Advanced Computing: Reinforcement Learning:

Game 1

Name: Carter Gunderson

Do not use your book or the Internet or another student. If you have questions, you can ask Richard.

• [2] What is the temporal difference equation for computing avg, where the new value is R[i]?  
avg = avg + alpha(reward - avg)

•

• [2] If you use the greedy algorithm for 10-bandits, at any given time t, how is the next action chosen?  
the greedy algorithm chooses the action with the highest estimated reward

• [2] If a gaussian with mean = 0 and sd = 1 is used to set the true means of the 10 bandits, how likely is it that some of the bandits will have a negative mean reward?

• 0.5

• Define the epsilon-greedy policy, where epsilon = 0.1.

• if random.random() <= 0.1 { exploratory move }

• else { greedy move }

• Compare epsilon-greedy for epsilon = 0.1 with epsilon = 0.01in terms of short-term and long-term behavior (total rewards).

short term: epsilon = 0.1 will find the most rewarding action faster

long term: epsilon = 0.01 will exploit the most rewarding action more often