

X

<https://swayam.gov.in>https://swayam.gov.in/nc_details/NPTEL

sai.doc45@gmail.com ✓

NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Business analytics and data mining
Modeling using R (course)



Course outline

How does an NPTEL online course work? ()

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

Week 4 ()

● Lecture 16
PERFORMANCE METRICS
(unit?
unit=42&lesson=43)

● Lecture 17
PERFORMANCE METRICS
PART-2 (unit?
unit=42&lesson=44)

● Lecture 18
PERFORMANCE METRICS

Week 4 : Assignment 4

The due date for submitting this assignment has passed.

Due on 2023-02-22, 23:59 IST.

Assignment submitted on 2023-02-22, 12:41 IST

1) Which of the following metrics measure the 'goodness of fit' of a regression model? **1 point**

- ☐ Mean absolute deviation
- ☒ Root mean squared error
- ☐ R-square
- ☐ Total sum of squared errors

No, the answer is incorrect.

Score: 0

Accepted Answers:

R-square

2) Which of the following statements is true about outliers in all scenarios? **1 point**

- ☐ Outliers should be identified and removed from a dataset.
- ☐ Outliers should be part of the training set but should not be present in the test set.
- ☐ Outliers should be part of the test set but should not be present in the training dataset.
- ☒ Nature of the problem determines how to deal with the outliers.

Yes, the answer is correct.

Score: 1

Accepted Answers:

Nature of the problem determines how to deal with the outliers.

3) Which of the following is the average positive difference between estimated and actual outcome values? **1 point**

PART-3 (unit?
unit=42&lesson=45)

Lecture 19
PERFORMANCE
METRICS
PART-4 (unit?
unit=42&lesson=46)

Lecture 20
PERFORMANCE
METRICS
PART- 5 (unit?
unit=42&lesson=47)

Quiz: Week 4
: Assignment
4
(assessment?
name=125)

Solution for
Week 4 :
Assignment 4
(unit?
unit=42&lesson=48)

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

Week 10 ()

Week 11 ()

Week 12 ()

Download
Videos ()

Weekly
Feedback ()

Text
Transcripts ()

- ☐ Root mean squared error
- ☐ Mean squared error
- ☒ Mean absolute error
- ☐ None of the above

Yes, the answer is correct.
Score: 1

Accepted Answers:
Mean absolute error

4) Which of the followings are common uses of unsupervised techniques like clustering? **1 point**

- ☒ Detect outliers
- ☒ Find a relevant set of input features for supervised learning
- ☒ Find meaningful relationships in the dataset
- ☐ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Detect outliers
Find a relevant set of input features for supervised learning
Find meaningful relationships in the dataset

5) Consider a classifier for screening of terrorists in a military application (True means action has to be taken). The confusion matrix is shown below. **1 point**

Which of the following situations would be acceptable if your classifier has it?

- ☐ FP >> FN
- ☐ FN >> FP
- ☒ TN >> FP
- ☐ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
FP >> FN

6) Which of the following are accurate axes labels for an ROC curve? **1 point**

- ☐ Vertical axis: % of true negatives; Horizontal axis: % of false negatives
- ☒ Vertical axis: % of true positives; Horizontal axis: % of false positives
- ☐ Vertical axis: % of false negatives; Horizontal axis: % of false positives

- ☐ Vertical axis: % of false positives; Horizontal axis: % of true negatives

Yes, the answer is correct.

Score: 1

Accepted Answers:

Vertical axis: % of true positives; Horizontal axis: % of false positives

7) Which of the following are correct scenarios to use oversampling? **1 point**

- ☐ When the cost of failing rare events are low.
- ☐ To de-emphasize rare events to the learning algorithm
- ☒ When a binary response variable includes significantly more items with one choice of response than the other choice, and we seek to accurately predict both the choices.
- ☐ When a binary response variable includes roughly the same number of items for each choice.

Yes, the answer is correct.

Score: 1

Accepted Answers:

When a binary response variable includes significantly more items with one choice of response than the other choice, and we seek to accurately predict both the choices.

8) Which is not included as a step when partitioning is done with oversampling. **1 point**

- ☐ Half the records from class 1 stratum are randomly selected into training partition.
- ☐ Randomly select class 0 records for training partition equal to no. of class 1 records.
- ☐ Randomly select class 0 records to maintain the original ratio of class 0 to class 1 records for validation partition.
- ☒ None of the above.

Yes, the answer is correct.

Score: 1

Accepted Answers:

None of the above.

9) Which of the following statements is true about prediction problems in general? **1 point**

- ☐ The output attribute must be categorical.
- ☐ The output attribute must be numeric.
- ☒ The resultant model is designed to determine future outcomes.
- ☐ The resultant model is designed to classify current behavior.

Yes, the answer is correct.

Score: 1

Accepted Answers:

The resultant model is designed to determine future outcomes.

10) Which of the following are true about asymmetric misclassification cost? **1 point**

- ☒ Considered when the misclassification error for a class of interest is more costly than for the other class.
- ☒ Goal of performance metrics changes to either minimization of costs or maximisation of profits.
- ☒ Performance metrics which incorporate asymmetric misclassification costs are used.

☐ None of the above.

Yes, the answer is correct.

Score: 1

Accepted Answers:

Considered when the misclassification error for a class of interest is more costly than for the other class.

Goal of performance metrics changes to either minimization of costs or maximisation of profits.

Performance metrics which incorporate asymmetric misclassification costs are used.