EE551000 System Theory
Homework 1: Multi-Armed Bandit
Due: October 20, 2020 23:59

Goal

The goal of this assignment helps you get familiar with basic action-value based methods in multi-armed bandit problems.

Todo

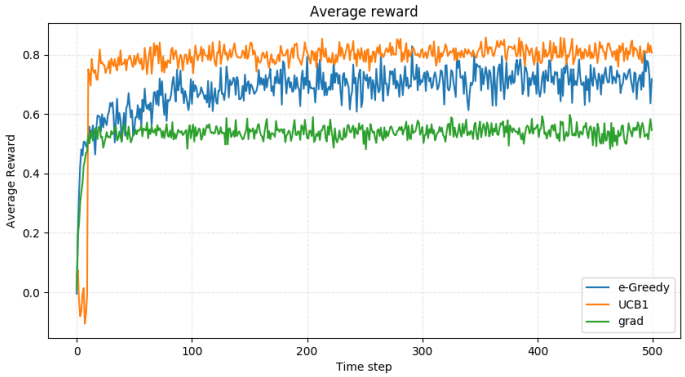
- Implement three algorithms:
  - ✓ ε-Greedy
  - ✓ upper confidence bound (UCB)
  - ✓ gradient bandit
- Get familiar with basic Python syntax.

Details

- File description
  - `env.py`: The bandit environment used in this assignment. We provide two kinds of distributions (Gaussian, Bernoulli) to randomly generate the reward function of each bandit. You should NOT modify this file.
  - `algo.py`: You'll implement three algorithms in the file. Please follow the instructions to complete your homework.
  - `utils.py`: Helper functions (such as plot) are implemented in this file. We strongly recommend to implement evaluation function or plotting function by your own in order to get familiar with plotting mechanism in Python. We provide an example plotting function as your reference.
  - `main.py`: main file for your implementation.

- You can show your implementation results after each method by running:
  `python main.py --algo [which_algo] --plot`
  This allows you to check the correctness of your implementation. If your implementation is correct, the result should be similar to that in the textbook.

- After you've done all the algorithms, you should implement plotting function on your own to analyze different settings. Or you can use the flag `--runAll` to show it out. For example:



- Please write a **README** file to explain how to run your code if you implemented extra functions.

Requirements and Installation

- Python version: 3.6
- Please run `pip install -r requirements.txt` to install necessary libraries.

Report

- **Title, name, student ID**
- **Implementation**
  - ✓ In ε-Greedy, how do you select action if the probabilities are equal?
  - ✓ In UCB, how do you select action when time steps < num of bandits?
  - ✓ Briefly describe your implementation.
- **Experiments and Analysis**
  - ✓ Plot the average reward curves of different methods into a figure.
  - ✓ Vary ε value with 0, 0.01, 0.1, 0.5 and 0.99. What happens? Why? Please plot it.
  - ✓ Vary the parameter c in UCB. What happens? Why? Please plot it.
  - ✓ Vary the number of bandits. What happens if the number of bandits is large? Please plot it.

Reminder

- Please upload your code and report.pdf to iLMS before 10/20 (Tue) 23:59. No late submission allowed.
- DO NOT zip your code into a single file.
- Please do not copy&paste the code from your classmates.