ISE 599 Data Mining Homework 4

Due on April 1, 2020

In order to build a feel for the epsilon-Greedy algorithm, you need to see how it behaves under a variety of circumstances. The following two questions are helpful to better understand the algorithm. For these questions use 1000 simulations for 250 visits for each case. Report the plots for optimal, average, and cumulative rewards. Consider $\epsilon = 0.1, 0.3, 0.5$ only.

- 1. (20 pts.) For the exercise solved in class, use a different number of arms than the five arms we have been working with. See how the algorithm performs if there are 2 arms and then if there are 20 arms. The best arm with probability 0.9, all other arms with probability 0.1). Submit one set of plots for 2 arms and other set for 20 arms.
- 2. For the exercise solved in class, change the probabilities of reward from the 5 Bernoulli arms as suggested below.
 - a) (10 pts.) How does epsilon-Greedy behave if the best arm probability is 0.6 and all other arms probabilities are 0.40?
 - b) (10 pts.) How does it behave when the probabilities for two arms are 0.60, for other two arms are 0.40, and for one arm is 0.5?
 - c) (10 pts.) How does it behave if all of the probabilities are 0.5?
- 3. (20 pts.) Test if the following web design have the same click-rates. If not, find which design(s) provides the largest click-rates.

	A	В	\mathbf{C}	D	Е
0110110	000	1000 1800	U _ U	200	880 1325

- 4. (30 pts.) In segmenting the market, a breakfast cereal manufacturer uses health and diet consciousness as the segmentation variable. Four segments are developed:
 - 1 =Concerned about eating healthy foods
 - 2 = Concerned primarily about weight
 - 3 = Concerned about health because of illness
 - 4 = Unconcerned

To distinguish between groups, a survey is conducted (see cereal.csv). In the survey, people are categorized as belonging to one of these groups. Each respondent was also asked the amount spent on breakfast cereal in an average month. Test if the amount spent by these four segments is the same. If not, use the t.test function to find CIs on the amount spend in an average month. Plot the CIs on the same graph and conclude which segments spend the most.

Report must show the student's last name, first name, and USC ID.