CS 229 Machine Learning, spring 2019

Homework 10:

Reinforcement Learning

Due Saturday May 18, 11:59pm

Submit by the **blackboard system**

**Question1: (70 points) Optimal values and policy**

Given the grid world in Figure 1, there are 4 deterministic actions: *up*, *down*, *left* and *right*. The goal is to reach the G, starting at S.

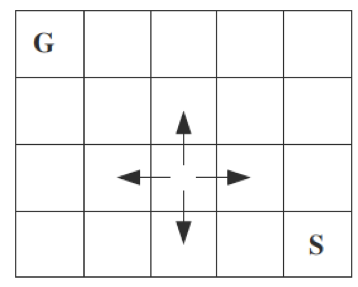


Figure 1 Grid world

The reward on reaching on the goal (G) is 100.

The reward on actions that would take the agent off the grid is -1 (agent stays still in this case).

The reward on other actions is 0.

The discount factor γ = 0.9,

Use **Q learning** to learn the optimal values of Q\* (s, a). Please submit your own code on calculating the Q\*(s,a).

--1 (20pts) What is the Q\*(s,a) for each pair of s and a?

--2 (5pts) What is the V\*(s) for each s?

--3 (5pts) What are the actions of optimal policy?

Use **Sarsa** algorithm to learn the optimal values of Q\* (s, a). Please submit your own code on calculating the Q\*(s,a).

--1 (20pts) What is the Q\*(s,a) for each pair of s and a?

--2 (5pts) What is the V\*(s) for each s?

--3 (5pts) What are the actions of optimal policy?

**Question2: (40 pts) A change on the grid world**

--1 (20pts) Add another goal state, e.g., to the lower-left corner.

How does the optimal policy change? Using either Sarsa or Q-learning.

--2 (20pts) Then, with these two goal states, if a state of reward **−**100 (avery bad state) is defined in the cell at last row of the second column. How does the optimal policy change? Using either Sarsa or Q-learning.