Assignment 9

Due on 2020-04-01, 23:59 IST.

This is the last assignment. The marks for this assignment have been passed, and the marks are cumulative. You must submit this assignment.

1. Do not use any programming language other than Python.

2. Select correct statement –

   a. Logistical regression model gives probability outputs
   b. Default cost for RQG score is 0.8
   c. The confusion matrix is used to compare actual outcomes with predicted data
   d. All of the above

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   a. Logistical regression model gives probability outputs

3. Which of the statements is true or false?

   a. A model is overfitting if it has a lower test error rate
   b. The test error rate is the percentage of the model’s predictions that are correct
   c. A model is underfitting if it has a higher test error rate
   d. All of the above

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   c. A model is underfitting if it has a higher test error rate

4. A model is overfitting if it has a lower test error rate.

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   c. A model is underfitting if it has a higher test error rate

5. Select the correct relationship between the threshold, sensitivity, and specificity –

   a. Low threshold value gives low specificity and high sensitivity
   b. Low threshold value gives high specificity and low sensitivity
   c. High threshold value gives specific high sensitivity and high specificity
   d. High threshold value gives high specificity and high sensitivity

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   c. High threshold value gives specific high sensitivity and high specificity

6. What is the meaning of “AUC” value?

   a. Social prediction
   b. Error prediction
   c. Perfect prediction
   d. None of these

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   c. Perfect prediction

7. Choose the best options for ROC graph –

   a. All points, which reside on the upper diagonal region are corresponding to good classifiers
   b. All points, which reside on the lower diagonal region are corresponding to noisy classifiers
   c. All points, which reside on the upper diagonal region have lower error than TPR
   d. All points, which reside on the upper diagonal region have more AUC than TPR

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   a. All points, which reside on the upper diagonal region are corresponding to good classifiers
   b. All points, which reside on the lower diagonal region are corresponding to noisy classifiers

8. Which method is used to convert categorical variables into numerical variables?

   a. Dummy variable
   b. Ideal variable
   c. Deselected value
   d. Classification report

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   a. Dummy variable

9. Which of the following is the relationship between dependent and independent variables by using an estimated regression equation?

   a. The relationship between dependent and independent variables is linear
   b. The relationship between dependent and independent variables is non-linear
   c. The relationship between dependent and independent variables is complex
   d. The relationship between dependent and independent variables is unknown

   Rs. the answer is incorrect.

   Marks: 1 point

   Accepted Answer:
   a. The relationship between dependent and independent variables is linear

10. Which of the following is the correct statement –

    a. More than two variables present, we can study the effect of one variable on the responses independently of the other variables
    b. More than two variables present, we have to study the joint effect of all the variables to draw a meaningful conclusion
    c. Both (a) and (b)
    d. None of these

    Rs. the answer is incorrect.

    Marks: 1 point

    Accepted Answer:
    c. Both (a) and (b)