Unit 4 - Week 2

Assignment 2

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2020-02-12, 23:59 IST.

1) Which of the following happens when using the UCB algorithm? Assume infinite time case. *1 point*

- Action with highest Q value will be chosen in every iteration
- A certain action may never be chosen again after a certain amount of time
- The confidence interval of a certain action decreases in size every time that action is not chosen
- The confidence interval of a certain action increases in size every time that action is chosen

No, the answer is incorrect.
Score: 0
Accepted Answers:
A certain action may never be chosen again after a certain amount of time

2) In UCB, the term $\sqrt{\frac{2\ln(n)}{n_j}}$ is added to each arm's Q value and the arm with the highest value after adding the 2 terms is chosen. Which one of the following would be an effect of adding $\sqrt{\frac{\ln(n)}{n_j}}$ instead of $\sqrt{\frac{2\ln(n)}{n_j}}$? *1 point*

- Sub-optimal arms would be chosen less frequently
- Sub-optimal arms would be chosen more frequently
- Makes no change on the frequency of picking sub-optimal arms
- Can't say any of the above

No, the answer is incorrect.
Score: 0
Accepted Answers:
Sub-optimal arms would be chosen less frequently

3) In a 4-arm bandit problem, after executing 100 iterations of UCB1 algorithm, the estimates of Q values are- $Q_{100}(1) = 1.73, Q_{100}(2) = 0.83, Q_{100}(3) = 1.19, Q_{100}(4) = 1.25$ and the number of times each of them are sampled are- $n_1 = 30, n_2 = 25, n_3 = 25, n_4 = 20$. Which arm will be sampled in next iteration? *1 point*

- Arm 1
- Arm 2
- Arm 3
- Arm 4

No, the answer is incorrect.
Score: 0
Accepted Answers:
Arm 1

4) We need 200 samples for getting an $(\epsilon, \delta) - PAC$ arm using naive $(\epsilon, \delta) - PAC$ algorithm in a 10-arm bandit problem with certain values of ϵ and δ. Now, the epsilon is halved keeping delta unchanged. How many samples would be needed to re-run naive $(\epsilon, \delta) - PAC$ algorithm? *1 point*

- 400
- 800
- 1600
- 100

No, the answer is incorrect.
Score: 0
Accepted Answers:
800

5) **Assertion:** The confidence bound of each arm in UCB1 algorithm cannot increase with iterations. *1 point*
Reason: The n_j term in denominator ensures that the confidence bound remains the same for unselected arms and decreases for selected arm.

- Assertion and Reason are both true and Reason is a correct explanation of Assertion
- Assertion and Reason are both true and Reason is not a correct explanation of Assertion
- Assertion is true and Reason is false
- Both Assertion and Reason are false

No, the answer is incorrect.
Score: 0
Accepted Answers:
Both Assertion and Reason are false

6) In UCB 1 Theorem proof when $l = \lceil 8 \ln m/\Delta_i^2 \rceil$, $q_*(a^*) - q_*(i) - 2C_{m,l} = 0$. Which of the following correctly explains $q_*(a^*) - q_*(i) - 2C_{m,T_i(s_i)}$ being greater than 0 for the same l. Note: $C_{m,l} = \sqrt{2\ln(m)/l}$.$q_*(a^*) - q_*(i) = \Delta_i$ *1 point*

- $T_i(s_i) < l$ thus $C_{m,T_i(s_i)} < C_{m,l}$
- $T_i(s_i) > m$ thus $C_{m,T_i(s_i)} < C_{m,l}$.
- $T_i(s_i) > l$ thus $C_{m,T_i(s_i)} < C_{m,l}$
- $T_i(s_i) < l$ thus $C_{m,T_i(s_i)} > C_{m,l}$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$T_i(s_i) > l$ thus $C_{m,T_i(s_i)} < C_{m,l}$

7) **Assertion:** If a sub optimal arm is replaced by some other arm whose expected reward is farther away from the optimal arm then the bound on regret($8 \sum_{i \neq a^*} \ln n/\Delta_i + (1 + \pi^2/3) \sum_{j=1}^{K} \Delta_i$) for UCB 1 algorithm will decrease(for $n > 3$) *1 point*
Reason: As the reward difference increases the arm will be selected more often

- Assertion and Reason are both true and Reason is a correct explanation of Assertion
- Assertion and Reason are both true and Reason is not a correct explanation of Assertion
- Assertion is true and Reason is false
- Both Assertion and Reason are false

No, the answer is incorrect.
Score: 0
Accepted Answers:
Assertion is true and Reason is false

8) In median elimination method of (ϵ, δ) PAC bounds, what should be the update rule for ϵ and δ if $\epsilon_1 = \frac{\epsilon}{3}$ and $\delta = \frac{\delta}{4}$? *1 point*

- $\epsilon_{l+1} = \frac{\epsilon_l}{2}$ and $\delta_{l+1} = \frac{\delta_l}{2}$
- $\epsilon_{l+1} = \frac{3\epsilon_l}{4}$ and $\delta_{l+1} = \frac{\delta_l}{2}$
- $\epsilon_{l+1} = \frac{2\epsilon_l}{3}$ and $\delta_{l+1} = \frac{3\delta_l}{4}$
- $\epsilon_{l+1} = \frac{2\epsilon_l}{3}$ and $\delta_{l+1} = \frac{\delta_l}{2}$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\epsilon_{l+1} = \frac{2\epsilon_l}{3}$ *and* $\delta_{l+1} = \frac{3\delta_l}{4}$

9) In median elimination method for (ϵ, δ) PAC bounds, we claim that for every phase l, for an event E_l, $Pr[E_l] > 1 - \delta_l$ *1 point*
Follow the notations as
S_l -- is the set of arms remaining in the l^{th} phase
A -- is the maximum of rewards of true best arm in S_l, i.e. in l^{th} phase
B -- is the maximum of rewards of true best arm in S_{l+1}, i.e. in $l + 1^{th}$ phase
Which of the following is E_l?

- $A + \epsilon_l \leq B$
- $A \leq B + \epsilon_l$
- $A \geq B$
- $A \geq B + \epsilon_l$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$A \leq B + \epsilon_l$

10) Which of the following statements is NOT true about Thompson Sampling or Posterior Sam- pling? *1 point*

- After each sampling of an arm, the distributions for the arm becomes thinner(considering bell shaped reward distributions for arms)
- It is shown that Thompson sampling generally gives better regret bound than UCB
- In Thompson sampling, eliminating arms is necessary for getting better complexity
- All are true

No, the answer is incorrect.
Score: 0
Accepted Answers:
In Thompson sampling, eliminating arms is necessary for getting better complexity