ELL729: Stochastic Control & Reinforcement Learning
Coding Assignment III

Maximum Marks: 7

December 25, 2020

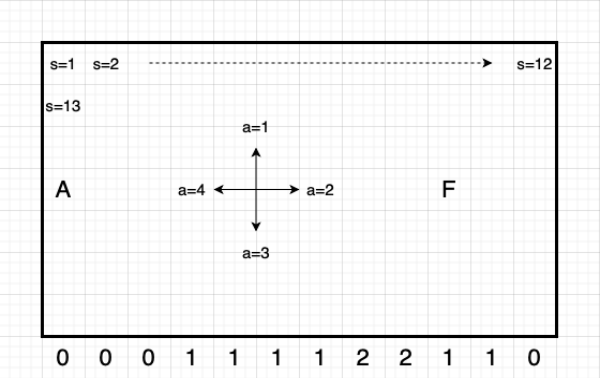

Figure 1

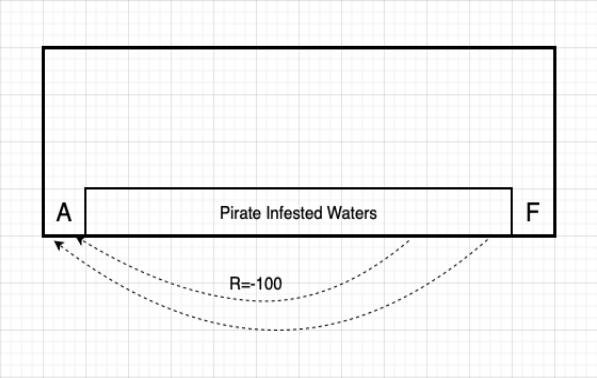

Figure 2

Troubled Waters Problem I

You are a sailor who wishes to take his ship from point A to point F in Figure 1. However on certain paths, there is a strong wind blowing which pushes your ship northwards. The severity of this wind is noted below each column. For each state, you can choose to go in either of the 4 directions. For example you are at s=16 and take a=2, then you'll end up at s=5 since wind strength is 1. You cannot go out of the grid. For example, if you are at s=2 and choose a=1, then you'll stay at s=2. You keep accumulating rewards of -1 till you reach F. This is an episodic task.

To Do List

- Use Q-Learning with different values of ϵ for an ϵ-greedy strategy to figure out the optimal path. Initialise all Q values to 0.
- Plot this optimal path on a similar diagram and include it in your report.
- Also plot the optimal policy(N, E, W or S) on such a similar diagram.

Troubled Waters Problem II

Similar to Problem I, you have to reach F starting from A. However a section of the sea is infested by pirates. If you move into this area, you are sent back to starting point A with a reward of -100. State-Action dynamics are the same as that in Problem I and you receive a reward -1 till you reach F.

To Do List

- Use Q-Learning with different values of ϵ for an ϵ-greedy strategy to figure out the optimal path. Initialise all Q values to 0.
- Plot this optimal path on a similar diagram and include it in your report.
- Also plot the optimal policy(N, E, W or S) on such a similar diagram.
- Is this task episodic? Does the ϵ-greedy strategy pose a problem here? Why? Do you see a discrepancy with the average rewards over episodes when the task is solved for 1000 episodes using the optimal or near-optimal Q-values.

Evaluation Criteria

- **You must create a class instance for both the environments.**
- 3 for each problem (report+code). 1 for the last to-do of problem II.
- Code without comments and proper indentation may not be evaluated.

Logistics

- **Deadline:** 16^{th} January. **NO EXTENSION AT ALL. NO LATE DAYS.**
- **Only** MATLAB or Python will be accepted. You may use **draw.io** to make diagrams.
- Any plagiarism detected will lead to a zero in the entire programming assignments section i.e. a zero on twenty.
- Switching to project will not be allowed if plagiarism is detected.
- Libraries for Q-Learning are not allowed. Numpy is allowed and encouraged for python.
- Make one .py or .m file for each of the two problems.
- All discussions pertaining to the assignment to be done on Piazza.
- Make one single zip file containing all the files and name it as 2017MTabcde_3.zip