← Go Back to Feature Selection, Model Selection and Tuning

Course Content

Weekly Quiz - Feature Engineering and Cross Validation

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	30m
Due Date	:	Nov 22, 2021, 8:30 AM
Your Score	:	15/15

Instructions

Attempt History

Attempt #1

Nov 22, 2021, 1:12 AM

Marks: 15

Q No: 1 Correct Answer

Marks: 2/2

If k=3 in the K-fold cross-validation, how many times each fold will be used in training?

1

2 You Selected

3

5

In K-fold cross-validation, the dataset is divided into k folds and in every execution, k-1 folds are used in training and 1 fold is used in testing. Hence, if k = 3 then every fold will be used 2 times in training and 1 time in testing.

Q No: 2 Correct Answer

Marks: 2/2

Which of the following techniques can be used to deal with a dataset having imbalanced classes?

1. Synthetic Minority Over-sampling Technique
2. Over-sampling, Random Undersampling
3. Linear regression
4. Cross-Validation

2 and 3

1 and 2 You Selected

1, 2 and 3

3 and 4

We resample the data when we have imbalanced data to balance out class distribution. Resampling can be done in two ways - Either we oversample the data or undersample the data. SMOTE is one of the over-sampling techniques. Hence, 1 and 2 are the correct options.

Q No: 3 Correct Answer

Marks: 2/2

In a city with a population of 1 million, 500 people have been diagnosed with cancer, whereas the rest of the people do not have cancer.

Such a class distribution is considered to be:

- ○ Balanced
- ◉ Imbalanced — You Selected

The ratio of the classes is 500 (Have cancer):10,00,000 (Do not have cancer). This clearly indicates that the data is imbalanced.

Q No: 4 Correct Answer

Marks: 1/1

SMOTE (Synthetic Minority Over-sampling Technique) uses which of the following algorithms to create synthetic data?

- ○ Decision trees
- ◉ KNN Algorithm (K- Nearest Neighbor) — You Selected
- ○ Linear Regression
- ○ Logistic Regression

SMOTE uses KNN to create synthetic data.

Q No: 5 Correct Answer

Marks: 1/1

Which of the following split of the data is used to evaluate the final machine learning model?

- Train data
- Validation data
- Test data — You Selected

We use a training set to train our model, a validation set to check the performance of the model so that we can tweak hyperparameters and perform tuning, and a test set is used at the last stage to evaluate the performance of our **final** model.

Q No: 6 Correct Answer

Marks: 1/1

What is the minimum value of 'K' we can use to perform K-fold Cross Validation,

- 1
- 2 — You Selected
- 3
- 4

Q No: 7 Correct Answer

Marks: 1/1

If there are 100,000,000 observations in the dataset then which of the following cross-validation techniques would be appropriate to use?

K-fold You Selected

LOOCV

Using LOOCV on a large dataset would be time-consuming. Hence, we would use the K fold.

Q No: 8 Correct Answer

Marks: 1/1

Which of the following methods is used to improve the performance when a linear regression model is overfitting?

1. Regularization methods
2. Building Pipeline

Only 1 You Selected

Only 2

Both 1 and 2

None of the above

Regularization (Shrinkage) is a technique that is used to control overfitting whereas building pipelines and, feature engineering are the techniques used in model building and improving model performance.

Q No: 9 Correct Answer

Marks: 2/2

Which of the following statements are true about contour plots?

1. Every ring represents the combination of coefficients and slope
2. Every point along a contour ring represents the same error value

3. The innermost ring gives the least error
4. The error value keeps on increasing as we move inwards

- (●) 1, 2 and 3 — You Selected
- ( ) 1, 2 and 4
- ( ) 2, 3 and 4
- ( ) 3 and 4

Every ring in the contour plot represents the combination of slope and coefficient and every point along a contour ring has the same error value. As we move inwards, the error decreases, and hence, the innermost ring gives the least error.

Q No: 10 Correct Answer

Marks: 2/2

Which of the following regularization techniques might make the coefficients zero that reduces the dimension of the data?

- ( ) Ridge
- (●) Lasso — You Selected

The penalty term in Lasso regression is raised to power 1 and this process makes the coefficients zero which results in the reduction of the dimension of data.

Comments:

+ Add comments

Previous

Next