UNIVERSITÄT BONN
AIS
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN
COMPUTER SCIENCE VI AUTONOMOUS INTELLIGENT SYSTEMS
Prof. Dr. Sven Behnke
Friedrich-Hirzebruch-Allee 8

Robot Learning

Assignment 2

Due Tuesday, April 30th, before class.

2.1) Given is a six-armed bandit, as introduced in the lecture.
The first arm shall sample its reward uniformly from the interval [1, 3).
The second arm shall sample its reward uniformly from [-3, 8).
The third arm shall sample its reward uniformly from the interval [2, 5).
The fourth arm shall sample its reward uniformly from [–2, 6).
The fifth arm shall sample its reward uniformly from [3, 4).
The sixth arm shall sample its reward uniformly from [-2, 2).

What is the expected reward when actions are chosen uniformly?

4 points

2.2) Implement the six-armed bandit from 2.1) and compute the sample average reward for 10 uniformly chosen actions!
Compare this to your expectation from 2.1)!

4 points

2.3) Initialize $Q(a_i)=0$ and chose 4000 actions according to an ε-greedy selection strategy (ε=0.1)!
Update your action values by computing the sample average reward of each action recursively!
For every 100 actions show the percentage of choosing arm 1, arm 2, arm 3, arm 4, arm 5, and arm 6 as well as the resulting average reward!

4 points

2.4) Redo the experiment, but after 2000 steps sample the rewards of the fourth arm uniformly from [5, 7) !
Compare updating action values by computing the sample average reward of each action recursively (as done in 2.3) with using a constant learning rate α=0.01 !
For every 100 actions show the percentage of choosing arm 1, arm 2, arm 3, arm 4, arm 5, and arm 6 as well as the resulting average reward!

4 points

2.5) Modify your implementation by using an optimistic initialization $Q(a_i)=5$ and a greedy action selection strategy, still using a constant learning rate α=0.01 !
For every 100 actions show the percentage of choosing arm 1, arm 2, arm 3, arm 4, arm 5, and arm 6 as well as the resulting average reward !
Compare this to your result from 2.4)!

4 points