

Q1-1: Which of the following statements are correct about RL?

- A. The agent gets rewards according to the state and action.*
- B. It's online in the sense that the data arrive in an online fashion.*
- C. The target of an agent is to maximize the rewards.*
- D. Reinforcement Learning is an unsupervised learning.*

- 1. A, B, C
- 2. B, C, D
- 3. A, C, D
- 4. A, B, C, D

Q1-1: Which of the following statements are correct about RL?

- A. The agent gets rewards according to the state and action.*
- B. It's online in the sense that the data arrive in an online fashion.*
- C. The target of an agent is to maximize the rewards.*
- D. Reinforcement Learning is an unsupervised learning.*

1. A, B, C



2. B, C, D

3. A, C, D

4. A, B, C, D

RL is neither an unsupervised nor a supervised learning.

Q1-2: Select the correct statement.

- A. Markov Assumption implies that given the present state and action, all following states are independent of all past states.*
- B. All Reinforcement Learning techniques adopts Markov assumption property.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q1-2: Select the correct statement.

- A. *Markov Assumption implies that given the present state and action, all following states are independent of all past states.*
- B. *All Reinforcement Learning techniques adopts Markov assumption property.*

- 1. Both the statements are TRUE.
- 2. Statement A is TRUE, but statement B is FALSE.
- 3. Statement A is FALSE, but statement B is TRUE.
- 4. Both the statements are FALSE.



Though markov assumption makes the analysis easier, it's not necessary to assume Markov property.

## Q2-1: Match the following about RL Terminology.

A. reward function	P. is the desired output of an RL algorithm
B. transition probability	Q. quantifies immediate success of agent
C. value function	R. quantifies the likelihood of landing a new state, given a state/action pair
D. optimal policy	S. gives the expected future discounted accumulated reward of a state

1. A - P, B - R, C - S, D - Q
2. A - Q, B - S, C - R, D - P
3. A - Q, B - R, C - S, D - P
4. A - S, B - R, C - Q, D - P

## Q2-1: Match the following about RL Terminology.

A. reward function	P. is the desired output of an RL algorithm
B. transition probability	Q. quantifies immediate success of agent
C. value function	R. quantifies the likelihood of landing a new state, given a state/action pair
D. optimal policy	S. gives the expected future discounted accumulated reward of a state

1. A - P, B - R, C - S, D - Q


2. A - Q, B - S, C - R, D - P

3. A - Q, B - R, C - S, D - P

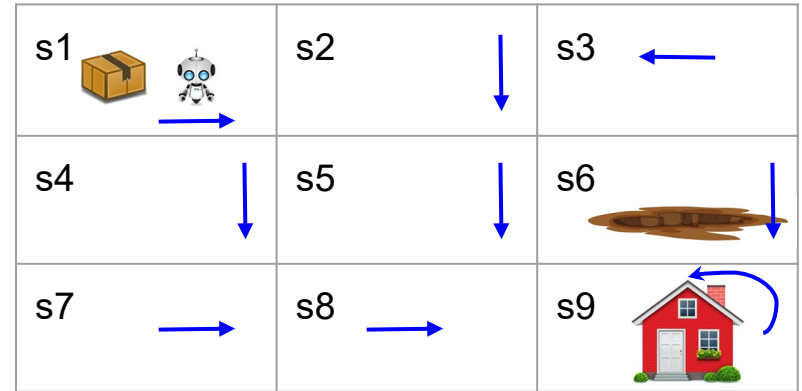


4. A - S, B - R, C - Q, D - P



Q2-2: A robot wants to deliver a package from warehouse at s1 to a home at s9. However, it wants to avoid trench (present at s6). Arrows shows the optimal policy. What is  $V^*(s1)$  approximately? Assume discount factor  $\gamma = 0.8$  and rewards as follows:


- $r(s, a) = -100$  if entering the trench 
- $r(s, a) = +100$  if entering home 
- $r(s, a) = 0$  otherwise

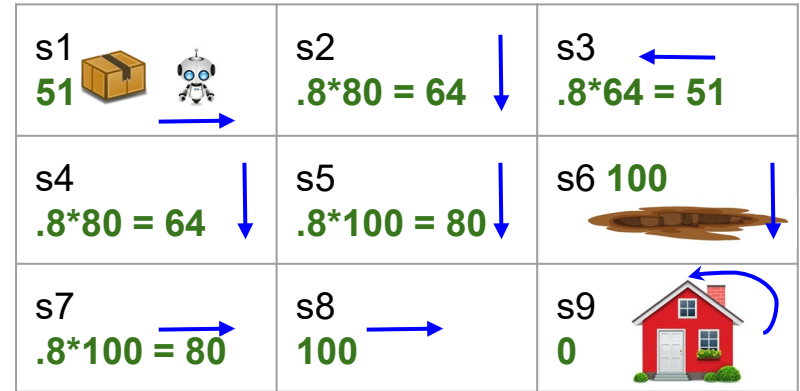
1. 32
2. 51
3. 64
4. 80



Q2-2: A robot wants to deliver a package from warehouse at s1 to a home at s9. However, it wants to avoid trench (present at s6). Arrows shows the optimal policy. What is  $V^*(s1)$  approximately? Assume discount factor  $\gamma = 0.8$  and rewards as follows:

- $r(s, a) = -100$  if entering the trench 
- $r(s, a) = +100$  if entering home 
- $r(s, a) = 0$  otherwise

1. 32
2. 51 
3. 64
4. 80







Q3-1: Select the correct statement.

- A. Q-learning is a “model-free” RL algorithm, i.e., uses no predictions of the environment response.*
  - B. In Q-learning, the agent does not need to know state transition probabilities.*
1. Both the statements are TRUE and statement B is a correct explanation for statement A.
  2. Both the statements are TRUE but statement B is NOT a correct explanation for statement A.
  3. Statement A is TRUE, but statement B is False.
  4. Statement A is FALSE, but statement B is TRUE.

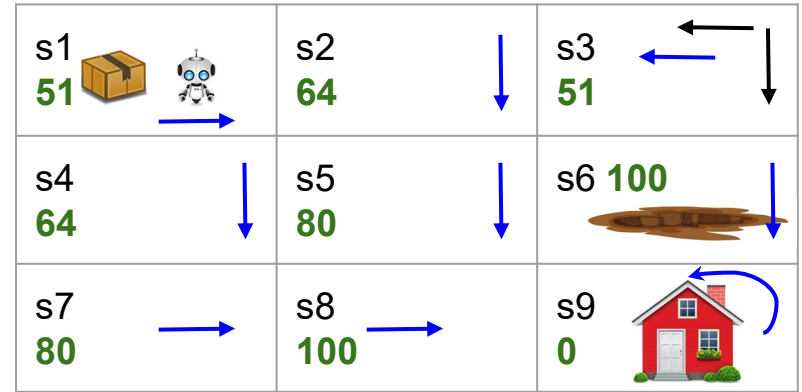
Q3-1: Select the correct statement.

- A. Q-learning is a “model-free” RL algorithm, i.e., uses no predictions of the environment response.*
- B. In Q-learning, the agent does not need to know state transition probabilities.*
1. Both the statements are TRUE and statement B is a correct explanation for statement A. 
  2. Both the statements are TRUE but statement B is NOT a correct explanation for statement A.
  3. Statement A is TRUE, but statement B is False.
  4. Statement A is FALSE, but statement B is TRUE.


Q3-2: A robot wants to deliver a package from warehouse at s1 to a home at s9. However, it wants to avoid trench (present at s6). In the figure, the green numbers are the optimal  $V^*(s)$ , the blue arrows are the optimal policy, and the black arrows are the possible actions from s3. How can you get  $V^*(s3)$  using  $Q(s, a)$ ? Assume discount factor  $\gamma = 0.8$  and rewards as follows:

- $r(s, a) = -100$  if entering the trench 
- $r(s, a) = +100$  if entering home 
- $r(s, a) = 0$  otherwise


1.  $\max \{51, 0\}$
2.  $\max \{51, -20\}$
3.  $\max \{51, -80\}$
4.  $\max \{51, -100\}$



Q3-2: A robot wants to deliver a package from warehouse at s1 to a home at s9. However, it wants to avoid trench (present at s6). In the figure, the green numbers are the optimal  $V^*(s)$ , the blue arrows are the optimal policy, and the black arrows are the possible actions from s3. How can you get  $V^*(s3)$  using  $Q(s, a)$ ? Assume discount factor  $\gamma = 0.8$  and rewards as follows:

- $r(s, a) = -100$  if entering the trench 
- $r(s, a) = +100$  if entering home 
- $r(s, a) = 0$  otherwise

1.  $\max \{51, 0\}$

2.  $\max \{51, -20\}$  

3.  $\max \{51, -80\}$

4.  $\max \{51, -100\}$

$Q(s3, \leftarrow) = 0 + 0.8 * 64 = 51$

$Q(s3, \downarrow) = -100 + 0.8 * 100 = -20$

$V^*(s3) = \max \{Q(s3, \leftarrow), Q(s3, \downarrow)\}$   
 $= \max \{51, -20\}$

