COMS 4030A

Adaptive Computation and Machine Learning

LAB EXERCISE 6

This lab is not for submission and not for marks.
However, you are encouraged to attempt it as it is a good way to understand VALUE ITERATION, SARSA and Q-LEARNING algorithms, and it is quite interesting to see these algorithms in action.

(1) Code up an MDP grid-world similar to those used in the lecture notes. You can make it about 10×10 in size, with the following rules:

• The start state is s_{00}, the goal state is s_{99}.
• In each state the available actions are ℓ, u, r and d (for *left*, *up*, *right* and *down*) that move the agent one cell in the chosen direction. If an action would take the agent off the grid, then the agent stays in the current cell.
• The reward function is as follows: an action that would take the agent off the grid gets a -5 reward, an action that gets the agent to the goal state gets a 100 reward and otherwise an action that moves the agent one cell gets a -1 reward.

(2) You can make the MDP more interesting by adding in some extra features, such as:
• *barriers* between cells that an agent cannot pass through that act as grid walls.
• *partial barriers* between cells that an agent can pass through, but for which there is a -3 reward.
• *teleport cells*, which are cells with the property that when the agent would move there from any direction, the agent is teleported directly to some cell, say c_1, with probability p and to some other cell, say c_2, with probability $1 - p$. The reward for such a move is -1.

(3) Code up the following algorithms for the above MDP:
• VALUE ITERATION,
• SARSA,
• Q-LEARNING.

(4) As output, print out a 10×10 grid with the optimal action for each cell, i.e., place an ℓ, u, r or d in each cell that indicates what the optimal action is for that cell.