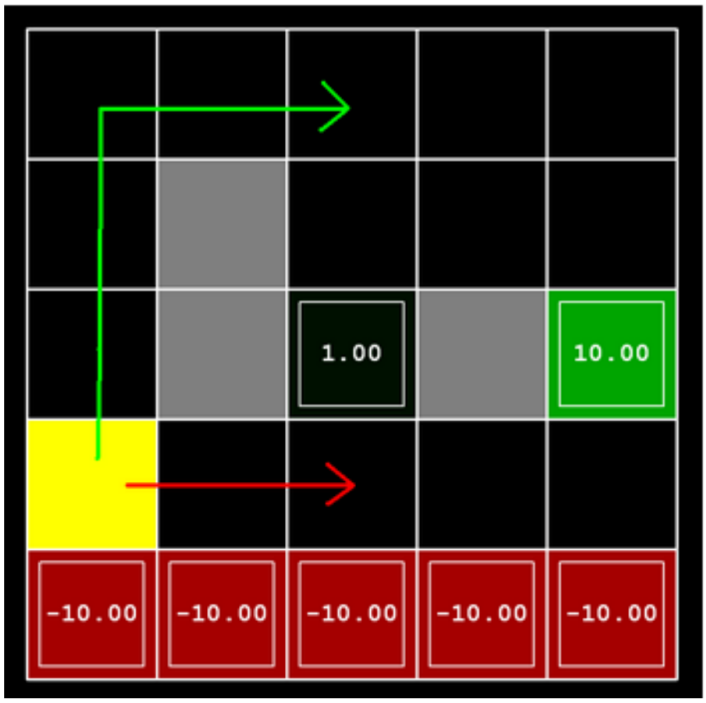
Problem 1

Explain how finding the shortest path in an acyclic weighted directed graph can be encoded into a Markov Decision Process. In other words, what are the states and actions, and how should you define the transition probabilities and rewards so that the optimal solution to the MDP will be the shortest path in the graph? Should there be any discount factor?

Problem 2

Consider the following grid layout. Imagine an agent starting from the yellow square. There are two terminal states with positive payoffs, a close exit with payoff +1 and a distant exit with payoff +10. The last row of this grid resemble a “cliff” region. Each terminal state in this region has payoff -10. The grey color indicates an obstacle. You would stay in the same state if you move toward an obstacle or the wall. All the other squares represents non-terminal states. We distinguish between two types of paths:

1. Paths that "risk the cliff" and travel near the bottom row of the grid; these paths are shorter but risk earning a large negative payoff, and are represented by the red arrow in the figure below.
2. Paths that "avoid the cliff" and travel along the top edge of the grid. These paths are longer but are less likely to incur huge negative payoffs. These paths are represented by the green arrow in the figure below.



You need to implement the following Matlab function:

```
function [optimal_policy]=find_the_optimal_policy(discount, livingReward, noise )
```

Where discount is the discount factor; livingReward is the immediate transition reward to any non-terminal state; and, noise denotes how often the agent ends up in an unintended successor state when it

performs an action. For example, with noise=0.2 and a move toward north, the agent would end up in the top successor state 80% of the time, 10% (noise×100/2) in the right, and 10% in the left state.

Part 1

Return the optimal policy for the aforementioned grid with the following parameters:

- discount = 0.9
- livingReward= 0.0
- noise = 0.2

The optimal policy is a 5×5 matrix which shows the direction to move at each state. There are four possible directions: east, north, west, and south. We want the directions to be represented by numbers in the following way:

- east is represented by 1
- north is represented by 2
- west is represented by 3
- south is represented by 4

Part 2

Modify the parameters = {discount, livingReward, noise} to achieve the following behaviors. What should be the values for these parameters in each case? If a behavior is not possible, say "It is impossible".

(a) Prefer the close exit (+1), risking the cliff (-10)
(b) Prefer the close exit (+1), but avoiding the cliff (-10)
(c) Prefer the distant exit (+10), risking the cliff (-10)
(d) Prefer the distant exit (+10), avoiding the cliff (-10)

Rules

- Your solution should be submitted through MyCourses portal.
- The deadline for the assignment is October 19.
- You need to submit two files, `find_the_optimal_policy.m` containing your solution and assignment2.pdf reporting the answers to the first problem and part 2 of the second problem.
- In the report, mention whether you are confident that your code returns the true optimal policy. Remember that you will earn a **negative grade (-2)** for the **wrong policy which you are sure is correct**. If you are not sure and your code is wrong, you won't lose any grade. On the other hand, if you are **sure** and your code is correct, **you will earn a +1 positive reward**.
- By submitting your solution, you affirm that you have developed the solution yourself.
- You are allowed to discuss the task with other students but you are not allowed to co-operate beyond discussion.

You can contact Murtaza Hazara (Murtaza.Hazara@aalto.fi) for clarifying the assignment or if you need support.

Grading

Maximum 20 points

- +5 points for the first problem
- +7 points for the Matlab function
- +3 points for correct optimal policy
- +1 for correct optimal policy of which you are confident
- -2 points for incorrect policy which you are confident about its correctness
- +4 for the part b of the second problem (+1 point for correct answer to each section)
- -1 point per each day the deadline is exceeded