

Поздравляем! Вы прошли тест!

для успешного прохождения 80% или выше

Продолжить обучение

оценка 84.61%

Graded Quiz

оценка последней работы 84.61%

1.	Which approach ensures continual exploration? (Select all that apply)	0/1 балл
	Exploring starts	
	Correct Correct! Exploring starts guarantee that all state-action pairs are visited an infinite number of times in the limit of an infinite number of episodes.	
	\square On-policy learning with a deterministic policy \square On-policy learning with an ϵ -soft policy	
	\checkmark Correct Correct! ϵ -soft policies assign non-zero probabilities to all state-action pairs.	
	Off-Policy learning with an ϵ -soft behavior policy and a deterministic target policy Off-Policy learning with an ϵ -soft target policy and a deterministic behavior policy	
	You didn't select all the correct answers	
2.	When can Monte Carlo methods, as defined in the course, be applied? (Select all that apply)	1/1балл
	 □ When the problem is continuing and there are sequences of states, actions, and rewards □ When the problem is continuing and there is a model that produces samples of the next state and 	
	reward When the problem is episodic and there are sequences of states, actions, and rewards	
	Correct Correct! Well-defined returns are available in episodic tasks.	
	When the problem is episodic and there is a model that produces samples of the next state and reward	
	 Correct Correct! Well-defined returns are available in episodic tasks. 	
3.	Which of the following learning settings are examples of off-policy learning? (Select all that apply) Learning about multiple policies simultaneously while following a single behavior policy	1/1 балл
	✓ Correct Correct! Off-policy learning enables learning about multiple target policies simultaneously using a single behavior policy.	
	✓ Learning the optimal policy while continuing to explore	
	 Correct Correct! An off-policy method with an exploratory behavior policy can assure continual exploration. 	
	✓ Learning from data generated by a human expert	
	Correct Correct! Applications of off-policy learning include learning from data generated by a non-learning agent or human expert. The policy that is being learned (the target policy) can be different from the human expert's policy (the behavior policy).	

4. If a trajectory starts at time t and ends at time T, what is its relative probability under the target policy π and the behavior policy b?

 $\bigcirc \prod_{k=t}^{T-1} \frac{\pi(A_k \mid S_k)}{b(A_k \mid S_k)}$

1 / 1 балл

Suffyr_1 Syr_1 Suffyr_1 Syr_2 Suffy	$\bigcirc \sum_{k=t}^{T-1} \frac{\pi(A_k \mid S_k)}{b(A_k \mid S_k)}$	
ord, (S) NA, (S) NA, (S) NA, (S) NA, (S) NA, (S) ✓ Carrect Correct City to be importance sampling ratio and is used to weight returns in off-policy Monte-carlo Policy (valuation. When is it possible to determine a policy that is greedy with respect to the value functions t ₁ , q, for the policy of defice all that apply) ✓ When state values v ₁ , and a model are available ✓ Carrect Correct(With state values and a model are available ✓ Carrect Correct(With state values and a model are available. ✓ When action values q, and a model are available. ✓ When action values q, and a model are available. ✓ When action values q, and a model are available. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action values are sufficient for choosing the best action in each state. ✓ Carrect Correct(Action methods in Reinforcement Learning sample and average returns much like beared methods anople and average recards. ✓ Carrect Correct(Deep subject of the environment. ✓ Carrect Correct(Deep values are probabilities under a must have a non-zero probability under b. ✓ Carrect Correct(Deep values taken under a must have a non-zero probability under b. ✓ Carrect Correct(Deep values action at for the value of a? ✓ Carrect Correct(Deep values action at for the value of a? ✓ Carrect Correct(Deep values action perform its first update? ✓ Carrect Correct(Deep values action perform its first update? ✓ Carrect Correct(Deep values action perform its first update? ✓ Carrect Correct (Deep values action perform its		
✓ Correct Correct This is the importance sampting ratio and is used to weight returns in off-policy Monte-Carlo Policy Evaluation. When is it possible to determine a policy that is greedy with respect to the value functions e ₁ , q ₂ , for the policy or (Celect all chat apply) ☑ When size values upon an amodel are available. ✓ Correct Correct With state values and a model are available. ☑ When action values q ₂ , and a model are available. ☑ When action values q ₂ , and a model are available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. ☑ When action values q ₂ , are available but no model is available. ✓ Correct Action values are sufficient for choosing the best action in each state. ☑ When action values q ₃ , are available but no model is available. ✓ Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by ○ Accreaging sample returns ○ For coach state a and action q ₄ if for [4] >> 0 then A(a [4] >> 0 ○ Accreases taste a and action q ₄ if for [4] >> 0 then A(a [4] >> 0 ○ Accreases taste a has been visited three times, with corresponding returns 8, 4, and 3. What is the current	$\frac{b(A_{T-1} \mid S_{T-1})}{b(A_{T-1} \mid S_{T-1})}$	
✓ Correct	$\bigcap_{t \in A, t \in S} \frac{\pi(A_t \mid S_t)}{b(A \mid S_t)}$	
When is it possible to determine a policy that is greedy with respect to the value functions v _{x1} q _x for the policy of delect at that apply) When is it possible to determine a policy that is greedy with respect to the value functions v _{x1} q _x for the policy of delect at that apply) When state values v _{x1} and a model are available ✓ correct Correct With state values and a model one can look ahead one step and see which action leads to the best combination of reward and not state. □ When state values v _{x2} are available but no model is available. □ When action values q _{x2} and a model are available ✓ correct Action values q _{x2} are available but no model is available. □ When action values q _{x2} are available but no model is available. □ When action values q _{x2} are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Avaraging sample returns Performing sneeps through the state set Pannings with a model of the environment ✓ correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bands may be a non-zero probability under the behavior policy for a target policy at a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy at a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy at a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy at the state and action a, if fac 1) > 0 then fac 1) > 0 All actions have non-zero probabilities under π ✓ Correct Correct How Monte Carlo estimate for the value of s? 3 15 5 5 5 6 5 3.5 When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once 8 At the end of the first episode	$\nu(\alpha_l \mid s_l)$	
When is it possible to determine a policy that is greedy with respect to the value functions v _e , q, for the policy or? Eelect all that apply) When size values v _e and a model are available ✓ Correct Or best combination of reward and next state. □ When state values v _e are available but no model is available. □ When action values q _e and a model are available ✓ Correct Correct Action values q _e are available but no model is available. ☑ When action values q _e are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Image: Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Beinforcement Learning work by A vierraging sample returns Partorning sweeps through the state set Planning with a model of the environment. ✓ Correct Monte Carlo methods in Beinforcement Learning sample and average returns much like bandit methods sample and average reverds. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy or? For each state a and action a, if fold x) > 0 then fire x) > 0 All ections have non-zero probabilities under x must have a non-zero probability under b. Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3.	✓ Correct	
When is it possible to determine a policy that is greedy with respect to the value functions e ₁ , q, for the policy of Gelect at that apply) When state values e ₂ , and a model are available ✓ correct Correct With state values and a model, one can look ahead one step and see which action leads to the best combination of reward and next state. □ When state values e ₂ , are available but no model is available. □ When action values q ₂ , and a model are available ✓ correct Correct Action values are sufficient for choosing the best action in each state. □ When action values q ₂ are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. □ When action values q ₂ are available but no model is available. ✓ Correct Correct Mone Carlo neithods in Reinforcement Learning work by Acting ging sample returns Performing sweeps through the state set Planning with a model of the environment ✓ correct Correct Mone Carlo methods in Reinforcement Learning sample and average returns much like bands methods sample and average revers. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy gr For each state a and action q ₂ if fa(1) > 0 them fa(1) > 0 All actions have non zero probabilities under π must have a non-zero probability under δ. Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s7 3 15 5 5 Correct Correct Correct The Monce Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? At the ond of the first episode		
the policy ≈? (Select all that apply) When state values v _i , and a model are available ✓ correct Correct With state values and a model, one can look ahead one step and see which action leads to the beat combination of reward and next state. When state values v _i , are available but no model is available. When action values q _i and a model are available ✓ correct Correct Action values are sufficient for choosing the best action in each state. When action values q _i are available but no model is available. ✓ correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Palanning with a model of the environment ✓ correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy are a target p	caro rolley evaluation.	
the policy ≈? (Select all that apply) When state values e, and a model are available ✓ correct Correct With state values and a model, one can look ahead one step and see which action leads to the beat combination of reward and next state. When state values e, are available but no model is available. When action values q, and a model are available. When action values q, and a model are available. ✓ correct Correct Action values are sufficient for choosing the best action in each state. When action values q, are available but no model is available. ✓ correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Parning with a model of the environment ✓ correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy are a target policy are rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy are rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy are a target policy are rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy are rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a farget policy are rewards. **Vision** Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a farget policy are rewards. **Vision** A correct Correct Every action taken under # must have a non-zero probability under b. **Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the st	When is it nossible to determine a policy that is greedy with respect to the value functions $v_{ij}(a)$ for	1/16200
✓ Cerrect Correct With state values and a model, one can look ahead one step and see which action leads to the best combination of reward and next state. □ When state values v ₀ are available but no model is available. ☑ When action values v ₀ are available but no model is available. ☑ When action values v ₀ , and a model are available but no model is available. ☑ When action values v ₀ , are available but no model is available. ☑ When action values v ₀ , are available but no model is available. ✓ Correct Correct Action values v ₀ , are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample returns Performing sweeps through the state set Planning with a model of the environment ✓ Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like build in reinford sample and average returns much like Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy s ² ⑤ For each state a and action a, if M(a s) > 0 then a(a s) > 0 ⑥ For each state a and action a, if A(a s) > 0 then a(a s) > 0 Ø For each state a and action a, if A(a s) > 0 then a(a s) > 0 <td></td> <td>1710</td>		1710
✓ Cerrect Correct With state values and a model, one can look ahead one step and see which action leads to the best combination of reward and next state. □ When state values v ₀ are available but no model is available. ☑ When action values v ₀ are available but no model is available. ☑ When action values v ₀ , and a model are available but no model is available. ☑ When action values v ₀ , are available but no model is available. ☑ When action values v ₀ , are available but no model is available. ✓ Correct Correct Action values v ₀ , are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample returns Performing sweeps through the state set Planning with a model of the environment ✓ Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like build in reinford sample and average returns much like Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy s ² ⑤ For each state a and action a, if M(a s) > 0 then a(a s) > 0 ⑥ For each state a and action a, if A(a s) > 0 then a(a s) > 0 Ø For each state a and action a, if A(a s) > 0 then a(a s) > 0 <td>$ec{oldsymbol{ec{v}}}$ When state values v_π and a model are available</td> <td></td>	$ec{oldsymbol{ec{v}}}$ When state values v_π and a model are available	
Correct With state values and a model, one can look ahead one step and see which action leads to the best combination of reward and next state. □ When action values y_a are available but no model is available. □ When action values y_a and a model are available. □ When action values y_a are available but no model is available. □ When action values y_a are available but no model is available. □ When action values y_a are available but no model is available. □ When action values y_a are available but no model is available. □ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandli methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy b ? For each state b and action b , b (b b > 0 then b (a b) > 0 then b (a) b 0 b 0		
when state values v ₀ , are available but no model is available. when action values q ₀ and a model are available. when action values q ₀ and a model are available. when action values q ₀ are available but no model is available. when action values q ₀ are available but no model is available. when action values q ₀ are available but no model is available. when action values q ₀ are available but no model is available. when action values q ₀ are available but no model is available. when action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Avaraging sample rewards whereaging sample rewards correct Wherea Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy ?? For each state a and action a, if f(a s) > 0 then g(a s) > 0 All actions have non-zero probabilities under π correct Correct Every action taken under π must have a non-zero probability under b. suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of ?? 3 15 5 6 6 After the first time step When does Monte Carlo prediction perform its first update? After the first time step When does Monte Carlo prediction perform its first update? After the first time step When does Monte Carlo selimate for the state value is the average of sample returns observed from that state.		
When action values <i>q_e</i> and a model are available ✓ Correct Correct Action values are sufficient for choosing the best action in each state. When action values <i>q_e</i> are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment ✓ Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bands methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy π? For each state <i>s</i> and action <i>a</i> , if <i>b</i> (<i>a</i> <i>s</i>) > 0 then <i>b</i> (<i>a</i> <i>s</i>) > 0 All actions have non-zero probabilities under π ✓ Correct Correct Every action taken under π must have a non-zero probability under <i>b</i> . Suppose the state <i>s</i> has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of <i>s</i> ? 3 15 5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
When action values <i>q_e</i> and a model are available ✓ Correct Correct Action values are sufficient for choosing the best action in each state. When action values <i>q_e</i> are available but no model is available. ✓ Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment ✓ Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bands methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy π? For each state <i>s</i> and action <i>a</i> , if <i>b</i> (<i>a</i> <i>s</i>) > 0 then <i>b</i> (<i>a</i> <i>s</i>) > 0 All actions have non-zero probabilities under π ✓ Correct Correct Every action taken under π must have a non-zero probability under <i>b</i> . Suppose the state <i>s</i> has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of <i>s</i> ? 3 15 5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
 ✓ correct Correct Action values are sufficient for choosing the best action in each state. ☑ When action values <i>q</i>, are available but no model is available. ✓ correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment ✓ correct	when state values v_π are available but no model is available.	
Correct Action values are sufficient for choosing the best action in each state. When action values q _x are available but no model is available. Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy π? For each state a and action α, if f(α s) > 0 then π(α s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under fb. Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of a? Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	$igspace$ When action values q_π and a model are available	
Correct Action values are sufficient for choosing the best action in each state. When action values q _x are available but no model is available. Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy π? For each state a and action α, if f(α s) > 0 then π(α s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under fb. Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of a? Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	. / Correct	
Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy δ for a target policy π? For each state a and action α, if b(α s) > 0 then π(α s) > 0 For each state a and action α, if π(α s) > 0 then h(α s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under δ. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	•	
Correct Correct Action values are sufficient for choosing the best action in each state. Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy δ for a target policy π? For each state s and action α, if b(α s) > 0 then b(α s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under δ. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	✓ When action values a₂ are available but no model is available	
Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	When action values q_π are available but no model is available.	
Monte Carlo methods in Reinforcement Learning work by Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bands methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy for a target policy π? For each state s and action a, if $f(a \mid s) > 0$ then $f(a \mid s) > 0$ All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	✓ Correct	
Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if t(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then h(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Correct! Action values are sufficient for choosing the best action in each state.	
Averaging sample rewards Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if t(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then h(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
 Averaging sample returns Performing sweeps through the state set Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state a and action a, if ta(a s) > 0 then π(a s) > 0 For each state a and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state a has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of a? 3 15 5 3.5 Correct Correct Correct If Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode 	Monte Carlo methods in Reinforcement Learning work by	1/1 балл
Performing sweeps through the state set Planning with a model of the environment ✓ Correct Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π ✓ Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 ✓ Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Averaging sample rewards	
Planning with a model of the environment Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
 ✓ Correct Correct Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 All actions have non-zero probabilities under π ✓ Correct	Performing sweeps through the state set	
Correct! Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.3.5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Planning with a model of the environment	
Correct! Monte Carlo methods in Reinforcement Learning sample and average returns much like bandit methods sample and average rewards. Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy b for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π Correct Correct! Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.3.5 Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	✓ Correct	
Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior policy of for a target policy π? For each state s and action a, if b(a s) > 0 then π(a s) > 0 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 When does Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Correct! Monte Carlo methods in Reinforcement Learning sample and average returns much like	
policy <i>b</i> for a target policy <i>π</i> ? For each state <i>s</i> and action <i>a</i> , if <i>b</i> (<i>a</i> <i>s</i>) > 0 then <i>π</i> (<i>a</i> <i>s</i>) > 0 For each state <i>s</i> and action <i>a</i> , if <i>π</i> (<i>a</i> <i>s</i>) > 0 then <i>b</i> (<i>a</i> <i>s</i>) > 0 All actions have non-zero probabilities under <i>π</i> Correct Correct Every action taken under <i>π</i> must have a non-zero probability under <i>b</i> . Suppose the state <i>s</i> has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of <i>s</i> ? 3 15 5 3.5 Vecrect Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	bandit metnods sample and average rewards.	
policy <i>b</i> for a target policy <i>π</i> ? For each state <i>s</i> and action <i>a</i> , if <i>b</i> (<i>a</i> <i>s</i>) > 0 then <i>π</i> (<i>a</i> <i>s</i>) > 0 For each state <i>s</i> and action <i>a</i> , if <i>π</i> (<i>a</i> <i>s</i>) > 0 then <i>b</i> (<i>a</i> <i>s</i>) > 0 All actions have non-zero probabilities under <i>π</i> Correct Correct Every action taken under <i>π</i> must have a non-zero probability under <i>b</i> . Suppose the state <i>s</i> has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of <i>s</i> ? 3 15 5 3.5 Verrect Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Which of the following is a requirement for using Monte Carlo policy evaluation with a behavior	1/16200
 For each state s and action a, if π(a s) > 0 then b(a s) > 0 All actions have non-zero probabilities under π ✓ Correct Correct Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 ✓ Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode 		17 1 08351
 All actions have non-zero probabilities under π ✓ Correct Correct! Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 ✓ Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode 	\bigcirc For each state s and action a , if $b(a\mid s)>0$ then $\pi(a\mid s)>0$	
 ✓ Correct Correct! Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 ✓ Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode 	$ igotimes $ For each state s and action a , if $\pi(a \mid s) > 0$ then $b(a \mid s) > 0$	
Correct! Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	$igcirc$ All actions have non-zero probabilities under π	
Correct! Every action taken under π must have a non-zero probability under b. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s? 3 15 5 3.5 Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	·	
current Monte Carlo estimate for the value of s? 3 15 5 3.5 ✓ Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
current Monte Carlo estimate for the value of s? 3 15 5 3.5 ✓ Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Suppose the state s has been visited three times, with corresponding returns $8,4$, and 3 . What is the	1/1 балл
 ↓ 15 ⑤ 5 ҈ 3.5 ✓ Correct Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once ⑥ At the end of the first episode 		
 	○ 3	
 3.5 ✓ Correct Correct The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode 		
Correct Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode		
Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	○ 3.5	
Correct! The Monte Carlo estimate for the state value is the average of sample returns observed from that state. When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	✓ Correct	
When does Monte Carlo prediction perform its first update? After the first time step When every state is visited at least once At the end of the first episode	Correct! The Monte Carlo estimate for the state value is the average of sample returns observed	
After the first time step When every state is visited at least once At the end of the first episode	from that state.	
After the first time step When every state is visited at least once At the end of the first episode	When does Monte Caule prediction parties in first and does	
When every state is visited at least once At the end of the first episode		1/1 балл
At the end of the first episode		
✓ Correct	At the efficient episode	
¥	✓ Correct	

10. In Monte Carlo prediction of state-values, memory requirements depend on (select all that apply)	1 / 1 балл
✓ The number of states	
Correct Correct! Monte Carlo Prediction needs to store the estimated value for each state.	
☐ The number of possible actions in each state	
✓ The length of episodes	
Correct Correct! Monte Carlo Prediction needs to store the sequence of states and rewards. during an episode	
For Monte Carlo Prediction of state-values, the number of updates at the end of an episode depend on	S (1/1 балл
The number of states	
The length of the exicate	
The length of the episode	
Correct Correct! Monte Carlo Prediction updates the estimated value of each state visited during the episode.	
12. Which approach can find an optimal deterministic policy? (select all that apply)	0/1балл
Exploring Starts	
\bigcirc ϵ -greedy exploration	
extstyle ext	
Correct Correct In this case, the behavior policy can maintain exploration while the target policy is deterministic.	
You didn't select all the correct answers	
13. In an ϵ -greedy policy over $\mathcal A$ actions, what is the probability of the highest valued action if there are no other actions with the same value?	е (1/1 балл
○ 1 − e	
$\bigcirc \epsilon$ $\bigcirc 1 - \epsilon + \frac{\epsilon}{A}$	
○ £ A	
- 4	
 Correct Correct! The highest valued action still has a chance of being selected as an exploratory action. 	