ASSIGNMENT-3

Deep Eigen	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, implement following problems

1.1 Tasks

- Using knowledge from Upper confidence bound (UCB), implement same method in case of bandits for c=2, 3, 10. Compare UCB for c=2 with epsilon greedy for epsilon = 0.1, 0.01. Do performance comparison for each of the cases.
- Implement Gradient Bandit Algorithm with Initial H(a) = 10 and H(a) = 100. Also consider baseline = 4 and baseline = 0 for gradient computation. Compare performance for both these methods.

2 To Submit

- **UCB_vs_epsilon_greedy.png** : Performance comparison for UCB approach and epsilon-greedy approach
- **gradient_bandits_10.png** : Average performance of gradient bandits for preference value = 10 initially with baseline = 4 and baseline = 10
- **gradient_bandits_100.png** : Average performance of gradient bandits for preference value = 100 initially with baseline = 4 and baseline = 10

You should share with us:
put all files in a folder named **username** where username is your username with which you signed up in Deep Eigen, e.g. **username_assignment_rl10y_2.zip**