به نام خدا

تمرین درس یادگیری ماشین – سری اول

Assume we have an agent that plays a 5-armed bandit, trying to maximize its reward. By pulling a lever, the agent will be given some reward. Each of 5 levers has its own reward distribution. You are not aware of reward distributions, but a MATLAB code that generates samples of each lever is given to you. Use the following function in the file "Bandit.p" to generate samples:

function Bandit(student_number, lever_number)

in which student_number tends to generate different samples for different students and lever_number is in {1,2,3,4,5}. For example

Bandit(810188100, 5)

1. Generate 10000 samples for each of levers. Draw probability density functions for each lever using histograms.
2. Implement epsilon-greedy action selection policy for action-value method. Use average rewards as action-values. Initialize action-values by a random uniform distribution in [0,100]. Consider three values of epsilon:
   a. Epsilon = .1
   b. Epsilon = .2
   c. Epsilon = .5

   For each value of epsilon, run the algorithm 100 times, each consisting of 1000 iterations. Draw average reward (average of 100 runs) in terms of iteration number. Compare and analyze the results: consider speed of convergence and asymptotic value of reward.
3. Repeat previous part using adaptive epsilons and compare the results:
   a. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \frac{\varepsilon_t}{1+0.1t}$
   b. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \frac{\varepsilon_t}{1+0.01t}$
   c. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \frac{\varepsilon_t}{1.1}$
   d. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \frac{\varepsilon_t}{1.01}$
   e. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \varepsilon_t e^{-0.01t}$
   f. $\varepsilon_0 = 0.5, \varepsilon_{t+1} = \varepsilon_t e^{-0.001t}$
4. Repeat previous part using reinforcement comparison and compare the results. Try these parameter values:
   a. $\alpha=0.1, \beta=0.1, r_0=0$
   b. $\alpha=\frac{1}{k}, \beta=0.1, r_0=0$
   c. $\alpha=0.1, \beta=\frac{1}{k}, r_0=0$
   d. $\alpha=\frac{1}{k}, \beta=\frac{1}{k}, r_0=0$
   e. $\alpha=\frac{1}{k}, \beta=0.1, r_0=100$
5. Compare the best results of parts 2, 3 and 4.