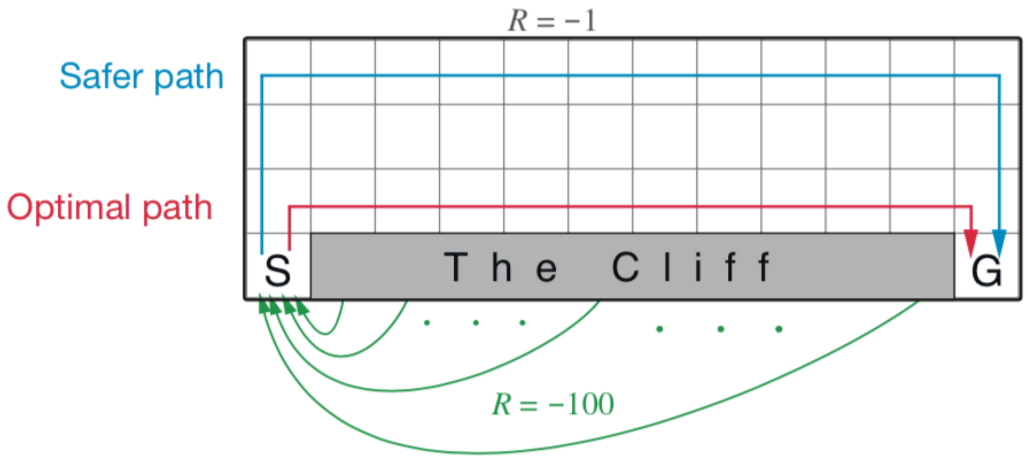
<u>Reinforcement Learning Mini-Assignment</u>

In this assignment, you will implement a value iteration agent, Sarsa, and Q-learning agents in a **variant** of the Cliff World environment. The agent starts in the bottom left corner of the grid-world below and takes actions that move it in the four directions. **The agent at the cliff incurs a reward of -100 and the run will terminate.** The reward for all other transitions is -1. When the agent reaches the bottom right corner it will receive a reward of 10 and the run will terminate.



We have provided the code skeleton including all the environment and simulator codes to get you started, which means you will only have to write the agent's functions. **Do not modify code in files other than 'valueIterationAgents.py' and 'qlearningAgents.py'.**

- Run manual control: ***python gridworld.py -m***
- Run random agent: ***python gridworld.py -a random***
- Run value agent with 10 iterations: ***python gridworld.py -a value -i 10***
- Run q learning agent for 10 episodes: ***python gridworld.py -a q --episodes 10***
- Run q learning agent with the update in place, under manual control:
  python gridworld.py -a q ---episodes 10 -m

<u>Submissions</u>

- ***<your metric number>_qLearningAgents.py***
  Your code for QLearningAgent.

  Note: For getPolicy, you should break ties randomly for better behavior. The random.choice() function will help. In a particular state, actions that your agent *hasn't* seen before still have a Q-value, specifically a Q-value of zero, and if all of the actions that your agent *has* seen before have a negative Q-value, an unseen action may be optimal.

- ***<your metric number>_valueIterationAgents.py***
  Your code for ValueIterationAgent.

  Your value iteration agent will be graded on a new grid. We will check your values, q-values, and policies after fixed numbers of iterations and at convergence (e.g. after 100 iterations).

- ***<your metric number>_RL.pdf***
  A short report describing the programming assignment and your solution.