Advanced Regression Exam [MCQ] [Timed] (Version: 0)

TEST

Correct Answer

(Answered in 200.9333333333 Minutes

Uploaded File: 1646058758-Advanced_Regression_Exam_MCQ-5665-Eric_Mbuthia.Zip

Question 1/20

You want to measure the difference between the true y-value of each data point and the predicted value, which of these is the best method to use

Minimum Residual Error



Error Difference Method

Maximum Error Method

Question 2/20

The equation $\frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2$ is used to calculate the:

Residual Sum of Squares



Logarithmic Residual Sum of Squares

R^2

Question 3/20 Both eta_0 and eta_1 in the multiple linear regression equation $Y=eta_0+eta_1X_1+eta_2X_2+...+eta_pX_p$ are known as coefficients.

Question 4/20

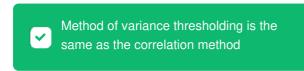
Which of the following is true regarding a multiple linear regression model trained on a dataset where all features have a positive correlation to Y (and there is no multicollinearity)?

- (i) All coefficients will be greater than zero
- (ii) Coefficients will be all equal
- (iii) The model error on the test set will be zero
- (iv) The resulting multiple linear regression model will have the general equation: $Y=\beta 0+\beta 1X1$

| i,ii and iv | |
|------------------|--|
| i and ii | |
| All of the above | |
| i only | |

Question 5/20

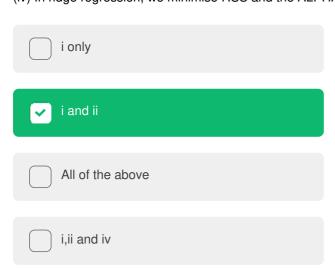
Which of the following is false regarding variables and variable selection?



Question 6/20

Which of the following statements are true?

- (i) L2_norm is known as the sum of the squares of model coefficients.
- (ii) L1_norm is known as the sum of the absolute values of model coefficients.
- (iii) We can use the shrinkage method in ridge regression to shrink coefficient's so that they are equal to zero
- (iv) In ridge regression, we minimise RSS and the ALPHA*(L1_norm)



Question 7/20

Variable selection and shrinkage methods (regularisation) are methods that can help with removing outliers in the data.



| True | |
|--|-------------------------------------|
| Question 8/20 | |
| Standardisation is a method of feature scaling which i | s more robust to handling outliers. |
| False | |
| True | |
| | |
| Question 9/20 | |
| LASSO regression can be considered as an impleme | ntation of which of the following: |

- (i) Shrinkage methods
- (ii) Feature selection
- (iii) Data Scaling
- (iv) Normalisation

| i, ii | | |
|----------|--|--|
| iv only | | |
| ii, iii | | |
| iii only | | |

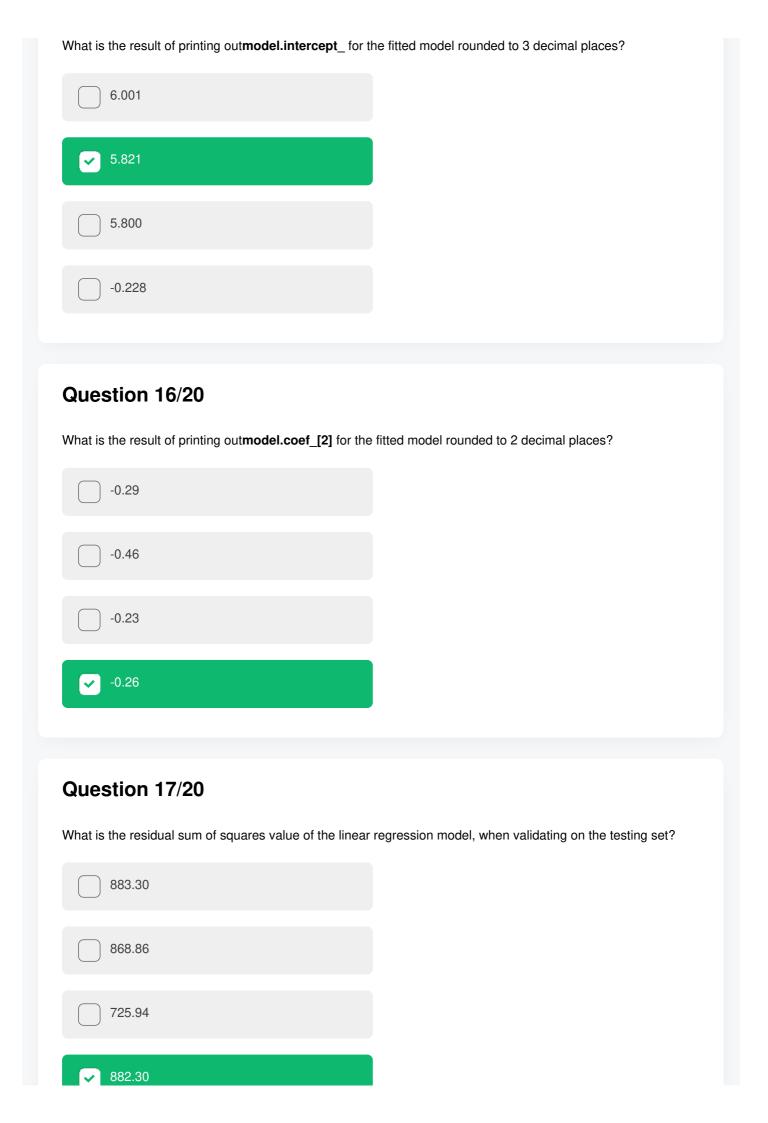
Question 10/20

Which of the following statements are true regarding Random Forest models:

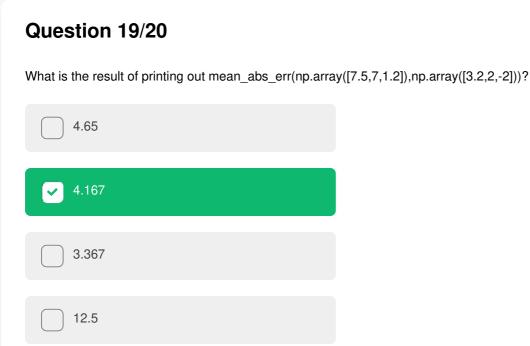
- (i) Random Forests are an example of a heterogeneous ensemble model
- (ii) Random Forests are trained using the boosting method
- (iii) Individual estimators are trained with different subsets of the data

| (iv) Random Forests are generally more prone to overfitting compared | to decision trees |
|--|-------------------|
| i, iii and iv | |
| iii only | |
| ii and iv | |
| i only | |
| | |
| Question 11/20 | |
| Practical Questions | |
| Questions 11 - 20, are practical questions based on the given jupyter of expected to fill in the missing code and use the resulting functions to a | |
| What is the result of printing out the 6th column and the 13th row of X_ | train? |
| -0.09303318078696134 | |
| -0.8282787380653501 | |
| -1.2508601954469347 | |
| 0.056380810844757615 | |
| | |
| Question 12/20 | |
| What is the result of printing out the 6th column and the 13th row of X_ | test? |
| 1.7170736234873545 | |
| -0.17191842758365714 | |

| 0.5415166925051395 |
|--|
| -1.5508940181797966 |
| Question 13/20 |
| |
| What is the result of printing out the 16th row of y_train |
| 4 |
| 6 |
| 5 |
| 7 |
| Question 14/20 |
| What is the result of printing out the 16th row of y_test? |
| 7 |
| |
| |
| 6 |
| |



Question 18/20 What is the residual sum of squares value of the decision tree regression model, when validating on the testing set? 1094.0 4622.0 1113.0



Question 20/20

Which regression model (linear vs decision tree) has the lowest mean absolute error?



