FPT Education

FPT UNIVERSITY

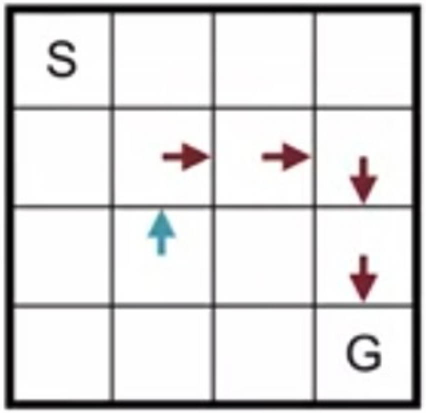
Exploration method for Monte- Carlo

Objectives

- ❑ Understand why exploring starts can be problematic in real problems
- ❑ Describe an alternative exploration method for Monte-Carlo control

Exploration for Monte-Carlo

- ❑ We can not always use Exploring Starts.
  - ❑ The situations where we cannot use exploring starts this algorithm must be able to start from every possible State action pair.
  - ❑ Otherwise the age of may not explore enough and could converge to a suboptimal solution in many problems.
  - ❑ It can be difficul ple an initial State action pair.



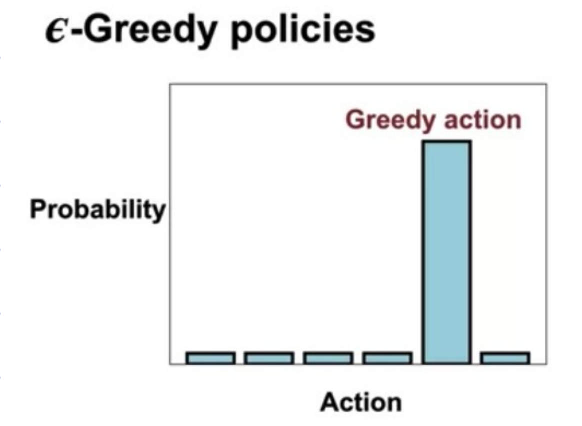
Exploration for Monte-Carlo

- ❑ Example: how would you randomly sample the initial State action pair for a self-driving car?

  - How could we ensure the agent can start in all possible States?
  - We would need to put the car in many different configurations in the middle of a busy freeway.
  - This would be dangerous and impractical.

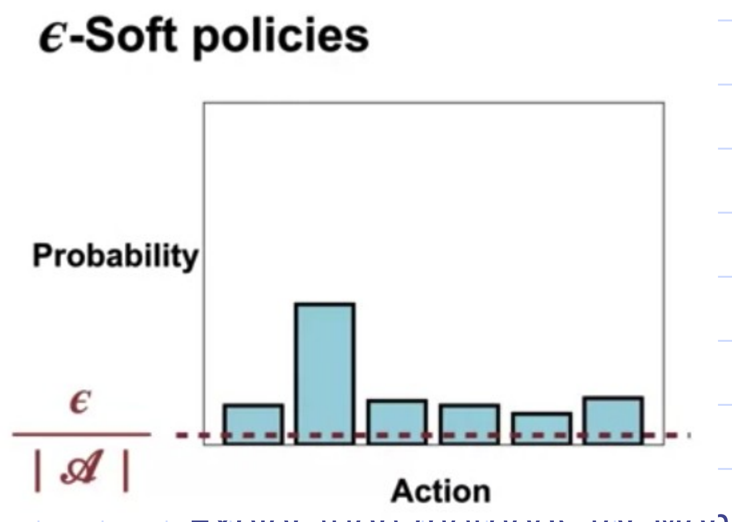
Exploration for Monte-Carlo

- ❑ How can we learn all the action values without exploring starts?
- ❑ We can use the Bandit with Monte Carlo to as a quick recap Epsilon greedy policies are stochastic policies.
- ❑ They usually take the greedy action, but occasionally take a random a



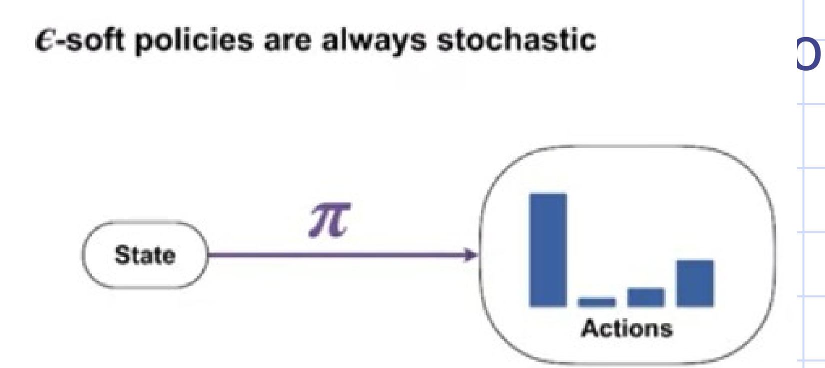
Exploration for Monte-Carlo

- Epsilon greedy policies are a subset of a larger class of policies called Epsilon soft policies Epsilon soft policies take each action with probability at least Epsilon over the number of actions.



Epsilon- Soft Policies

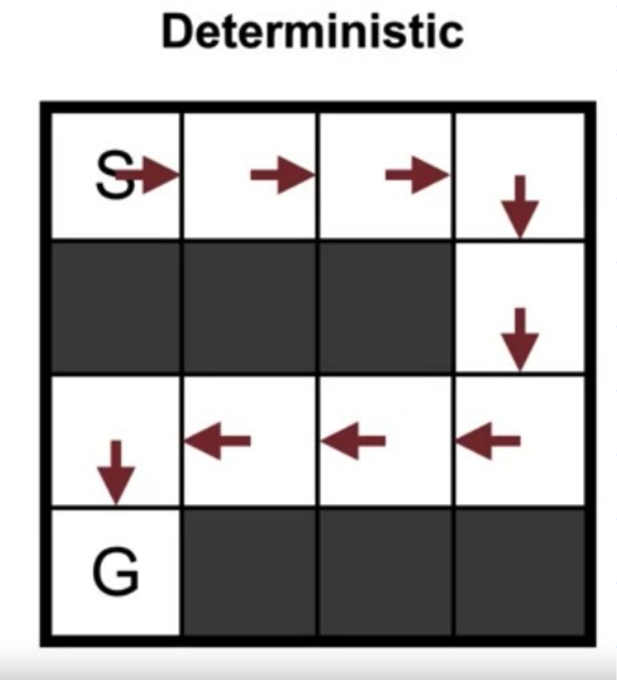
- □ Epsilon soft policies are always stochastic deterministic policy specify a single action to take in each state stochastic policies instead specify the probability of taking action in each state in epsilon.
- □ All actions have a on over the number c

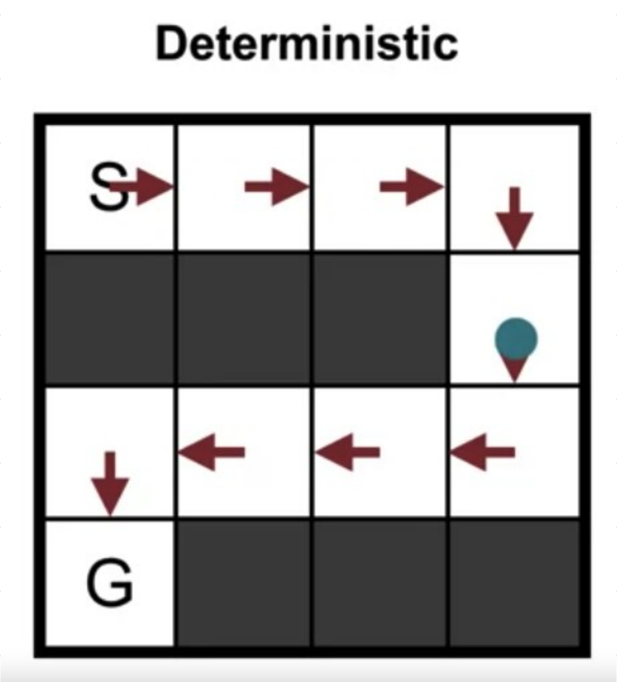


Epsilon- Soft Policies

- ❑ Example of an Epsilon greedy policy and a deterministic policy.
  - ❑ We have a grid rolled with the arrows representing the deterministic policy.
  - ❑ From the start State the agent will follow the exact same trajectory through the grip world.

Epsilon- Soft Policies





Epsilon- Soft Policies

- The Epsilon greedy policy has more arrows because every action has some small probability of being selected accordingly.
- The agent will probably follow a slightly different trajectory every episode.

Exploration for Monte-Carlo

- The Epsilon greedy policy has more arrows because every action has some small probability of being selected accordingly. The agent will probably follow a slightly different trajectory every episode.

Summary

- Understand why exploring starts can be problematic in real problems
- Describe an alternative exploration method for Monte-Carlo control

Q & A