

CS 6700: Programming Exercise 1

You are to conduct experiments on the 10-arm bandit test bed described in Section 2.2 of the book. Please hand in the code listing and the necessary graphs. Label the graphs clearly, with axes labels, parameter values, question numbers etc. Ensure that the code is adequately commented. Also turn in a short writeup of your observations from the experiments and answers to any questions asked below.

Remember: The graphs are to run for 1000 plays, with each curve being the average of the performance on 2000 different bandit problems, generated as per the description in the book.

The points will be given according to the following criteria:

- Correct generation of the bandit problems
- Correct coding of the learning algorithms
- Correct code for gathering data to plot the graphs
- Correct performance of the learning algorithms
- Neatness of the graphs - correctly labelled etc. and well commented code

Note: You can program in any language you desire. But if you need any help later on, it will have to be in a language I can understand. So check with me before you start.

1. (10 marks) First implement the 10-arm bandit testbed. Reproduce the graphs given in Figure 2.1 on page 29 of the text book. Specifically, implement the ϵ -greedy method on the 10-armed bandit testbed, for different values of ϵ .

Use graphing software of your choice, (I suggest either Matlab or Gnuplot), to produce the graphs. Ensure that you have labelled the graphs correctly.

2. (8 marks) How does the softmax action selection method using the Gibbs distribution fare on the 10-armed testbed? Implement the method and run it at several (at least 3) temperatures to produce graphs similar to those in Figure 2.1. Note that now you are required to sample from the softmax distribution, and **not** take the action with the maximum probability!
3. (7 marks) Implement the UCB1 algorithm. Compare the performance with that of epsilon greedy and softmax. What do you observe? Why do you think this is so?
4. (5 marks) What happens as the number of arms grows? Run experiments on a 1000 arm bandit setup and report comparison results as in the previous question.

The due date for the homework is: **Friday, February 24th.**