

X

NPTEL

reviewer2@nptel.iitm.ac.in ▼

Courses » Introduction to Data Analytics

Announcements

Course

Ask a Question

Progress



Unit 8 - Week 7 - Association Rule Mining and Big Data

Course outline

How to access the portal

Week 1 - Course Overview and Descriptive Statistics

Week 2 - Probability Distributions & Inferential Statistics

Week 3 - Inferential Statistics

Week 4 - Machine Learning

Week 5 - Supervised Learning (Regression and Classification Techniques) - I

Week 6 : Supervised Learning (Regression and Classification Techniques)-II

Week 7 - Association Rule Mining and Big Data

Associative Rule Mining

Association Rule Mining (cont'd)

Big Data, A small introduction

Assignment 7

The due date for submitting this assignment has passed. **Due on 2017-09-13, 23:55 IST**
As per our records you have not submitted this assignment.

The assignment also has some questions on the concept of **confusion matrix**. Reading the following page will help you to answer the questions of assignment:

https://en.wikipedia.org/wiki/Confusion_matrix

1) Imagine, you are working with NPTEL course management team and you want to develop a **1 point** machine learning algorithm which predicts the number of views on the courses. Your analysis is based on features like the name of instructor, number of courses taught by the same instructor on NPTEL in the past and a few other features. Which of the following evaluation metric would you choose in that case?

- mean square error
- classification accuracy
- F1 score
- precision
- recall

No, the answer is incorrect.

Score: 0

Accepted Answers:

mean square error

2) Imagine, you are solving a multiclass classification problem with highly imbalanced class. **1 point**
The distribution of the classes is such that, you observed the majority class 99% of the times in the training data. Your model has 99% accuracy after taking the predictions on test data. Which of the following is true in such a case?

- 1) Accuracy is not a good metric for imbalanced class problems.
- 2) Accuracy is a good metric for imbalanced class problems.
- 3) Precision and Recall are good metrics for imbalanced class problems.
- 4) Precision and Recall are not good metrics for imbalanced class problems.

- 1 and 2
- 2 and 3
- 1 and 3
- 2 and 4
- 3 and 4
- 1 and 4

No, the answer is incorrect.

Score: 0

Accepted Answers:

1 and 3

- Big Data - A small introduction (cont'd)
- Quiz : Assignment 7
- Feedback for week 7
- Assignment 7: Solution

Week 8 - Clustering Analysis and Prescriptive Analytics

Course Summary+ Insight into the Final Exam

3) Imagine you are working on a project which is a binary classification problem. You trained a model on training dataset and get the below confusion matrix on validation dataset. **1 point**

	n=165	
	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Based on the above confusion matrix, choose which option(s) is true among the following? The values are approximated up to 2 decimal points.

- 1) Accuracy is 0.91
- 2) Misclassification rate is 0.91
- 3) False positive rate is 0.95
- 4) True positive rate is 0.95

- 1 and 2
- 2 and 3
- 1 and 3
- 2 and 4
- 3 and 4
- 1 and 4

No, the answer is incorrect.

Score: 0

Accepted Answers:

1 and 4

4) In identifying frequent itemsets in a transactional database, we find the following to be the frequent 3-itemsets: {B, D, E}, {C, E, F}, {B, C, D}, {A, B, E}, {D, E, F}, {A, C, F}, {A, C, E}, {A, B, C}, {A, C, D}, {C, D, E}, {C, D, F}, {A, D, E}. Which among the following 4-itemsets can possibly be frequent? **1 point**

- {A, B, C, D}
- {A, B, D, E}
- {A, C, E, F}
- {C, D, E, F}

No, the answer is incorrect.

Score: 0

Accepted Answers:

{C, D, E, F}



5) Consider the following transactional database of 10 transactions.

1 point

Transaction ID	Item set
T1	AB
T2	BCD
T3	ACDE
T4	ADE
T5	ABC
T6	ABCD
T7	BA
T8	ABC
T9	ABD
T10	BCE



Making use of the apriori property, find the number of frequent item sets, for a minimum support of 4 (an item set with support greater than or equal to 4 is frequent)

- 7
 8
 10
 6

No, the answer is incorrect.

Score: 0

Accepted Answers:

8

6) Consider the following transactional data.

1 point

Transaction ID	Items
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

Assuming that the minimum support is 2, what is the number of frequent 2-itemsets (i.e., frequent items

sets of size 2)?

- 2
 4
 6
 8

No, the answer is incorrect.

Score: 0

Accepted Answers:

6

7) For the same data as above (Question 6), what are the number of candidate 3-itemsets and frequent 3-itemsets respectively? **1 point**

- 1,1
 2,2
 2,1
 3,2

No, the answer is incorrect.

Score: 0

Accepted Answers:

2,2

8) Continuing with the same data, how many association rules can be derived from the frequent itemset {A, B, E}? (Note: for a frequent itemset X, consider only rules of the form $S \rightarrow (X-S)$, where S is a non-empty subset of X.) **1 point**

- 3
 6
 7
 8

No, the answer is incorrect.

Score: 0

Accepted Answers:

6

9) For the same frequent itemset as mentioned above, which among the following rules have a minimum confidence of 60%? **1 point**

- $A \wedge B \Rightarrow E$
 $A \wedge E \Rightarrow B$
 $E \Rightarrow A \wedge B$
 $A \Rightarrow B \wedge E$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$A \wedge E \Rightarrow B$

$E \Rightarrow A \wedge B$

10) Which of the following statements are true, about frequent item-sets in the context of transactional databases? **1 point**

- Every maximal frequent itemset is a closed frequent itemset.
 Every closed frequent itemset is a maximal frequent itemset.
 We can recover all frequent itemsets given all maximal frequent itemsets
 We can recover the frequencies of all frequent itemsets, given the frequencies of all maximal frequent itemsets.

No, the answer is incorrect.

Score: 0

Accepted Answers:

Every maximal frequent itemset is a closed frequent itemset.

We can recover all frequent itemsets given all maximal frequent itemsets

◀ Previous Page

End ▶



© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

A project of



In association with



Funded by



Powered by

