International Institute of Information Technology

Assignment: 1

---

1. Question 1: OpenAI CliffWalking-v0 (Value Iteration and Policy Iteration)

Use CliffWalking-v0 from OpenAI gym.

1. Create two agents to find the optimal policy using Policy Iteration and Value Iteration.
   For PolicyIteration, break down your agent's update function into a function to evaluate policy and a function to improve policy. Also, create a ConfusedAgent, which randomly picks an action available from a given state (No need to train this one)
2. Test-run and visualizing learning.
   (a) Compare and plot the agents' learning using the return obtained vs training iterations. Are the 2 learned agents performing better than ConfusedAgent (compare the final return of all three agents) ?
   (b) Compare differences in paths obtained by setting γ as [0, 0.1, 0.5, 0.75, 1] while learning. How does γ affect the final path (or the policy learnt)?

2. Question 2: OpenAI Taxi-v3 (Monte Carlo methods and TD methods)

Use Taxi-v3 from OpenAI gym. (Pick suitable learning-rate and discount-factor)

1. 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 for some appropriate N) vs episodes during training. Also, plot the number of unique states covered in the rollout so far versus return.
2. Prepare and train two more agents using i) Q-Learning and ii) SARSA. Plot the rewards (over N runs) vs episodes during training.
3. Which of the above four methods performed better? Compare how many episodes each method took to learn the best policy.

3. Submission Instructions

1. Submit your source code in **main_QuestionNumber.ipynb** also exported into a **main_QuestionNumber.html** inside a zipped file **rollnumber_A1.zip** . No trained models.

2. Do not use any RL/DL libraries.
3. Write well-commented code to describe the methods you have implemented.
4. Answer the questions asked in the ipynb file itself.

4. 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.