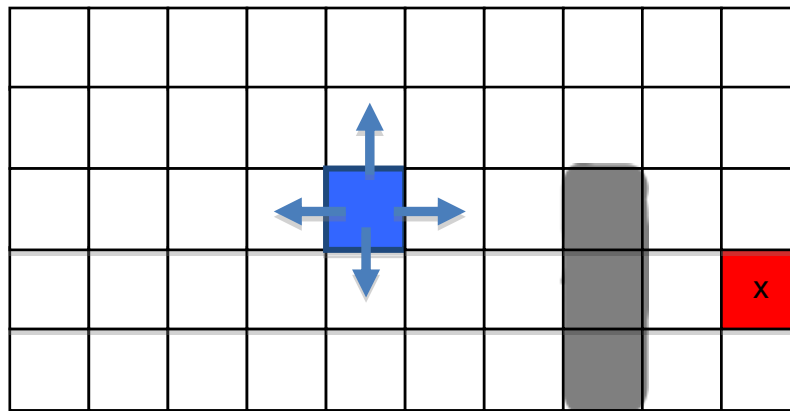


ROB 537
Learning Based Control
Fall 2021
HW #3: Reinforcement Learning
Due 11/2/2021 at 11:59 p.m.

1 - Use your favorite programming language to implement a **temporal difference reinforcement** learning algorithm in a 5x10 Gridworld. There is:

- A door on the red square (x) with a reward of 20;
- A solid wall (gray) the agent cannot move across;
- A reward of -1 for every time the agent is in a state other than the red door state.

The agent starts at a random location and has five actions (move in four directions or stay in place), the system state is the location of the agent (x, y), and an episode is 20 time steps. Use the ϵ -greedy action selection and show how the system learns.



2 - Implement a **Q-learning algorithm** and use it to solve the same, 20 step problem. How did the algorithm perform? How did solutions compare to the simple temporal difference algorithm tried for problem 1? Discuss the implications of your results.

3 - Now change the environment such that the **red door moves randomly by 1 cell** every time step. Keep the initial starting location of the door the same as before. Use the **EXACT same algorithm** from problem 2 to solve this problem. How does the performance of the agent compare to problem 2? Does the agent learn a good policy? Describe the results you obtain and hypothesize why your agent performs the way it does. Speculate on how you may improve the performance of the agent.

Please submit your assignment as a PDF, and submit your code in a SEPARATE zip file.