Monte Carlo Learning (MC Learning)

esgl Hu

November 27, 2017

Monte Carlo methods require only *experience* — sample sequences of states, actions, and rewards from actual or simulated interaction with an environment. It requires no prior knowledge of the environment's dynamics, yet can still attain optimal behavior.

If a model is not avaliable, then it is particularly useful to estimate *action* values (the values of state-action pairs) rather than *state* values. With a model, state values alone are sufficient to determine a policy; one simply looks abead one step and chooses whichever action leads to the best combination of reward values and next state. Without a model, however, state values alone are not sufficient.

Algorithm 1 First-visit MC policy evaluation (returns $V \sim v_\pi$)

```
Initialize π ← policy to be evaluated
Initialize V ← an arbitrary state-value function
Initialize Returns(s) ← an empty list, for all s ∈ S
while True do
    Generate an episode using π
    for each state s appearing in the episode do
        G ← return following the first occurrence of s
        Append G to Returns(s)
        V(s) ← average(Returns(s))
    end for
end while
```