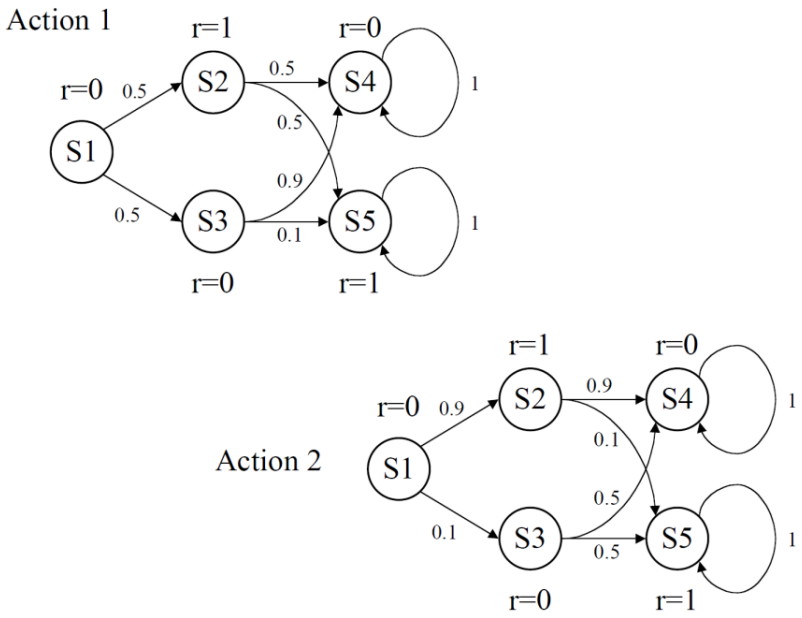
COEN 266 Artificial Intelligence

Homework #6

Guideline: Please complete the following problems and submit the answers as a single PDF file to Camino.

Problem 1. MDP.



Calculate the optimal values of the states in the above MDP by value iteration, assuming $\gamma = 0.9$, and the initial values of the states are zeros. Let the stopping criterion threshold be $\epsilon = 10^{-4}$. Describe how you solve the problem, with proper math expressions, then give the final result (optimal state values).

Problem 2. MDP.

Consider the following problem: Consider a rover that operates on a slope and uses solar panels to recharge. It can be in one of three states: high, medium and low on the slope. If it spins its wheels, it climbs the slope in each time step (from low to medium or from medium to high) or stays high. If it does not spin its wheels, it slides down the slope in each time step (from high to medium or from medium to low) or stays low. Spinning its wheels uses one unit of energy per time step. Being high or medium on the slope gains three units of energy per time step via the solar panels, while being low on the slope does not gain any energy per time step. The robot wants to gain as much energy as possible.

a) Draw the MDP graphically. Use arrows to represent state transitions, specify the action and the reward on each arrow.

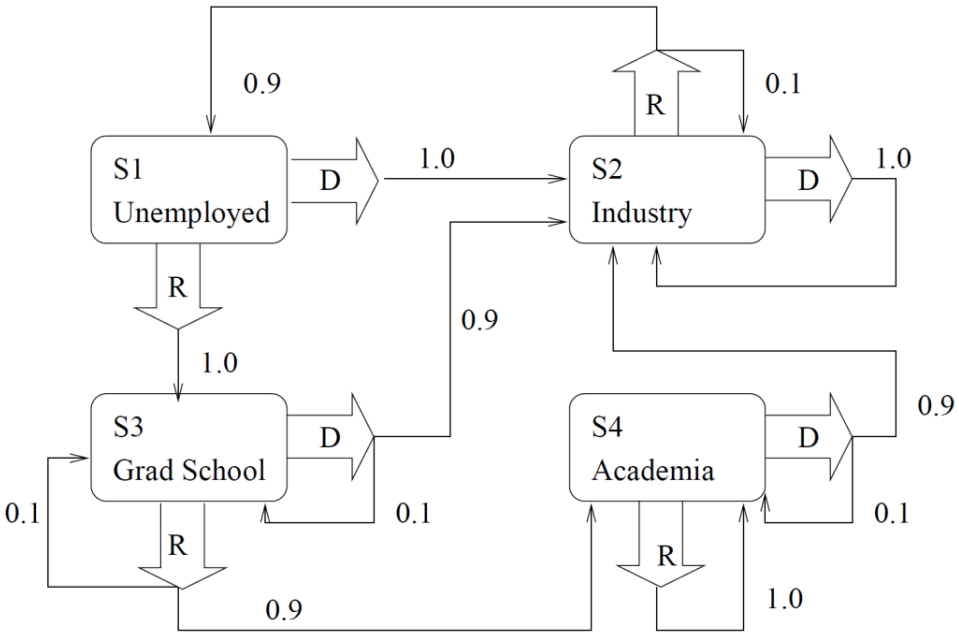
b) Solve the MDP using value iteration with a discount factor of 0.8. Start with 0 values. Let the stopping criterion threshold be $\epsilon = 10^{-4}$. What are the state values after two iterations? What are the state values at convergence? Round the results to 2 decimal places.

c) Determine the optimal policy by using the optimal values obtained in question **b)**. What method do you use to obtain the optimal policy? Write out the math expression, and the optimal policy.

d) Now answer the three questions **a)** – **c)** above for the following variant of the robot problem: If it spins its wheels, it climbs the slope in each time step (from low to medium or from medium to high) or stays high, all with probability 0.3. It stays where it is with probability 0.7. If it does not spin its wheels, it slides down the slope to low with probability 0.4 and stays where it is with probability 0.6. Everything else remains unchanged from the previous questions.

Problem 3. Non-deterministic world.

Consider the Markov Decision Process below. Actions have nondeterministic effects, i.e., taking an action in a state always leads to one next state, but which state is the one next state is determined by transition probabilities. These transition probabilities are shown in the figure attached to the transition arrows from states and actions to states. There are two actions out of each state: D for development and R for research.



Consider the following deterministic *ultimately-care-only-about-money* reward for any transition *starting* at a state:

REWARD			
S1	S2	S3	S4
0	100	0	10

Let π^* represent the optimal policy which is *given* to you, namely, for $\gamma = 0.9$, $\pi^*(s) = D$, for any $s = S1, S2, S3, S4$.

3.a. Compute the optimal value for each state, namely $V^*(S1), V^*(S2), V^*(S3), V^*(S4)$, according to this policy. **Show your work.**

3.b. Calculate the Q-value: $Q^*(S2, R)$. **Show your work.**