REINFORCEMENT LEARNING

UNIT-II

BITS

1. What property is important in reinforcement learning that allows decisions to be made based solely on the current state?

Ans: Markov property

2. The _____is a decision-making process by which it is possible to make the best decisions without reference to a history of prior states.

Ans: Markov Decision Process (MDP)

3. _____ policy returns a probability distribution over actions for each state.

Ans: Stochastic policy

4. _____ function tells how good it is to take a specific action in a state.

Ans: Q (action-value) function

5. _____function tells how good it is to be in a state.

Ans: Value function

6. The _____policy is the best possible policy that maximizes the expected cumulative reward for the agent.

Ans: optimal policy

7. Q-learning update rule Q(s, a) = _____

Ans: Q(s, a) = $\underline{Q(s, a) + \alpha [r + \gamma \max(Q(s', a')) - Q(s, a)]}$

8. If the discount factor (γ) is set close to 0, what will the agent prioritize?

Ans: immediate rewards

9. The _____controls how much the Q-learning algorithm adjusts the Q-value estimates after each update.

Ans: learning rate

10._____learning rate leads to slower learning, but it helps in stabilizing the learning process.

Ans: small learning rate

11. The _____is a parameter that determines how much the agent values future rewards over immediate rewards.

Ans: discount factor

12. _____learning rate can speed up learning but may cause instability.

Ans: large learning rate

13. A discount factor (γ) value close to 1 means the agent will heavily consider _____ rewards.

Ans: future rewards

14. Which of the following modes Gridworld game initializes only the player at a random position, while keeping other objects static?

Ans: random

15. Numpy's built-in _____ function, which takes in an array, finds the largest value in the array, and returns its index position.

Ans: argmax function

16. The neural network model as the Q function to play Gridworld, an output layer that produces a ___ length vector of Q values for each action, given the state.

Ans: fixed-length

17. In the context of experience replay, how are experiences stored?

Ans: In a deque

18. To address instability in DQNs, DeepMind introduced the use of a second network called the _____ network.

Ans: target network

19. A _____is simply where we use a deep learning algorithm as the model in Q-learning.

Ans: Deep Q-Network (DQN)

20. The Q-learning update rule is used to:

Ans: Estimate the future rewards and adjust the Q-values

21. What is the range of values for the discount factor in Q-learning?

Ans: 0 to 1

22. What are the modes available in Gridworld game?

Ans: Static Mode, Random Mode and Player Mode

23. What does the output layer of the Q-network represent?

Ans: Q-values for actions

24. What does experience replay help with in reinforcement learning?

Ans: Preventing catastrophic forgetting

25. What is the main purpose of using a target network in DQN?

Ans: To mitigate instability during training

REINFORCEMENT LEARNING

UNIT-I

BITS

1. Making decisions through interaction with an environment is the primary focus of _____

Ans: Reinforcement learning

2. Components of the RL framework?

Ans: Agent, action, environment, state, reward

3. The learner or decision-maker describes the ____ in the context of reinforcement learning.

Ans: agent

4. To provide an immediate benefit of the agent's action is the purpose of the _____ in reinforcement learning.

Ans: reward

5. The _____is the potentially dynamic conditions in which the agent operates

Ans: environmen

6. Which command is used to reset the environment in OpenAI Gym?

Ans: reset()

7. Which function in PyTorch is used to automatically compute gradients?

Ans: loss.backward()

8. What technique is used to solve complex problems by breaking them down into smaller subproblems?

Ans: Dynamic Programming

9. Dynamic Programming assumes that the agent has _____knowledge of the environment, allowing it to apply a structured approach to solve problems efficiently.

Ans: complete (maximum) knowledge

10. What type of methods involves learning through random trial and error?

Ans: Monte Carlo methods

11. Monte Carlo Methods are more suited for situations where the agent has _____ knowledge of the environment, as they explore the environment through random actions to learn about it.

Ans: limited (minimum) knowledge

12. What is the primary goal of an agent in Reinforcement Learning?

Ans: To maximize cumulative rewards

13. What type of function can deep neural networks approximate in DRL?

Ans: Value functions

14. What type of bandits better models in real-world ad placements?

Ans: Contextual bandits

15. _____ bandits are a type of reinforcement learning problem where an agent must select actions based on contextual information in order to maximize some reward signal.

Ans: Contextual bandits

16. Epsilon-Greedy strategy primarily focus on _____ in Multi-Arm Bandit problem.

Ans: balancing exploration and exploitation

17. Softmax selection policy primarily address _____ in Multi-Arm Bandit problem.

Ans: exploration-exploitation trade-off

18. _____is the output of the softmax function.

Ans: probability distribution

19. _____is the primary purpose of the torch.nn module in PyTorch.

Ans: building blocks (layers) for creating and training neural networks.

20. _____diagrams are a type of flow-like diagram adapted from category theory, a branch of mathematics.

Ans: String Diagrams

21. What is the primary focus of reinforcement learning models represented by string diagrams?

Ans: Agent-environment interactions

22. _____refers to the process of trying out new actions in order to learn more about the environment and potentially discover better policies.

Ans: Exploration

23. What does adjusting the temperature parameter (Ä) in the Softmax Policy control?

Ans: The exploration-exploitation balance

24. What does a high temperature parameter (tau) in the Softmax policy promote?

Ans: Exploration

25. What does the 'requires_grad=True' argument do in PyTorch?

Ans: It tracks and computes gradients