CHENNAI MATHEMATICAL INSTITUTE

Reinforcement learning

Assignment 2.
Date: April 2, 2021. Due date: April 12, 2021.

(1) Use TD to estimate the value of the uniform random policy (the policy which always select a uniform action, probabilityt 1/4 for L,R,U,D for the 6 x 6 grid problem from Assignment 2. Initiqalize all staes with value zero. Make plots with x-xis being step size (using a logarithmic scale, and any reasonable range that includes 0.1, 0.01, and 0.001), and where the vertical axis is the mean squared TD-error of the value function output by TD after 100 episodes. To estimate this stop updating the weights after 100 episodes, and run an additional 100 episodes. During these extra 100 episodes, compute the TD error at each time step, square it to obtain the squared TD error, and report the average value of these TD errors as a function of step size.

(2) Implement Q-learning and SARSA using a tabular estimate of the Q function and also using linear approximation. When selecting a state use ϵ-greedy. You may need to optimize ϵ. Apply this algorithm to the 6x6 grid world problem. Plot the learning curves as a function of the number of episodes. The y-axis is the reward obtained. On the x-axis you will plot the episode number and the average of the reward obtained in the i-th episode when you run say 1000 trials. A trial has 200 episodes. You always start at $(1, 1)$-always. Rewards are as in assignment 2 and the discount factor is 0.9.

Also mention how many linear features you selected for each state and the mapping of states to vectors.

(3) Implement example 6.6 from the book and plot the graphs your code gives.