Assignment 3

Due on 2020-02-18, 23:59 IST.

Week 6 - Week 3

Due Assignment

Assignment Details

1. Consider the following policy: select an action for each state in your library.
   \[ a_t = \arg\max_{a} \sum_{t=1}^{T} r_t \]
   where \( T \) is the total number of states. Which of the following is true for the above algorithm?

   a) It is an \( \epsilon \)-greedy algorithm.
   b) It always explores the best state.
   c) It has a fixed probability of selecting the best state.
   d) It will explore even if the best state has a fixed probability of selecting the best state.

   
   2. Assume that a Contextual Bandit can be modeled as a full reinforcement learning problem. What is the action? (This can be an \( \epsilon \)-greedy method used to select actions corresponding to the arm in each bandit, with a single action leading to terminal states, going to a new state which corresponds to the same arm in each bandit, and selecting an action from the arm in the new state.)

   a) An \( \epsilon \)-greedy method.
   b) A Non-Binary Policy.
   c) A Bi-Binary Policy.
   d) No, there is no \( \epsilon \) in this problem.

   3. Which of the following statements are true? (True or False)

   a) We assume that the agent determines the reward based on the current state and action.
   b) Our agent is not guaranteed to get a positive reward.
   c) At any time step we can perform only one action.
   d) Zero rewards are possible.

   4. The baseline in the REINFORCE update can depend on the following factors, which of the following are correct?

   a) \( A_t \)
   b) \( G_t \)
   c) \( G_t \neq \sum_{t=0}^{T} r_t \)
   d) All of the above

   5. Which of the following is true for HDP?

   a) \( P(x|\pi) = P(\pi|\alpha)P(x|\pi, \alpha) \)
   b) \( P(x|\pi) = P(\pi|\alpha)P(x|\pi, \alpha) \)
   c) \( P(x|\pi) = P(\pi|\alpha)P(x|\pi, \alpha) \)
   d) None of the above

   6. Assume that a Contextual Bandit can be modeled as a full reinforcement learning problem. What is the action? (This can be an \( \epsilon \)-greedy method used to select actions corresponding to the arm in each bandit, with a single action leading to terminal states, going to a new state which corresponds to the same arm in each bandit, and selecting an action from the arm in the new state.)

   a) An \( \epsilon \)-greedy method.
   b) A Non-Binary Policy.
   c) A Bi-Binary Policy.
   d) No, there is no \( \epsilon \) in this problem.