Reinforcement Learning

Jonathan Somer

April 28, 2018

Introduction

SAR:

- Start in state S_t
- Apply action A_t
- Get reward R_t

SAS:

- Start in state S_t
- Apply action A_t
- Move to state S_{t+1}

Multi-Arm Bandit

Epsilon-Greedy Explore-Exploit:

Algorithm 1 Epsilon-Greedy Explore-Exploit

```
for turn do
    draw a random p ∈ [0,1]
    if p < ϵ then
        explore()
    else
        exploit()
    end if
end for
```

Efficient Mean Update:

$$\bar{X}_N = \frac{N-1}{N}\bar{X}_{N-1} + \frac{1}{N}X_N$$