ELTE | FACULTY OF INFORMATICS

REVIEW QUIZE

Deep Reinforcement Learning

Balázs Nagy, PhD

ELTE | IK

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Question 1.

What is the goal of a reinforcement agent?

- Generalize knowledge
- Learn the hidden structure of a dataset
- Maximize long term reward
- Learn from experience

Question 1.

What is the goal of a reinforcement agent?

- Generalize knowledge
- Learn the hidden structure of a dataset
- **Maximize long term reward**
- Learn from experience

Question 2.

What does the Markov property express?

- Changing to a specific state is not reliant on the present state
- Changing to a specific state is reliant only on the present state
- Changing to a specific state is reliant on the present and preceding states
- Changing to a specific state is reliant only on the preceding states

Question 2.

What does the Markov property express?

- Changing to a specific state is not reliant on the present state
- **Changing to a specific state is reliant only on the present state**
- Changing to a specific state is reliant on the present and preceding states
- Changing to a specific state is reliant only on the preceding states

Question 3.

What are the main elements of an MDP?

- Environment, Agent, Value function, Policy, State, Action, Reward
- Model, Agent, State, Reward, Action, Observation, Policy
- Environment, Agent, State, Action, Reward, Model, Policy
- Model, Dynamic of the environment, Policy

Question 3.

What are the main elements of an MDP?

- Environment, Agent, Value function, Policy, State, Action, Reward
- Model, Agent, State, Reward, Action, Observation, Policy
- **Environment, Agent, State, Action, Reward, Model, Policy**
- Model, Dynamic of the environment, Policy

Question 4.

What does it mean if a task is Associative?

- Does not involve learning to act in more than one situation
- Action are taken in more than one situation
- Independent of action taken
- Uses training information that evaluates the action taken

Question 4.

What does it mean if a task is Associative?

- Does not involve learning to act in more than one situation
- **Action are taken in more than one situation**
- Independent of action taken
- Uses training information that evaluates the action taken

Question 5.

What represents $q_*(a)$?

- Estimated action value of action a
- Calculated action value of action a
- Expected action value of action a
- True action value of action a

Question 5.

What represents $q_*(a)$?

- Estimated action value of action a
- Calculated action value of action a
- Expected action value of action a
- **True action value of action a**

Question 6.

- **What does the Law of Large Number states?**
- The average of the results obtained from a large number of trials should be close to the mean value
- The sum of the results obtained from a large number of trials should be close to the cumulated sum
- The average of the results obtained from a large number of trials should be close to the true value
- The average of the results obtained from a large number of trials should be close to the expected value

Question 6.

- **What does the Law of Large Number states?**
- The average of the results obtained from a large number of trials should be close to the mean value
- The sum of the results obtained from a large number of trials should be close to the cumulated sum
- The average of the results obtained from a large number of trials should be close to the true value
- **The average of the results obtained from a large number of trials should be close to the expected value**

Question 7.

Which of the following statements is true?

- At any timestep there is only one greedy action.
- At any timestep there is at least one greedy action.
- At any timestep there are always multiple greedy actions.
- At any timestep sometimes there is no greedy action.

Question 7.

Which of the following statements is true?

- At any timestep there is only one greedy action.
- **At any timestep there is at least one greedy action.**
- At any timestep there are always multiple greedy actions.
- At any timestep sometimes there is no greedy action.

Question 8.

Which of the following statements is true?

- In case of exploitation always a greedy action is selected
- In case of exploration always a greedy action is selected
- In case of exploitation always a non-greedy action is selected
- In case of exploration always a non-greedy action is selected

Question 8.

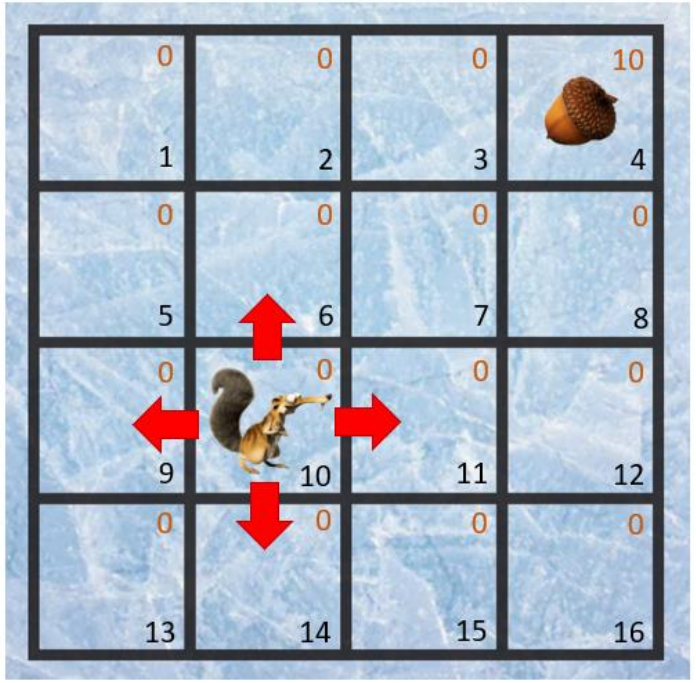
Which of the following statements is true?

- **In case of exploitation always a greedy action is selected**
- In case of exploration always a greedy action is selected
- In case of exploitation always a non-greedy action is selected
- In case of exploration always a non-greedy action is selected

Question 9.

In case of the following stochastic gridworld what cannot be the value of $p(14, 0 \mid 10, \text{up})$?

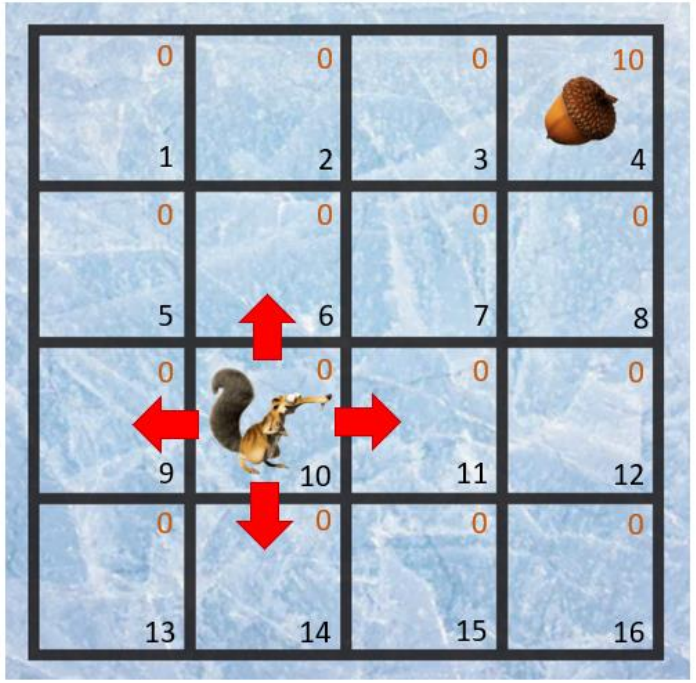
- 1
- 0
- 0.8
- 0.1



Question 9.

In case of the following stochastic gridworld what cannot be the value of $p(14, 0|10, \text{up})$?

- **1**
- 0
- 0.8
- 0.1



Question 10.

What is the Policy evaluation?

- Making the policy greedy with respect to the current value function
- **Making the value function consistent with the current policy**

Question 11.

What is true for DP?

- Utilize bootstrapping and require a model of the environment
- Not utilize bootstrapping and require a model of the environment
- Not utilize bootstrapping and not require a model of the environment
- Utilize bootstrapping and not require a model of the environment

Question 11.

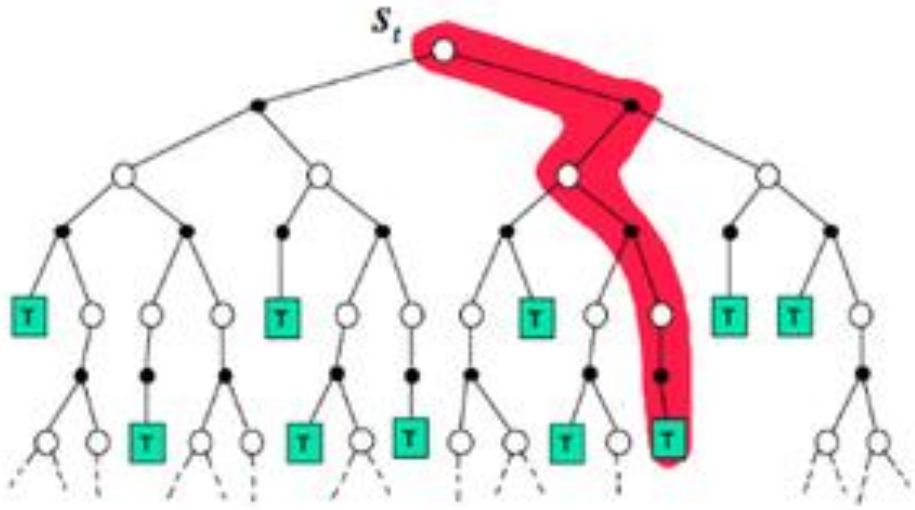
What is true for DP?

- **Utilize bootstrapping and require a model of the environment**
- Not utilize bootstrapping and require a model of the environment
- Not utilize bootstrapping and not require a model of the environment
- Utilize bootstrapping and not require a model of the environment

Question 12.

Which methods backup diagram can be seen in the Figure?

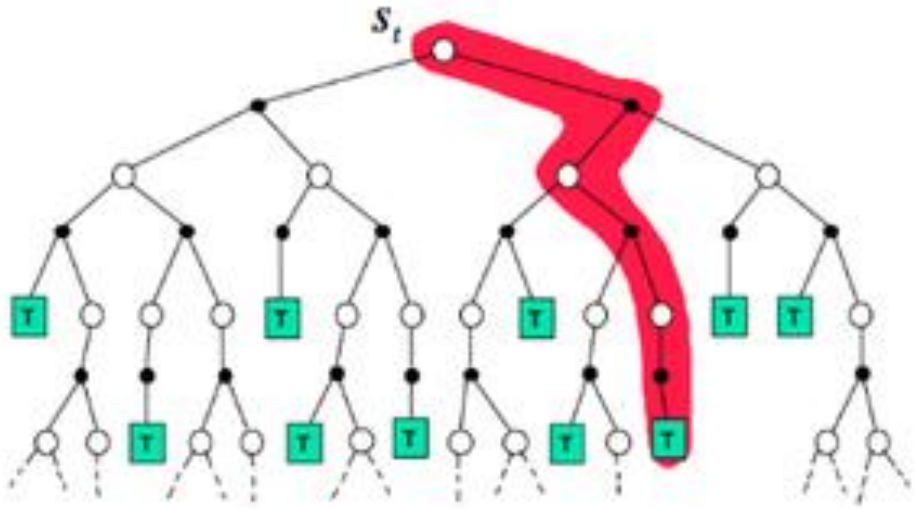
- DP
- MC
- TD
- Sarsa



Question 12.

Which methods backup diagram can be seen in the Figure?

- DP
- **MC**
- TD
- Sarsa



Question 13.

What are the terms used in connection with off-policy prediction?

- Target policy and Behavior policy
- Update policy and Behavior policy
- Target policy and Control policy
- Update policy and Control policy

Question 13.

What are the terms used in connection with off-policy prediction?

- **Target policy and Behavior policy**
- Update policy and Behavior policy
- Target policy and Control policy
- Update policy and Control policy

Question 14.

Which algorithm's value update is the following?

$$V(S_t) \leftarrow V(S_t) + \alpha \Big[R_{t+1} + \gamma V(S_{t+1}) - V(S_t) \Big]$$

- DP
- MC
- TD
- Sarsa

Question 14.

Which algorithm's value update is the following?

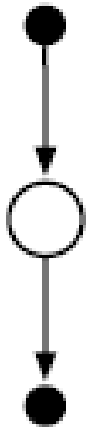
$$V(S_t) \leftarrow V(S_t) + \alpha \Big[R_{t+1} + \gamma V(S_{t+1}) - V(S_t) \Big]$$

- DP
- MC
- **TD**
- Sarsa

Question 15.

Which algorithm's backup diagram is the following?

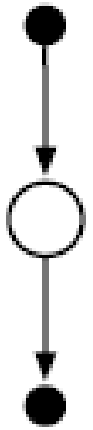
- DP
- MC
- TD
- Sarsa



Question 15.

Which algorithm's backup diagram is the following?

- DP
- MC
- TD
- **Sarsa**



Question 16.

What is the missing component?

$$G_{t:t+n} \doteq R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{n-1} R_{t+n} + \gamma^{n} \; ?$$

- $V_{t+n-1}(S_{t+n})$
- $V_{t+n}(S_{t+n})$
- $V_{t+n}(S_{t+n-1})$
- $V_{t+n-1}(S_{t+n-1})$

Question 16.

What is the missing component?

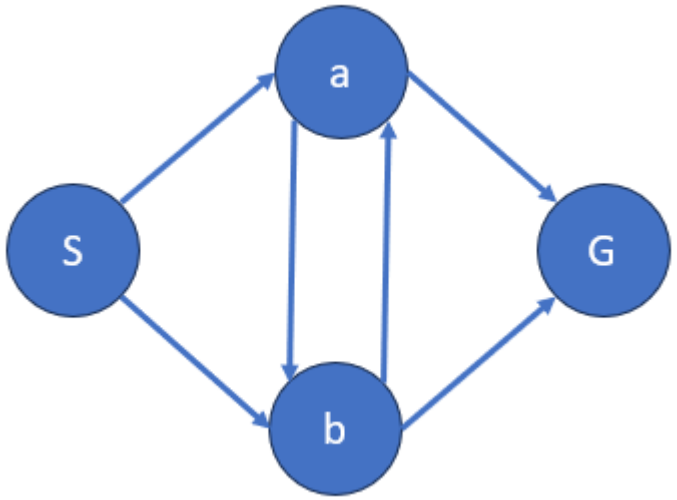
$$G_{t:t+n} \doteq R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{n-1} R_{t+n} + \gamma^{n} V_{t+n-1}(S_{t+n})$$

- **$V_{t+n-1}(S_{t+n})$**
- $V_{t+n}(S_{t+n})$
- $V_{t+n}(S_{t+n-1})$
- $V_{t+n-1}(S_{t+n-1})$

Question 17.

What is the solution that the Depth First Search algorithm provides on the following graph starting from state S and reaching the goal state G?

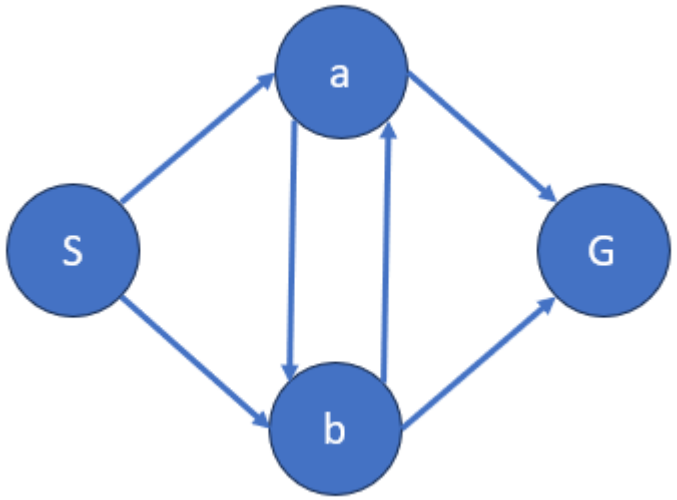
- SaG
- SbG
- SabG
- No solution



Question 17.

What is the solution that the Depth First Search algorithm provides on the following graph starting from state S and reaching the goal state G?

- SaG
- SbG
- SabG
- **No solution**



Question 18.

What is the relation between the dimensionality of weights(w) and the number of states (S) in case of a function approximation?

- S>>d
- S>d
- S<d
- S<<d

Question 18.

What is the relation between the dimensionality of weights(w) and the number of states (S) in case of a function approximation?

- **S>>d**
- S>d
- S<d
- S<<d

Question 19.

Which algorithm is equivalent if the λ parameter is set to 1 in TD(λ)?

- TD(0)
- TD(1)
- MC
- DP

Question 19.

Which algorithm is equivalent if the λ parameter is set to 1 in TD(λ)?

- TD(0)
- TD(1)
- **MC**
- DP

Question 20.

What is λ?

- Discount factor
- Step size parameter
- Trace-decay parameter
- Weight vector

Question 20.

What is λ?

- Discount factor
- Step size parameter
- **Trace-decay parameter**
- Weight vector

ELTE | FACULTY OF INFORMATICS

Thank you for your attention!