Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653103759
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 148 Question Id : 640653699353 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DEGREE LEVEL :REINFORCEMENT LEARNING (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?

CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406532335295. ✔ YES

6406532335296. ✖ NO

Question Number : 149 Question Id : 640653699354 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

Note:

1. Always enter your answer correct upto 3 decimal places without rounding off for numerical questions.

Options :

6406532335297. ✔ Useful Data has been mentioned above.

6406532335298. ✖ This data attachment is just for a reference & not for an evaluation.

Sub-Section Number :	2
Sub-Section Id :	640653103760
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 150 Question Id : 640653699355 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

Assertion: Taking exploratory actions is important for RL agents

Reason: If the rewards obtained for actions are stochastic, an action which gave a high reward once, might give lower reward next time.

Options :

6406532335299. ✖ Assertion and Reason are both true and Reason is a correct explanation of Assertion.

6406532335300. ✔ Assertion and Reason are both true and Reason is not a correct explanation of Assertion.

6406532335301. ✖ Assertion is true and Reason is false

6406532335302. ✖ Both Assertion and Reason are false.

Question Number : 151 Question Id : 640653699367 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

Match the methods with their corresponding characteristics.

Options :

6406532335329. ✔ DP: bootstraps, full backups
MC: does not bootstrap, full backups
TD: bootstraps, full backups

6406532335330. ✖ DP: bootstraps, full backups
MC: does not bootstrap, sample backups
TD: bootstraps, sample backups

6406532335331. ✖ DP: bootstraps, sample backups
MC: bootstraps, sample backups
TD: does not bootstrap, full backups

6406532335332. ✖ DP: does not bootstrap, full backups
MC: bootstraps, full backups
TD: bootstraps, sample backups

6406532335333. ✖ None of these

Question Number : 152 Question Id : 640653699379 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

Which of the following is correct formulation of advantage function(A)?

Options :

6406532335383. ✔ $q_\pi = A_\pi(s,a) + v_\pi(s)$

6406532335384. ✖ $A_\pi(s,a) = q_\pi + v_\pi(s)$

6406532335385. ✖ $v_\pi(s) = q_\pi + A_\pi(s,a)$

6406532335386. ✖ None of these.

Question Number : 153 Question Id : 640653699382 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

Which of the following is the most suitable way to solve for highest reusability?

Options :

6406532335398. ✖ Hierarchical optimal

6406532335399. ✖ Flat Optimal.

6406532335400. ✔ Recursive optimal.

6406532335401. ✖ None of these

Sub-Section Number :	3
Sub-Section Id :	640653103761
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 154 Question Id : 640653699360 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following assertion-reason pair:

Assertion: Monte Carlo value function approximation methods need knowledge of model to be implemented.

Reason: Monte Carlo value function approximation methods require a way to sample trajectories from the environment.

Options :

6406532335309. ✖ Assertion and Reason are both true and Reason is a correct explanation of Assertion.

6406532335310. ✖ Assertion and Reason are both true and Reason is not a correct explanation of Assertion.

6406532335311. ✖ Assertion is true but Reason is false.

6406532335312. ✔ Assertion is false but Reason is true.

Question Number : 155 Question Id : 640653699368 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is the correct definition of accumulating eligibility traces?

Options :

6406532335334. ✔ $$e_t(s) = \begin{cases} \gamma\lambda e_{t-1}(s), & \text{if } s \neq s_t. \\ \gamma\lambda e_{t-1}(s) + 1, & \text{if } s = s_t. \end{cases}$$

6406532335335. ✖ $$e_t(s) = \begin{cases} \lambda e_{t-1}(s), & \text{if } s \neq s_t. \\ \lambda e_{t-1}(s) + 1, & \text{if } s = s_t. \end{cases}$$

6406532335336. ✖ $$e_t(s) = \begin{cases} \gamma e_{t-1}(s), & \text{if } s \neq s_t. \\ \gamma e_{t-1}(s) + 1, & \text{if } s = s_t. \end{cases}$$

6406532335337. ✖ $$e_t(s) = \begin{cases} \lambda e_{t-1}(s), & \text{if } s \neq s_t. \\ \lambda e_{t-1}(s) \cdot 10, & \text{if } s = s_t. \end{cases}$$

Question Number : 156 Question Id : 640653699370 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following statement(s) are correct about the TD(λ) algorithm?

Options :

6406532335342. ✖ For λ = 1, the algorithm will behave like SARSA.

6406532335343. ✖ For λ = 1, the algorithm will behave like Q-learning.

6406532335344. ✖ For λ = 1, the algorithm will behave like Double Q-learning.

6406532335345. ✔ For λ = 1, the algorithm will behave like Monte Carlo and will update incrementally instead of waiting for the trajectory to end.

6406532335346. ✖ None of these

Question Number : 157 Question Id : 640653699371 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

In the naive method of state aggregation such as we have in a grid world, the assumption that states that are close together have similar values breaks down for which of the following

scenarios?

Options :

6406532335347. ✖ A pair of states that are in the middle of some cell.

6406532335348. ✔ A pair of states that are close to the border with both of them lying in the same cell.

6406532335349. ✖ A pair of states that are close to the border with both of them lying in different cells.

6406532335350. ✖ None of these

Question Number : 158 Question Id : 640653699372 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following learning algorithms exhibit maximization bias?

Options :

6406532335351. ✔ Q-learning

6406532335352. ✖ Double-Q learning

6406532335353. ✖ SARSA

6406532335354. ✖ Expected SARSA

6406532335355. ✖ TD(λ) with eligibility traces

6406532335356. ✖ None of these

Question Number : 159 Question Id : 640653699373 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

As a consequence of the proposed fix for the non-stationarity of the target in DQN, the target network changes at a _____ rate than the online network.

Options :

6406532335357. ✖ faster

6406532335358. ✔ slower

6406532335359. ✖ None of these

Question Number : 160 Question Id : 640653699374 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is the correct update rule for θ (representing the policy) with the policy gradient method?

Options :

6406532335360. ✔ $\theta \leftarrow \theta + \alpha \nabla J(\theta)$

6406532335361. ✖ $\theta \leftarrow \theta - \alpha \nabla J(\theta)$

6406532335362. ✖ $\theta \leftarrow \theta + \nabla J(\theta)$

6406532335363. ✖ $\theta \leftarrow \theta - \nabla J(\theta)$

6406532335364. ✖ None of these

Question Number : 161 Question Id : 640653699375 Question Type : MCQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is the correct formulation of cost function for multi arm bandit problem?

Options :

6406532335365. ✔ $J(\theta) = \sum_a q^*(a)\pi_\theta(a)$

6406532335366. ✖ $J(\theta) = \sum_a q(a)\pi_\theta(a)$

6406532335367. ✖ $J(\theta) = q^*(a)\pi_\theta(a)$

6406532335368. ✖ $J(\theta) = \sum_a q^*(a)\pi(a)$

Question Number : 162 Question Id : 640653699376 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is the correct formulation to estimate cost function for multi arm bandit problem from N samples? r_i is the reward received after pulling an arm at i^{th} timestamp.

Options :

6406532335369. ✔ $\hat{\nabla} J(\theta) = \frac{1}{N} \sum_{i=1}^{N} r_i \frac{\nabla \pi_a(\theta)}{\pi_a(\theta)}$

6406532335370. ✖ $\hat{\nabla} J(\theta) = \frac{1}{N} \sum_{i=1}^{N} r_i \nabla \pi_a(\theta)$

6406532335371. ✖ $\hat{\nabla} J(\theta) = \sum_{i=1}^{N} r_i \frac{\nabla \pi_a(\theta)}{\pi_a(\theta)}$

6406532335372. ✖ $\hat{\nabla} J(\theta) = \frac{1}{N} \sum_{i=1}^{N} \frac{\nabla \pi_a(\theta)}{\pi_a(\theta)}$

6406532335373. ✖ $\hat{\nabla} J(\theta) = \frac{1}{N} \sum_{i=1}^{N} r_i \frac{\pi_a(\theta)}{\nabla \pi_a(\theta)}$

Question Number : 163 Question Id : 640653699377 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is correct formulation of policy gradient theorem with baseline?

Options :

6406532335374. ✔ $\nabla J(\theta) \propto \sum_s \mu(s) \sum_a (q_\pi(s,a) - b(s)) \nabla \pi(a|s,\theta)$

6406532335375. ✖ $\nabla J(\theta) \propto \sum_s \mu(s) \sum_a (v_\pi(s) - b(s)) \nabla \pi(a|s,\theta)$

6406532335376. ✖ $\nabla J(\theta) \propto \sum_a \mu(s) \sum_s (q_\pi(s,a) - b(s)) \nabla \pi(a|s,\theta)$

6406532335377. ✖ $\nabla J(\theta) \propto \sum_s \mu(s) \sum_a (q_\pi(s,a) - b) \nabla \pi(a|s,\theta)$

6406532335378. ✖ None of these

Question Number : 164 Question Id : 640653699385 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Choose the correct formula for UCT:

Options :

6406532335410. ✖ $\text{argmax}_{a \in A(s)} \left[Q(s,a) - 2c\sqrt{\frac{2lnN(s)}{N(s,a)}} \right]$

6406532335411. ✔ $\text{argmax}_{a \in A(s)} \left[Q(s,a) + 2c\sqrt{\frac{2lnN(s)}{N(s,a)}} \right]$

6406532335412. ✖ $\text{argmax}_{a \in A(s)} \left[Q(s,a) + 2c\sqrt{\frac{2lnN(s,a)}{N(s)}} \right]$

6406532335413. ✖ $\max_{s'} \left[v(s') + 2c\sqrt{\frac{2lnN(s')}{N(s',a)}} \right]$

Question Number : 165 Question Id : 640653699386 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is the correct formulation of an RL agent to maximize influence in an unknown network?
Note:
G_t: the sub-graph discovered after t queries.
G_T: discovered graph.
G^*: Entire graph.
$I_{G^*}(G_T)$: Number of nodes influenced with the discovered graph.

Options :

6406532335414. ✖ **State**: The discovered graph G_t.
Action: all nodes in G_t that have been queried.
Reward: at the end based on the number of nodes influenced in G^* after discovering graph G_T (denoted by $I_{G^*}(G_T)$)

6406532335415. ✖ **State**: The complete graph G_t.
Action: all nodes in G_t that have not been queried.
Reward: at the end based on the number of nodes influenced in G^* after discovering graph G_T (denoted by $I_{G^*}(G_T)$)

6406532335416. ✖ **State**: The undiscovered graph G_t.
Action: all nodes in G_t that have not been queried.
Reward: at every step, small reward proportional to number of nodes discoverd.

6406532335417. ✔ **State**: The undiscovered graph G_t.
Action: all nodes in G_t that have not been queried.
Reward: at the end based on the number of nodes influenced in G^* after discovering graph G_T (denoted by $I_{G^*}(G_T)$)

6406532335418. ✖ None of these

Sub-Section Number : 4

Sub-Section Id : 640653103762

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 166 Question Id : 640653699359 Question Type : MSQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

If π and π' are two policies for an MDP both of which are optimal, which of the following statements is true? Assume that the MDP has more than two optimal policies.

Options :

6406532335305. ✔ The value functions for both π and π' are equal and each of them is equal to the optimal value function.

6406532335306. ✖ At least one of π or π' has to be deterministic.

6406532335307. ✔ For state s, $a_1, a_2 \in A(s)$, if $q_\pi(s, a_1) = q_{\pi'}(s, a_2)$, then $\pi(a_1|s) = \pi'(a_2|s)$

6406532335308. ✖ None of these

Question Number : 167 Question Id : 640653699361 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider Monte-Carlo approach for policy evaluation. Suppose the states are $S_1, S_2, S_3, S_4, S_5, S_6$ and *terminal_state*. You sample one trajectory as follows - $S_1 \rightarrow S_3 \rightarrow S_5 \rightarrow S_2 \rightarrow terminal\_state$.
Which among the following states can be updated from this sample?

Options :

$S_1 \rightarrow S_3 \rightarrow S_5 \rightarrow S_2 \rightarrow S_1$

S_4 ✗ S_6 ✗

6406532335313. ✔ S_1

6406532335314. ✔ S_2

6406532335315.

✖ S_6

6406532335316. ✖ S_4

6406532335317. ✖ None of these

Question Number : 168 Question Id : 640653699363 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose the correct qualifiers for SARSA from the options given below.

Options :

6406532335322. ✔ On-policy

6406532335323. ✖ Off-policy

6406532335324. ✔ TD-control

6406532335325. ✖ TD-prediction

6406532335326. ✖ None of these

Question Number : 169 Question Id : 640653699378 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

What are the correct changes you have to make in the standard policy gradient theorem to come up with MC policy gradient with baseline update rule?

Options :

6406532335379. ✔ $\mu(s)$ is used to sample the states according to policy π.

6406532335380. ✔ Sample actions from policy π.

6406532335381. ✖ Sample states and actions from policy π.

6406532335382. ✔ Replace $q_\pi(S_t, A_t)$ with G_t

Question Number : 170 Question Id : 640653699380 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose the correct statement regarding the A3C algorihtm:

Options :

6406532335387. ✔ It uses parallel processing.

6406532335388. ✖ It is sequential in natures.

6406532335389. ✔ All the individual learners have to produce only small updates to the gradient.

6406532335390. ✖ All the individual learners can produce small and large updates to the gradient.

6406532335391. ✔ It uses n-step TD error to update the policy parameters.

6406532335392. ✖ None of these

Question Number : 171 Question Id : 640653699384 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following are characteristics of MCTS?

Options :

6406532335406. ✔ Aheuristic.

6406532335407. ✔ Anytime.

6406532335408. ✖ Symmetric.

6406532335409. ✖ None of these.

Sub-Section Number :	5
Sub-Section Id :	640653103763
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 172 Question Id : 640653699362 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Select the correct statements from the options below:

Options :

6406532335318. ✔ Asynchronous DP is a type of generalized policy iteration.

6406532335319. ✖ Value iteration algorithm is not a type of generalized policy iteration.

6406532335320. ✔ Policy iteration algorithm is a type of generalized policy iteration.

6406532335321. ✔ If an algorithm is some form of generalized policy iteration, it is guaranteed to converge to an optimal policy.

Question Number : 173 Question Id : 640653699369 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following are correct about the notion of eligibility traces?

Options :

6406532335338. ✔ It is the backward view of the TD(λ) algorithm.

6406532335339. ✖ It is the forward view of the TD(λ) algorithm.

6406532335340. ✔ The motivation behind eligibility traces is, if an agent receives a reward/punishment at any time step, then which decisions in the past are eligible to get credit for it.

6406532335341. ✖ None of these

Question Number : 174 Question Id : 640653699381 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose the correct components of an option:

Options :

6406532335393. ✔ Initiation.

6406532335394. ✔ Policy.

$O(I_o, \Pi_o, T_o)$

6406532335395. ✔ Termination.

6406532335396. ✖ Set of actions.

6406532335397. ✖ None of these

Question Number : 175 Question Id : 640653699383 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following is true about Markov and Semi Markov Options?

Options :

6406532335402. ✔ In a Markov Option the option’s policy depends only on the current state.

6406532335403. ✔ In a Semi Markov Option the option’s policy can depend only on the current state.

6406532335404. ✔ In a Semi Markov Option, the option’s policy may depend on the history since the execution of the option began.

6406532335405. ✖ A Semi-Markov Option is always a Markov Option but not vice versa.

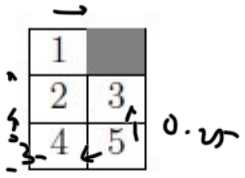
Sub-Section Number :	6
Sub-Section Id :	640653103764
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653699356 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (176 to 177)

Question Label : Comprehension

Consider following grid world:



- All transitions cost -1 reward.
- The agent can take 4 actions i.e. $\{left, right, up, down\}$.
  An action that takes the agent outside of the grid world,
  leaves the state unchanged.
- All transitions are deterministic.
- Gray cell represents terminal state.
- Discounting factor $\gamma = 1$
- π^*, $v_{\pi^*}(s)$ and $q_{\pi^*}(s, a)$ represent optimal policy and corresponding state and action value functions, respectively.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 176 Question Id : 640653699357 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the 2 step value function for state 2, computed in synchronous manner, following equiprobable random policy?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

-2

Question Number : 177 Question Id : 640653699358 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the 3 step value function for state 5, computed in synchronous manner, following equiprobable random policy?

Response Type : Numeric

Evaluation Required For SA : Yes

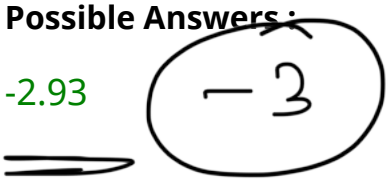
Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

-2.93



Question Id : 640653699364 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (178 to 179)

Question Label : Comprehension

Consider an MDP with a single nonterminal state and a single action that transitions back to the nonterminal state with probability p and transitions to the terminal state with probability $1 - p$. Let the reward be $+1$ on all transitions, and let $\gamma = 1$. Suppose you observe one episode that lasts 10 steps, with a return of 10.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 178 Question Id : 640653699365 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the first-visit estimator of the value of the nonterminal state?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

10 10

Question Number : 179 Question Id : 640653699366 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the every-visit estimator of the value of the nonterminal state?

Response Type : Numeric

Evaluation Required For SA : Yes

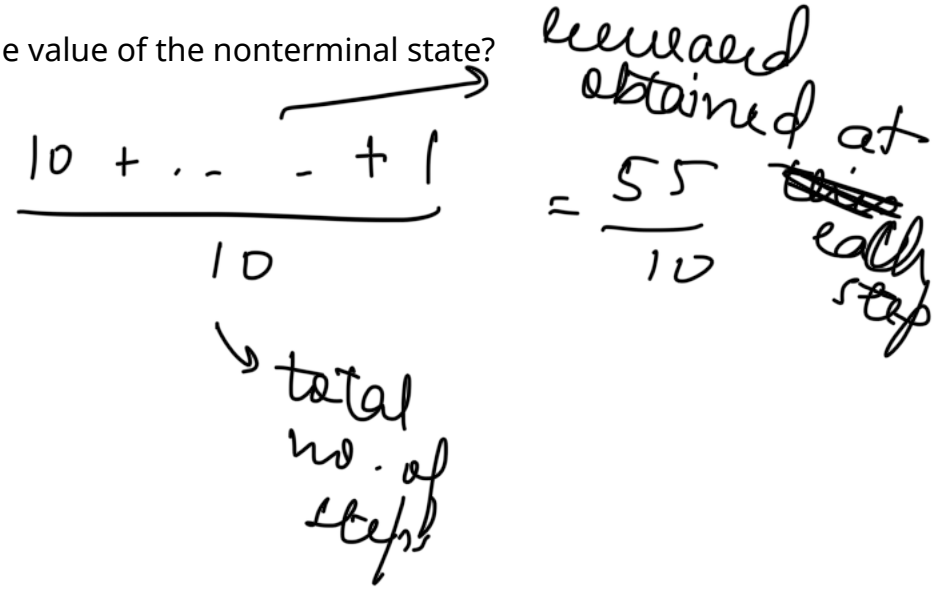
Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

5.5



Intro to BigData

Section Id :	64065349327
Section Number :	8
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	24
Number of Questions to be attempted :	24
Section Marks :	50