

Exercise

1. Gridworld domains are some of the most common in RL. Implement a 7x7 gridworld as an MDP:
 - Given a state and an action, you should be able to execute the action and return a new state and reward.
 - The actions are the cardinal directions and are deterministic. The states are just the cell the agent is currently in. Assume the third row from the top is all obstacles, except the right-most cell. The initial state is the bottom left, and the goal the top left.
 - There is no discounting. The reward for reaching the goal is 20, and -1 for each action.
2. Run a random agent on this domain that terminates after 50 steps.
3. Compute the optimal value function by hand and run a greedy agent using this value function.

By next week's lecture, submit on Moodle (groups of up to 4):

1. A bar graph of the returns accumulated by the two agents, averaged over 20 runs each.
2. A side-by-side illustration of a sample trajectory taken by each agent.
3. Your code