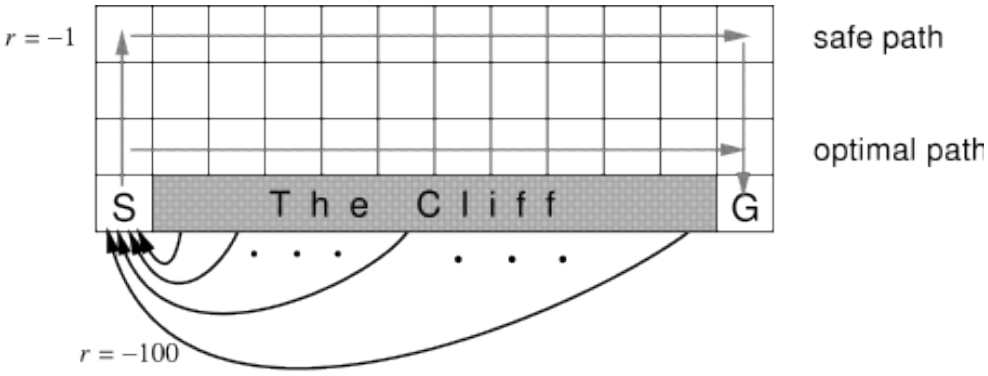
Reinforcement Learning – exercise 04

Hung Ngo & Vien Ngo
Machine Learning & Robotics lab, University of Stuttgart
Universittsstrae 38, 70569 Stuttgart, Germany

April 27, 2016

SARSA vs. Q-learning

Consider the following *Cliff Walking* problem (from example 6.6, Sutton & Barto's book):



This is a standard undiscounted, episodic task, with start (S) and goal (G) states, and the usual actions causing deterministic movement up, down, right, and left. Reward is -1 on all transitions except into the the region marked "The Cliff." Stepping into this region incurs a reward of -100 and sends the agent instantly back to the start.

a) Implement SARSA and Q-learning methods using ϵ-greedy action selection strategy with *fixed* $\epsilon = 0.1$. Choose a small learning rate, e.g., $\alpha = 0.1$.

– Evaluate the learned *behavior* policies of the two methods.

– Evaluate the learned *control* policies of the two methods.

Note: For all comparisons, plot the learning graph having *episode* as x-axis, *reward per episode* as y-axis (confer the results of Figure 6.13 in the textbook).

b) Propose a schedule that gradually reduces ϵ, starting from e.g. $\epsilon = 1$. Be creative! Then redo question a) using such schedule of ϵ instead of the fixed value.