



Modules



Grades



Inbox



Calc

☐ Pre-Processing

☐ Model Building

☒ Model Building Assessment
Assignment

Model Building Assessment

The due date for submitting this assignment has passed.

Due on 2024-03-10, 23:59 IST.

You may submit any number of times before the due date. The final submission will be considered for grading.

You have last submitted on: 2024-03-09, 18:27 IST

Note: For numerical type questions, always enter the answer correct upto 3 decimal places without rounding off, unless otherwise stated.

1) Load the dataset.

1 point

The file has no index column.
The last column is the target column.
The first row of the file has column ids

Click here to view the dataset

Which dataset are you using for this exam?

- ☐ V1
- ☒ V2
- ☐ V3
- ☐ V4
- ☐ V5

Yes, the answer is correct.

Score: 1

Accepted Answers:

V2

2) Split the dataset into train dataset and test dataset in the following manner

Use train_test_split to split the dataset into train and test dataset with test size equal to 0.3(i.e.30%) and random_state equal to 42. Let other parameters have default values.

Columns except the last column should be the feature matrix (X_train or X_test)

Last column will be the label vector.

(Common instructions for Q.2, Q.3 and Q.4)

Train the ridge model on the training data with the following parameters:

alpha = 10
solver = 'saga'
tol = 1e-4
random_state = 42

Enter the value of R^2 score on the test dataset.

0.6846

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 0.661,0.671

7 points

3) What is the index of most important feature? Note the index starts from 0. Ignore the intercept for this question.

10

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Numeric) 9

2 points

4) What is the index of least important feature? Note the index starts from 0. Ignore the intercept for this question.

9

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Numeric) 2

2 points

6 points

5) (Common Instructions for Q.5 and Q.6)
 Take SGDRegressor(random_state = 42) estimator with GridSearchCV. Hyperparameter tuning to be done over the following parameters:
 penalty as ['l1', 'l2']
 alpha values as [1e-5, 1e-4, 1e-3, 1e-2, 1e-1]
 values of tol as [1e-4, 1e-3, 1e-2, 1e-1]
 Use cross-validation = 5
 Set scoring as neg_mean_absolute_error
 Use the best model from above hyper parameter tuning process to answer following questions:

What is the best penalty?

- ☐ l1
- ☒ l2

Yes, the answer is correct.

Score: 6

Accepted Answers:

l2

6) What will be value of mean absolute error on the test dataset?

3.5919

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 3.685,3.695

7 points

7) (Common Instructions for Q.7 and Q.8)
 Create a pipeline of the PCA() as transformer and Lasso as an estimator.
 Use GridSearchCV for tuning the hyperparameters of the created pipeline on training dataset.
 Values of n_components for PCA to be [0.9, 0.95]
 lasso alpha value to be taken as : [10, 1, 0.01, 0.001]
 scoring : neg_mean_absolute_error.
 Use 5 fold cross validation.
 n_jobs = -1 (negative one) [it helps in using all the computational power to run this job]
 (Note: Kindly ignore the warning.)

If we fit the pipeline on the training dataset, what will be the R2 score on the test dataset?

0.6230

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 0.595,0.605

6 points

8) How much variance is explained by the first principle component?

1.1855

Yes, the answer is correct.

Score: 7

Accepted Answers:

(Type: Range) 1.185,1.195

7 points

9) Create a pipeline of the PolynomialFeatures as transformer and Lasso as an estimator with the following parameters:
 - For PolynomialFeatures:
 - interaction_only = False
 - degree = 2
 - For Lasso:
 - alpha = 1
 - warm_start = True
 - random state as 0

Fit the pipeline on the training dataset and find the R^2 score on the test dataset.

0.1751

Yes, the answer is correct.

Score: 7

Accepted Answers:

(Type: Range) 0.168,0.178

7 points