Principles of Online Decision-Making (CS-303)

Problem Set 7

Problem 1

We continue working with the Gymnasium Taxi environment. In this homework, we focus on learning a policy using the Monte Carlo method.

(a) Fill the blanks marked by `# PODMexercise` to complete the on-policy first-visit MC control algorithm with epsilon-soft policy (slide 23 of week 10th and [SB, page 101]).

(b) Set `epsilon_decay` to be 0. Run the code with `epsilon` = 0.01, `epsilon` = 0.1, `epsilon` = 0.4. Which one performs the best in terms of the average reward in the test runs? Can you reason why the given epsilon works the best?

(c) Reason why setting `epsilon_decay` to be a positive value helps the performance.