ASSIGNMENT-2A

DeepEigen | COURSE: RL-1.0Y: Reinforcement Learning | YEAR: 2022 | INSTRUCTOR: SANJEEV SHARMA

1 Problem Statement

Consider a multi-armed bandit with 15 arms. True action value (q*(a)) for each arm is sampled from a Gaussian distribution with mean = 0 and variance = 1. Rewards are sampled from normal distribution with mean = q*(a) and variance 0.1. Following sample averages method for bandit problems, try to implement following problems

1.1 Tasks

- Draw a violin plot for reward distribution of each arm with mentioned reward statistics and true action values.
- Implement epsilon-greedy methods for action selection and greedy action selection with different values of epsilon=0, 0.01, 0.1. Compare average performance and optimal action selection for both of these methods for number of steps = 10000.
- Compare Incremental Implementation for epsilon-greedy Bandit problem with Sample Averages method by estimating optimal action values for each arm. Mention differences in magnitude following each method.
- Implement an Optimistic Initial Values method for greedy and epsilon-greedy method, where optimistic initial value of q*(a) = 5 is provided for greedy bandit and q*(a) = 0 for epsilon-greedy. Compare the differences between both of these methods.

1.2 To Submit

- **solutions.txt** : Observations and parameters taken (as applicable) for all tasks and subtasks section wise for the whole assignment.
- **violin_plot.jpg** : Rewards Distribution for all arms.
- **average_performance_greedy_epsilon_greedy.jpg** : Average Performance for Greedy vs Epsilon-Greedy Bandits for epsilon=0,0.10,.01
- **optimal_actions_greedy_epsilon_greedy.jpg** : Optimal Action Selection for Epsilon-Greedy and Greedy Bandits for epsilon=0,0.10,.01.
- **incremental_vs_sample_averages_optimalactions.txt** file with comments on comparison of both methods.
- **incremental_vs_sample_averages_optimalactions.jpg** : Optimal Action Selection in incremental and sample averages plot.
- **optimistic_initials_vs_epsilon_greedy.jpg** : Plot of average performance for epsilon greedy method and optimistic initials greedy method for different values, i.e. q*(a) = 0 and q*(a) = 5.

1.3 Instructions

- To submit the assignment put all files in a folder named **username** where username is your username which you used to sign up with Deep Eigen, e.g. **username_assignment_rl10y_2.zip**