Northeastern University
CS 7180 – Special Topics in AI (Reinforcement Learning)
Fall 2018, Robert Platt

Bandits Assignment

Name: _____

Problem	Points
Plotting recency-weighted averages	/57
Bandit Example	/12
Non-constant step sizes	/6
Total	/75

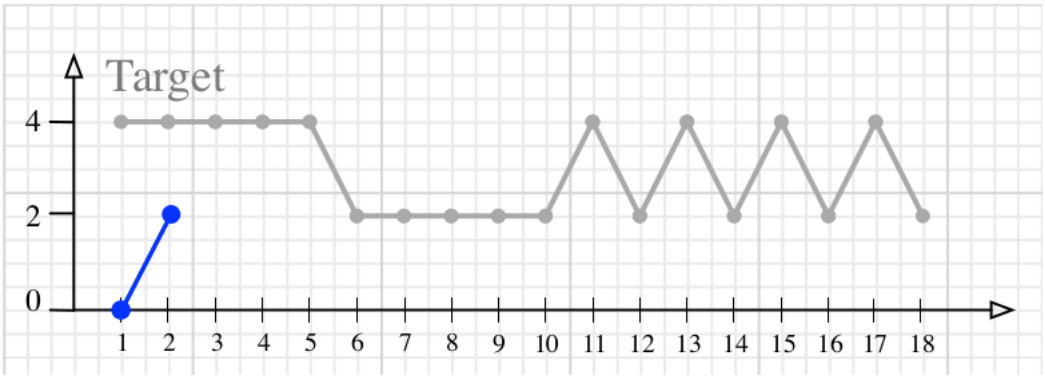
Instructions

- Don't cheat!
- Please scan your answers and submit them in PDF form.

(57 pts.) Plotting recency-weighted averages

Equation 2.5 (from the SB textbook, online 2nd edition) is a key update rule we will use throughout the course. This question will give you a better hands-on feel for how it works. Do all the plots in this question by hand. You should not even need a calculator, as the numbers are easy to place by eye (using fractions instead of floating points might help). You will get a better feel for it if you place the points manually. To make it easy for you, Ill include some graphing area and a start on the first plot here, so you should just be able to print these pages out and draw on them. This question has 7 parts.

a. (15 pts) Suppose the target is 4.0 for five steps, then 2.0 for five steps, and then alternates between 2.0 and 4.0 for 8 more steps, as shown by the grey line in the graph below. Suppose the initial estimate is 0.0, and that the step-size (in the equation) is 0.5. Your job is to apply Equation 2.5 iteratively to determine the estimates for time steps 2-19. Plot them on the graph below, using a blue pen, connecting the estimate points by a blue line. The first two estimate points are already marked and connected in dark blue (or black) below:



$\alpha = \frac{1}{2}$

What is the estimate after 5 updating steps (i.e., on step 6)? Please give the number as a symbolic expression.

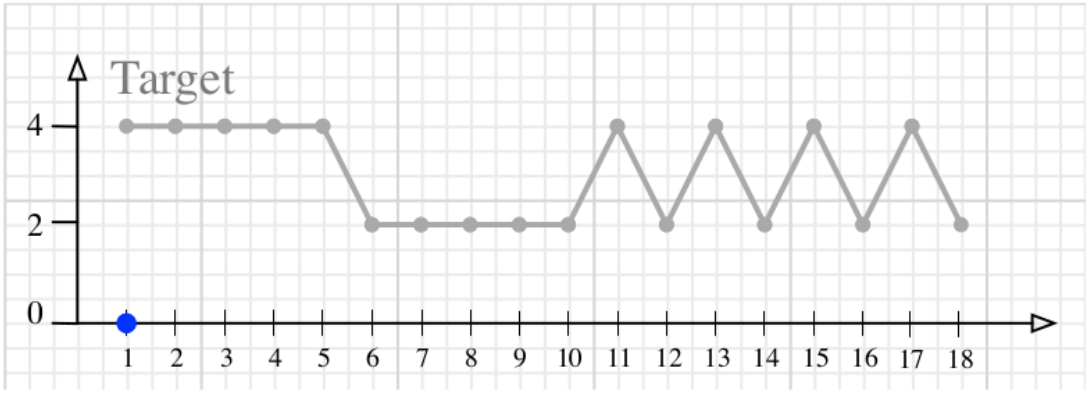
What would it be after 10 steps, if the target continued to be 4.0? Please give the number as a symbolic expression.

After 20 steps of a target of 4.0? Please give the number as a symbolic expression.

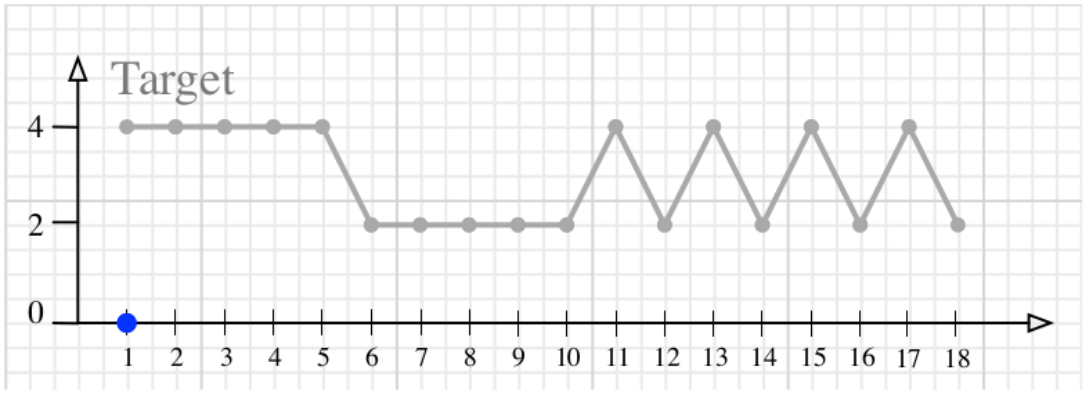
b. (6 pts) Repeat the graphing/plotting portion of Part 1, this time with a step size of 1/8.

c. (6 pts) Repeat with a step size of 1.0.

d. (6 pts) Which of these step sizes would produce estimates of smaller absolute error if the target continued alternating for a long time?



$\alpha = \frac{1}{8}$



$\alpha = 1$

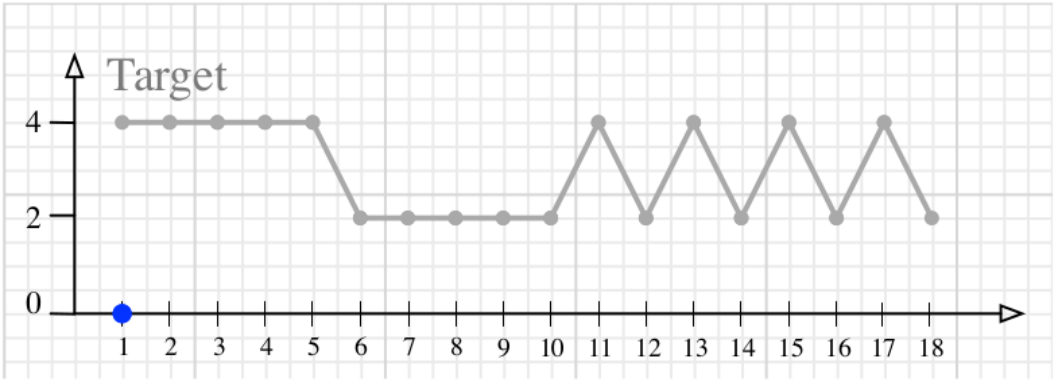
Which of these step sizes would produce estimates of smaller absolute error if the target remained constant for a long time?

e. (6 pts) Now suppose the target R_t moved randomly, such that $R_{t+1} = R_t + N(0,1)$ where $N(0,1)$ is a normally distributed random variable with mean 0 and variance 1. In this case, which of these step sizes would produce estimates of smaller absolute error?

f. (6 pts) Now suppose the target was of the form $R_t = N(\mu_t, 1)$ a normally distributed random variable with mean μ_t and variance 1, and the mean moved randomly according to $\mu_{t+1} = \mu_t + N(0,1)$. Then which of these step sizes would produce estimates of smaller absolute error? If you are not completely sure which is best, then say what you are sure of about the best step size and explain your reasoning.

g. (12 pts) Repeat with a step size of $\frac{1}{(t-1)}$ (i.e., the first step size you will use is 1, the second is 1/2, the third is 1/3, etc.).

Based on all of these graphs, why is the $\frac{1}{(t-1)}$ step size appealing?



$\alpha = \frac{1}{t\text{-}1}$

Why is the $\frac{1}{(t-1)}$ step size not always the right choice?

(12 pts.) Bandit Example

Consider a multi-arm bandit problem with k = 5 actions, denoted 1, 2, 3, 4, and 5. Consider applying to this problem a bandit algorithm using ϵ-greedy action selection, sample-average action-value estimates, and initial estimates of $Q_1(a) = 0$ for all a. Suppose the initial sequence of actions and rewards is $A_1 = 1, R_1 = 2, A_2 = 2, R_2 = 3, A_3 = 3, R_3 = 1, A_4 = 2, R_4 = 2, A_5 = 3, R_5 = 0, A_6 = 4, R_6 = 5$. On some of these time steps the ϵ case may have occurred causing an action to be selected at random. On which time steps did this definitely occur? On which time steps could this possibly have occurred?

(6 pts.) Non-constant step sizes

Exercise 2.4 in SB. Use extra paper as needed.