CS 7642: Reinforcement Learning and Decision Making **Homework #5**

Bar Brawl

1 Problem

1.1 Description

You are the proprietor of an establishment that sells beverages of an unspecified, but delicious, nature. The establishment is frequented by a set P of patrons. One of the patrons is the instigator and another is the peacemaker.

On any given evening, a subset $S \subseteq P$ is present at the establishment. If the instigator is in S but the peacemaker is not in S, then a fight will break out. If the instigator is not in S or if the peacemaker is in S, then no fight will occur.

Your goal is to learn to predict whether a fight will break out among the subset of patrons present on a given evening, without initially knowing the identity of the instigator or the peacemaker.

1.2 Procedure

Develop a KWIK learner for this problem (see Li et al. 2011). Your learner will be presented with S, the patrons at the establishment, and the outcome (fight or no fight) of that evening. Your learner will attempt to predict whether a fight will break out, or indicate that it doesn't know, and should be capable of learning from the true outcome for the evening.

For each problem, the following input will be given:

- `at_establishment`: a Boolean two-dimensional array whose rows represent distinct evenings and whose columns represent distinct patrons. Each entry specifies if a particular patron is present at the establishment on a particular evening: a 1 means present and a 0 means absent.
- `fight_occurred`: a Boolean vector whose entries are the outcomes (a 1 means "FIGHT" and a 0 means "NO FIGHT") for that particular evening.

Specifically:

- For each episode (evening), you should present your learner with the next row of `at_establishment` and the corresponding row of `fight_occurred`.
- If your learner returns a 1 (for "FIGHT") or a 0 (for "NO FIGHT"), you may continue on to next episode.
- If your learner returns a 2 (for "I DON'T KNOW"), then you should present the pair (`at_establishment`, `fight_occurred`) to you learner to learn from.

You will submit a string of integers corresponding to the returned values of each episode.

The test case will be considered successful if no wrong answers are returned **and** the number of "I DON'T KNOW"s does not exceed the maximum allowed by the autograder.

2 Examples

The following example can be used to verify that your agent is implemented correctly.

- `at_establishment = [[1,1], [1,0], [0,1], [1,1], [0,0], [1,0], [1,1]]`
- `fight_occurred = [0, 1, 0, 0, 0, 1, 0]`
- Output: `0200010` (No incorrect answers and only one "I DON'T KNOW".)

3 Resources

The concepts explored in this homework are covered by:

3.1 Readings

- Li-Littman-Walsh-2008.pdf Li et al. 2011

4 Submission Details

The due date is indicated on the Canvas page for this assignment.
Make sure you have set your timezone in Canvas to ensure the deadline is accurate.
Submit your answers on Canvas, as outlined in section 1.2. You will have a total of 10 submission attempts—only the highest score is kept.

References

[Li+11] Lihong Li et al. "Knows what it knows: a framework for self-aware learning". In: *Machine learning* 82.3 (2011), pp. 399–443.