SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

19IT601– Data Science and Analytics

III Year / VI Semester

Unit 3 – PREDICTIVE MODELING AND MACHINE LEARNING

Topic 9: Reinforcement Learning

Reinforcement Learning

Reinforcement Learning(RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences.

The goal is to find a suitable action model that would maximize the total cumulative reward of the agent.

Q-learning
Q-Learning is a Reinforcement learning policy that will find the next best action, given a current state.

It chooses this action at random and aims to maximize the reward.

Q-learning is a model-free, off-policy reinforcement learning that will find the best course of action, given the current state of the agent.

Depending on where the agent is in the environment, it will decide the next action to be taken.

Reinforcement Learning

The objective of the model is to find the best course of action given its current state.

To do this, it may come up with rules of its own or it may operate outside the policy given to it to follow.

This means that there is no actual need for a policy, hence we call it off-policy.

Important Terms in Q-Learning

- States: The State, S, represents the current position of an agent in an environment.
- Action: The Action, A, is the step taken by the agent when it is in a particular state.
- Rewards: For every action, the agent will get a positive or negative reward.
- Episodes: When an agent ends up in a terminating state and can't take a new action.
- Q-Values: Used to determine how good an Action, A, taken at a particular state, S, is. Q (A, S).
- Temporal Difference: A formula used to find the Q-Value by using the value of current state and action and previous state and action

Reinforcement Learning

Algorithm
Start with a set of environmental states of the agent called as S
A set of possible actions that can take in those states, called as A
Value for each state/action pair that we'll call Q;

At each time step, the agent observes the current state of the environment and selects an action to take based on the Q-value of each possible action in that state.

The Q-value is an estimate of the expected cumulative reward of taking a particular action in a particular state and following the optimal policy thereafter

The Q-values are learned through a process called temporal difference (TD) learning, where the agent updates its estimate of the Q-value based on the difference between the observed reward and its previous estimate of the Q-value.

Reinforcement Learning

The update rule for Q-learning is given by:

$$Q(s, a) = Q(s, a) + \alpha * (r + \gamma * \max(Q(s', a')) - Q(s, a))$$

where Q(s, a) is the estimated Q-value for taking action a in state s, r is the observed reward for taking action a in state s, s' is the resulting state after taking action a, α is the learning rate, and γ is the discount factor.

The max(Q(s', a')) term represents the maximum Q-value for any action a' in the resulting state s'.

Through repeated interactions with the environment, Q-learning converges to the optimal Q-values and optimal policy.

Q-learning is widely used in various applications, such as game playing, robotics, and autonomous driving.

Reinforcement Learning

RL algorithms can be categorized into
value-based,
policy-based, and
actor-critic methods.

Value-based methods learn the optimal value function, which estimates the expected cumulative reward for each state-action pair.

Policy-based methods learn the optimal policy directly, without estimating the value function.

Actor-critic methods combine both value-based and policy-based methods by learning an actor, which selects actions based on a learned policy, and a critic, which estimates the value function.

THANK YOU