# COMP4200 / COMP 5430 AI: Homework V
# Markov Decision Processes

UMASS - LOWELL

Name: ______________________________

Student ID: ______________________________

| Question | Points |
|---|---|
| **1** | 13 |
| **2** | 12 |
| **3** | 10 |
| **Total** | |

**Instructions:**

1. This examination contains 6 pages, including this page.
2. Write your answers in this booklet. If you must write on the back page, please indicate **very** clearly on the front of the page that you have written on the back of the page.
3. You **may** use any resources, including lecture notes, books, other students or other engineers, but you should provide a reference.
4. You may use a calculator. You may not share a calculator with anyone.

## Question 1: MDP: Fight or Run

[12 pts] A boy is being chased around the school yard by bullies and must choose whether to Fight or Run.

a. There are three states:
   - Ok (O), where he is fine for the moment.
   - Danger (D), where the bullies are right on his heels.
   - Caught (C), where the bullies catch up with him and administer noogies.

b. He begins in state O 75

c. He begins in state D 25

The graph of the MDP is given here:

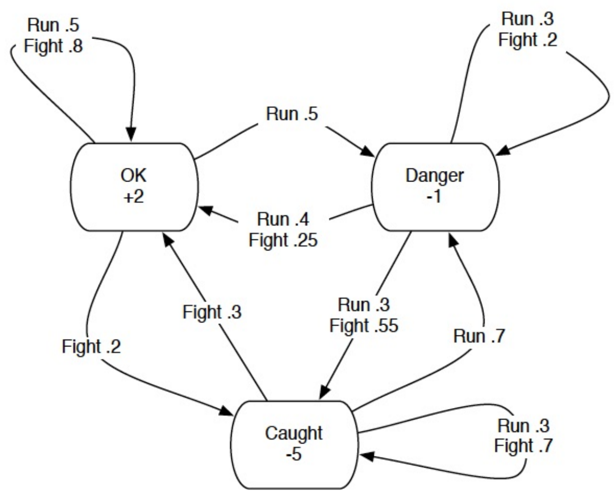


Figure 1: MDP: implausible robot

(a) (6 pts) Fill out the table with the results of value iteration at t =2 with a discount factor $\gamma = 0.9$. Note; to calculate values at t =1, I used initial values as 0.

| t | $J^t(O)$ | $J^t(D)$ | $J^t(C)$ |
|---|---|---|---|
| 1 | 2 | $-1$ | $-5$ |
| 2 | | | |

(b) (6 points) At t = 2 with $\gamma = 0.9$, what policy would you select? Is it necessarily true that this is the optimal policy? At t = 3 what policy would you select? Is it necessarily true that this is the optimal policy??

## Question 2: MDPs: Implausible robot

[13 pts] You are a wildly implausible robot who wanders among the four areas depicted below. You hate rain and get a reward of -30 on any move that starts in the deck and -40 on any move that starts in the Garden. You like parties, and you are indifferent to kitchens

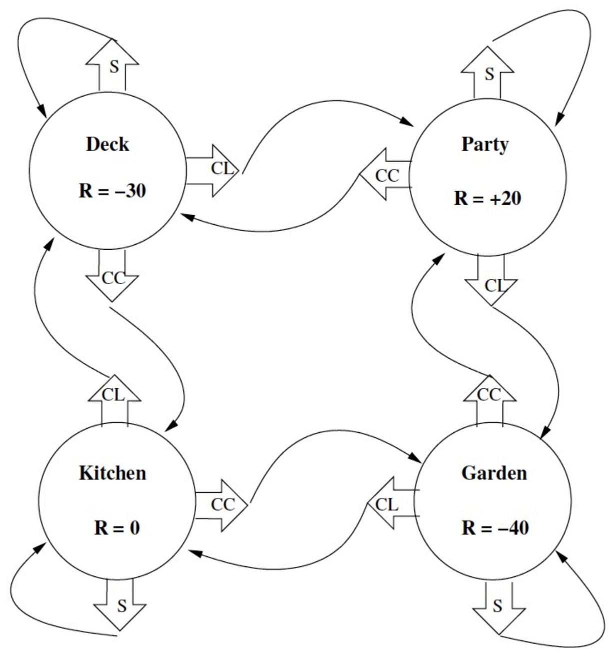


Figure 2: MDP: implausible robot

Actions: All states have three actions: Clockwise (CL), Counter-Clockwise (CC), Stay (S). Clockwise and Counter-Clockwise move you through a door into another room, and Stay keeps you in the same location. All transitions have are deterministic (probability 1.0).

(a) (3 pts) How many distinct policies are there for this MDP?

(b) (5 pts) Let $J^*($ Room$)=$ expected discounted sum of future rewards assuming you start in "Room" and subsequently act optimally. Assuming a discount factor $\gamma$=0.5 , give the $J^*$ values for each room.

(c) (7 pts) The optimal policy when the discount factor, $\gamma$ , is small but non-zero (e.g. $\gamma = 0.1$) different from the optimal policy when is large e.g. $\gamma = 0.9$). If we began with $\gamma = 0.1$ , and then gradually increased , what would be the threshold value of above which the optimal policy would change?

## Question 3: MDP: Solving Bellman Equations

[10 pts]

(a) (5 points) Suppose you have a robot trying to reach a goal and avoid cliffs in a small grid world. It can only move North, South, East, or West, but occasionally fails to move in the intended direction. If you were to model this using an MDP and were trying to solve it optimally, should you use value iteration or policy iteration? Justify your answer.

(b) (5 points) Now suppose that the robot can teleport to any grid cell but the teleportation causes it to land in neighboring grid cells near the target with some probability. Of you were to model this using an MDP and were trying to solve it optimally should you use value iteration or policy iteration? Justify your answer.