← Go Back to Advanced Machine Learning

Course Content

Weekly Quiz - Model Tuning

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	30m
Due Date	:	Aug 12, 3:30 AM EDT
Your Marks	:	10/10

Instructions

Attempt History

Attempt #1
Aug 11, 4:20 PM

Marks: 10

Q No: 1 Correct Answer

Marks: 1/1

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: 1/1

Which of the following techniques can be used to handle a class imbalance in a dataset?

1. Synthetic Minority Over-sampling Technique (SMOTE)
2. Random Undersampling
3. 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: 1/1

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
- Random Forest

SMOTE uses KNN to create synthetic data.

Q No: 5 Correct Answer

Marks: 1/1

While tuning hyperparameters, the data should be split into three parts - train, validation, and test - to avoid data leakage.

- (●) True — You Selected
- ( ) False

Data leakage happens when a certain part of the data is already seen in the training process. That's why it is always advised to keep the test dataset away and use it only for final evaluation. When we impute the missing values for the entire data and then split the data into train-test then a certain part of the data is leaked in the training process. Regularization is used to deal with overfitting. Hence, the best measure to avoid data leakage is to split the data into three sets.

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

Which of the following statements are correct for GridsearchCV?

1. Grid search looks only at a randomly selected subset of hyperparameters from the parameter grid.

2.GridSearchCV evaluates the model for each combination of hyperparameters in the parameter grid using the Cross-Validation method.

- ( ) Only 1
- ( ) 1 and 2
- (●) Only 2 — You Selected
- ( ) None

Grid search divides the domain of the hyperparameters into a discrete grid. Then every combination of values of this grid tried by it and performance metrics using cross-validation are calculated.

Q No: 8 Correct Answer

Marks: 1/1

Which of the following statements are correct about hyperparameter tuning?

1. Hyperparameters tuning is done on the test set.
2. Grid search and randomized search methods can be used to perform hyperparameter tuning.
3. Tuning does not have a significant effect on the model's performance
4. Choosing optimal hyperparameters can lead to improvements in the overall model's performance

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

Hyperparameter tuning is done on the training set with the help of a randomized search or grid search method. It has a significant impact on the model's performance.

Q No: 9 Correct Answer

Marks: 1/1

Consider a RandomizedSearchCV with the following parameters

```
param_grid={'n_estimators':[50,100,150,200],
            'learning_rate':[0.01,0.1,0.2,0.3]}

gb = GradientBoostingClassifier(random_state=1)

#Calling RandomizedSearchCV
clf = RandomizedSearchCV(estimator=gb, param_distributions=param_grid, cv=5, random_state=1, n_iter = 10)

clf.fit(X_train,y_train)
```

How many times will the randomized search fit the model?

- ◉ 50 — You Selected
- ○ 9
- ○ 135
- ○ 27

Here, n_iter = 10, so the Random search will try only 10 unique models.

CV = 5, so every model will be fit 5 times

Therefore, the model will be fit 10*5 = 50 times

Q No: 10 Correct Answer

Marks: 1/1

Which of the following statements are correct?

1. We should tweak the hyperparameters based on the model's performance on the validation set
2. Hyperparameter tuning can help us deal with both overfitting and underfitting

- ○ Only 1
- ○ Only 2
- ◉ 1 and 2 — You Selected
- ○ None

The hyperparameters are tweaked based on the performance of the validation set. Tuning helps in model performance which helps in dealing with the underfitting and overfitting of the model. As we can tweak hyperparameters during the training process, there is no fixed set of values to be used for tuning.

Comments:

+ Add comments

< Previous Next >