

CISC 8999 Reinforcement Learning

Spring 2019

Download and read the code for the 10-arm bandit experiment reported in Sutton Fig. 2.2

Make sure you can run this software and that you understand it.

Modify the software as follows:

- (1) Investigate and graph a few values of epsilon between 0 and 0.2
- (2) Investigate the effect of three different initial values and epsilon
- (3) Implement upper bound selection and compare on $\epsilon=0.1$, $C=2$
- (4) What if you use an unbiased step size?
- (5) Add a gradient bandit algorithm and compare. Use unbiased step size

Submit code and graphs for each of (1) through (5).