Taylor Stapus

CS 370

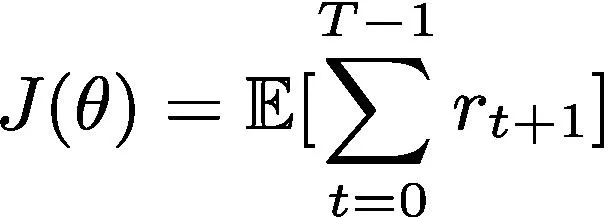
Module 6

04/10/2025

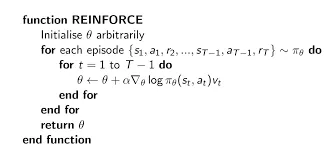
**Cartpole Revisited**

The Cartpole problem is a reinforcement learning task where the goal is to balance a pole on a cart by applying a certain amount of force to the left or right. The environment is usually completed implementing policy gradient methods and the reinforced algorithm is one of the simplest ones. Reinforce is a policy gradient method where it learns by sampling full episodes, computing the returns and then updates the policy based on the return.

**The objective function for policy gradients is defined as:**



**Here is the pseudo code for REINFORCE**:



**The policy usually has parameters that are simple neural network that outputs action probabilities.**

A computer screen shot of a program

AI-generated content may be incorrect.

A2C stands for Advanced Actor-Critic which is a algorithm that combines policy-based and value-based methods that are designed to address the limitations of each method when used individually. The actor makes the decision by selecting the action while the critic estimates the expected return from the state.

**Actor Critic Algorithm Objective Function**:

Policy Gradient (Actor): ∇*θ*​*J*(*θ*)≈*N*1​∑*i*=0*N*​∇*θ*​log*πθ*​(*ai*​∣*si*​)⋅*A*(*si*​,*ai*​) Value Function Update (Critic): ∇*w*​*J*(*w*)≈*N*1​∑*i*=1*N*​∇*w*​(*Vw*​(*si*​)−*Qw*​(*si*​,*ai*​))2

**Example Code of defining Actor and Critic Networks**:

A screen shot of a computer code

AI-generated content may be incorrect.

Policy gradient approaches differ from value-based by how they learn. Policy gradient approaches are designed to directly optimize the policy parameters by using feedback from the rewards. They are best used for handling continuous actions. Value-based approaches use the Bellman equation and derive a policy by greedily selecting the best action. They are more sample efficient and easier to implement in action spaces. Actor-Critic methods are a blend of both policy and value-based approaches. It is designed to reduce variance and stabilize learning by using the strength of both methods

**References:**

*Actor-Critic Algorithm in Reinforcement Learning*. (2023, December 13). GeeksforGeeks. <https://www.geeksforgeeks.org/actor-critic-algorithm-in-reinforcement-learning/>

Yoon, C. (2019, May 23). *Deriving Policy Gradients and Implementing REINFORCE*. Medium. <https://medium.com/@thechrisyoon/deriving-policy-gradients-and-implementing-reinforce-f887949bd63>

**‌**