Robot Learning	**Assignment 1**
CS 4756 Spring 2026	**Due 11:59pm Thursday, February 19**

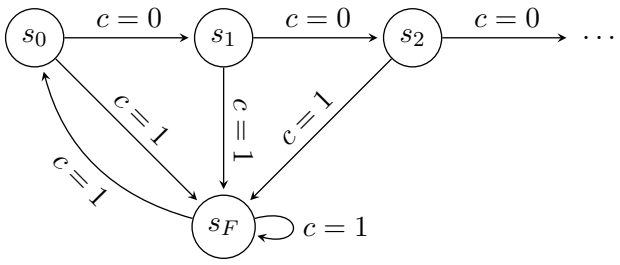
(1) Cliff Recovery: Behavior Cloning and DAgger (10 points)

In this question, we are going to be thinking about robots falling off of cliffs and trying to get back on. We will compare how well behavior cloning (BC) performs compared to DAgger.

We consider a Cliff MDP, where the robot (a.k.a. the learner) begins at state s_0. There is a path on the cliff consisting of infinitely many "safe" states. From each of these safe states, the robot can also fall off the cliff and land at the bottom, which we denote as reaching state s_F. From the bottom of the cliff s_F, there is a recovery action that brings the robot back to s_0, and another that indicates the robot stays there.

In class, we learned that BC and DAgger are trained using demonstrations from an expert. Remember that the expert will always follow an optimal trajectory, taking actions incurring the least cost at each state. In this setting, at each state the learner (both BC and DAgger) visits that the expert would also visit (i.e. all the safe states), the learner will make a mistake with probability ϵ and fall off the cliff. Once the the BC learner falls off the cliff, it does not know how to recover. In contrast, the DAgger learner will query the expert to determine the optimal action, allowing it to recover from the ditch back onto the cliff with probability 1.

We will consider an infinite horizon setting with a discount factor of γ.



Express the following quantities in terms of only ϵ, γ and constants.

- $J(\pi_{BC})$**: Expected total discounted sum of costs for Behavior Cloning.**
- $J(\pi_{DAgger})$**: Expected total discounted sum of costs for DAgger.**

Important tips:

- Try formulating two mutually recursive equations for $J_{\text{cliff}}(\pi)$ and $J_{\text{ditch}}(\pi)$, the expected total discounted sum of costs when the learner either starts on the cliff or starts in the ditch, respectively. In other words, you can write $J_{\text{cliff}}(\pi)$ and $J_{\text{ditch}}(\pi)$ as functions of themselves and each other. Think about how these relate to $J(\pi)$.

(2) Exploring Markov Decision Processes (5 points)

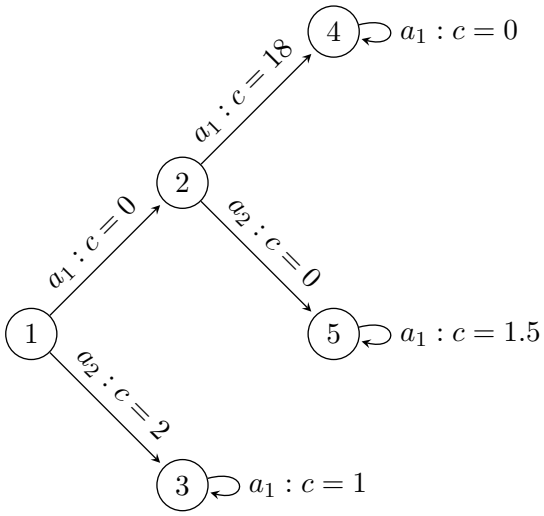


Figure 1: MDP for Problem 2

Compute the optimal value function V^* and the corresponding optimal policy π^* for each state in Fig. 1 for a discount factor of $\gamma = 0.9$ in the infinite horizon setting.

Notes:

- Initial State is always State 1.
- Each edge of the MDP is labeled in the following format: "{action} : {cost of action to complete transition}". Thus, the problem formulation involves a minimization of cost, rather than a maximization of a reward as may be seen elsewhere.
- Action a_1 at states 3, 4, and 5 must be taken infinitely if those states are ever reached.