Topics in Machine Learning (CSE975) IIIT Hyderabad
Instructor: Dr. Naresh Manwani

Assignment 1

Due: 18 October 2021, 11:59 pm

Question 1: OpenAI CliffWalking-v0 (Value Iteration and Policy Iteration) [10 marks]

Use CliffWalking-v0 from OpenAI gym.

a. Create two agents which provides a policy using dynamic programming methods and bootstrapping(DPAgent: PolicyIteration, ValueIteration). For PolicyIteration break down your agent's update function into a function to evaluate policy and a function to update policy. Also, create a ConfusedAgent, which randomly picks an action available from a given state (No need to train this one)

b. Test-run and visualizing learning.

   (a) Compare and plot the agents' learning using the return obtained v/s training iterations. Indicate mean, min and max return over N runs in the plot. Are the 2 learned agents performing better than ConfusedAgent (compare average return of all three agents) ?

   (b) Compare differences in paths obtained by setting γ [0, 0.1, 0.5, 0.75,1] while learning. How does γ affect the final path (or the policy learnt)?

Question 2: OpenAI Roulette-v0 (Monte Carlo methods and TD methods) [10 marks]

Use Roulette-v0 from OpenAI gym. (Pick suitable parameters for training).

a. Prepare and train your agent using i) On-Policy Monte Carlo and ii) Off Policy Monte-Carlo using Important Sampling. Plot the rewards (over N runs) vs episodes. Also plot number of unique state covered in the rollout so far versus return.

b. Prepare and train another agent using i) Q-Learning and ii) SARSA. Plot the rewards (over N runs) vs episodes.

c. Which among the above four methods performed better ? Compare how many episodes each method took to learn the best policy.

Question 3: OpenAI Taxi-v2 (TD methods) [15 marks]

Use Taxi-v2 from OpenAI gym. Plot the rewards (over N runs) vs episodes for all four methods and compare (Pick suitable learning-rate and discount-factor).

(1) Q-learning (2) Double Q-learning (3) SARSA (4) Expected SARSA

Submission Instructions

- Submit your source code in main_QuestionNumber.ipynb also exported into a main_QuestionNumber.html. No trained models.
- Do not use any DL/RL libraries.
- Write suitable comments to describe the methods you have implemented.

Plagiarism Policy: Plagiarism detection software is guaranteed to be run before any evaluation. Trying to beat any such software will make your code significantly unreadable and easy to prove malicious intent. In case of heavy plagiarism - all parties involved (giver, taker) will get a 0.