← Go Back to Model Tuning

Course Content

Weekly Quiz - Feature Engineering and Cross Validation

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	30m
Due Date	:	Jan 29, 1:30 AM CET
Your Marks	:	15/15

Instructions

Attempt History

Attempt #1
Jan 25, 1:17 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 testing?

- (●) 1 — You Selected
- ( ) 2
- ( ) 3
- ( ) 5

In K-fold cross-validation, the dataset is divided into K folds and in each execution, K-1 folds are used in training and 1 fold is used in testing. As such, each fold is used K-1 times for training and 1 time for testing. Hence, if K = 3, then each 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 handle a class imbalance in a dataset?

Synthetic Minority Over-sampling Technique (SMOTE)
Random Undersampling
Cross-Validation

- ○ Only 1
- ○ Only 2
- ◉ 1 and 2 — You Selected
- ○ 1, 2, and 3

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 and Random Undersampling is one of the under-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' that can be used to perform K-fold Cross-Validation?

- ○ 1
- ◉ 2 — You Selected
- ○ 3
- ○ 4

We divide the dataset into k folds such that k-1 folds are used for training and the remaining 1 is used for testing. Hence, the minimum value of k could be 2 where 1 fold is used in training and another one in testing.

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 cross-validation technique.

Q No: 8 Correct Answer

Marks: 1/1

We can use regularization methods to deal with the curse of dimensionality.

True You Selected

False

Regularization is a technique that can be used to deal with the curse of dimensionality and helps us in improving model performance by shrinking the model coefficients.

Q No: 9 Correct Answer

Marks: 2/2

Which of the following statements are true about contour plots?

Every ring represents the combination of coefficients and slope
The innermost ring gives the least error
The error value keeps on increasing as we move inwards

Only 1

1 and 3

1 and 2 You Selected

1, 2, and 3

Every ring in the contour plot represents the combination of slope and coefficient. 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.

Previous Next

Comments: