← Go Back to Feature Selection, Model Selection and Tuning

Course Content

Weekly Quiz - ML Pipeline and Hyperparameter Tuning

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	30m
Due Date	:	Dec 06, 2021, 11:30 AM
Your Score	:	15/15

Instructions

Attempt History

Attempt #1
Dec 06, 2021, 9:21 AM

Marks: 15

Q No: 1 Correct Answer

Marks: 2/2

State whether the following statement is true or false.

'While hyperparameter tuning, it is mandatory to tune all the hyperparameters associated with an algorithm.'

True

False — You Selected

It is not mandatory to tune all the hyperparameters associated with an algorithm. Tuning certain hyperparameters help us in achieving desired results.

Q No: 2 — Correct Answer

Marks: 1/1

Which of the following is the correct way to define a pipeline object using sklearn Pipeline() function if the steps are defined as -

steps = [('scaler', MinMaxScaler()), ('model', LogisticRegression())]

pipeline = Pipeline.steps

pipeline = steps(Pipeline)

pipeline = Pipeline(steps) — You Selected

pipeline = steps.Pipeline

Pipeline() is a function and we need to pass arguments to the function. The pipeline takes a list of tuples that is passed as an argument. Hence, Pipeline(steps) is the correct answer.

Q No: 3 — Correct Answer

Marks: 1/1

Which of the following measures can be taken to avoid data leakage?

Options-

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

2. Using regularization on the test set
3. Imputing missing values for the entire data before splitting the data into train and test

- Both 2 and 3
- Only 1 — You Selected
- Only 3
- Only 2

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: 4 — Correct Answer

Marks: 2/2

Which of the following statements are true about Randomized search CV?

1. It doesn't always guarantee to give the best parameters combination
2. It evaluates all the possible combinations available in the grid
3. It usually executes faster than Grid search CV
4. Only a fixed number of hyperparameter values are tried out from the provided parameter grid

1, 2 and 3

1, 2 and 4

2, 3 and 4

1, 3 and 4 You Selected

Random search CV tries random combinations and not all the possible combinations are tried out. Due to this, it doesn't always guarantee to give the best results. Random search cv tries some random combinations based on n_iter value and hence, the execution is comparatively faster than grid search.

Hence, Option 1, 3, and 4 are Correct.

Q No: 5 Correct Answer

Marks: 1/1

Which of the following is NOT a hyperparameter for their corresponding model?

Shrinkage factor in Ridge Regression

Number of estimators in Random Forest

Weight coefficients in a Linear Regression model You Selected

Depth of a tree in a Decision Tree

Hyperparameter is a parameter value that can be controlled in the learning process. The number of estimators, depth of the tree, and shrinkage factors are the parameters that can be controlled while tuning the model but weights are the values that need to be optimized(learned while training process). We tune the parameters to get the optimal weights.

Q No: 6 Correct Answer

Marks: 1/1

State whether the following statement is true or false.

'We should use a randomized search to reduce time complexity while searching for the combination of hyperparameters in a large space of hyperparameters.'

- (●) True — You Selected
- ( ) False

Searching in the grid for a large number of hyperparameters could be time-consuming. We should use a randomized search in this case.

Q No: 7 Correct Answer

Marks: 2/2

Which of the following functions creates a pipeline and automatically names each step so that we don't need to specify the names?

- ( ) pipeline.fit()
- (●) make_pipeline() — You Selected
- ( ) pipeline.transform()
- ( ) Pipeline()

make_pipeline() is a function that creates a pipeline and automatically names each step whereas in Pipeline() we need to specify the names of each step in a tuple.

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

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

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: 2/2

Consider a RandomizedSearchCV with the following parameters

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

gb = GradientBoostingClassifier(random_state=1)

#Calling RandomizedSearchCV
clf = RandomizedSearchCV(estimator=gb, param_distributions=param_grid,

clf.fit(X_train,y_train)
```

How many times will the randomized search fit the model?

- 27
- 135
- 9
- 50 — You Selected

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: 2/2

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
3. There is a fixed defined set of hyperparameters that should only be used for tuning every model

1 and 2 You Selected

Only 1

1, 2 and 3

2 and 3

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