Grid World of (4X13)

Step1.Identify start state and goal state in the grid and assign initial state and action value

Step2. Associate reward for every action

Step3. Calculate rewards for both Sarsa and Q-learning over the stateActionValue

Step4. Average the reward sum from 10 successive episodes. Though result is expected from single run but I have done this over 15 independent runs instead to draw the figure to get smooth curve. However, the optimal policy converges with a single run