Lesson 8 Quiz

4 questions

Correct

1 / 1 points

1. Which of the following algorithms is not designed for frequent pattern mining in stream data with approximation?

Space saving algorithm

Sticky sampling algorithm

FP-growth

Lossy counting algorithm

Correct

1 / 1 points

2. A data scientist is applying the lossy counting algorithm to a transactional data stream in order to obtain the counts of different items. If the bucket size is set to 1000, the total length of the transactional data stream is 10000, and the true count of an item A is 100, which of the following could be the possible outputs of item A’s count by lossy counting? Select all that apply.

98

90

80

102

1  
point

3. In CP-Miner, we use constraint-based sequential pattern mining to obtain the frequent sequences. Let us consider a source file, which has been transformed into a sequence DB after tokenization and hashing. If we set the max gap to 2 (the index difference between two items is no larger than 2) and the support threshold to 0.6, which of the following can be the frequent sequences output by CP-Miner? Select all that apply.

(1) <1, 2, 1, 3>

(2) <2, 3, 4, 1>

(3) <1, 2, 4, 3>

(4) <3, 2, 4, 3>

(5) <1, 2, 5, 4>

<1, 2>

<2, 4>

<3, 4>

<1, 3>

1  
point

4. Which of the following are designed for preserving data privacy? Select all that apply.

σ-frequent

t-closeness

K-anonymity

Differential privacy

Correct

1 / 1 points

5. Which of the following algorithms is not designed for frequent pattern mining in stream data with approximation?

Space saving algorithm

Sticky sampling algorithm

CloSpan

Lossy counting algorithm

Correct

1 / 1 points

6. A data scientist is applying the lossy counting algorithm to a transactional data stream in order to obtain the counts of different items. If the bucket size is set to 1000, the total length of the transactional data stream is 10000, and the true count of an item A is 100, which of the following could be the possible outputs of item A’s count by lossy counting?

110

95

105

85

1  
point

7. In CP-Miner, we use constraint-based sequential pattern mining to obtain the frequent sequences. Let us consider a source file, which has been transformed into a sequence DB after tokenization and hashing. If we set the max gap to 2 (the index difference between two items is no larger than 2) and the support threshold to 0.6, which of the following can be the frequent sequences output by CP-Miner? Select all that apply.

(1) <1, 2, 1, 3>

(2) <2, 3, 1, 4>

(3) <1, 2, 4, 3>

(4) <3, 2, 4, 3>

(5) <1, 2, 5, 4>

<1, 2>

<1, 3>

<2, 4>

<2, 3>

Submit Quiz