Kahoot!

Machine Learning: Summary

2 plays · 8 players

A public kahoot

Questions (26)

1 - Quiz

60 sec

In 5-fold cross-validation, how many times is each observation part of the testing set?

- 1 time ✓
- 4 times ✗
- 5 times ✗
- It's not possible to say ✗

2 - Quiz

20 sec

Which of the following model evaluation procedures takes the most time to run?

- Train/test split ✗
- 5-fold cross-validation ✗
- 10-fold cross-validation ✓

3 - Quiz

20 sec

Which of these is used to "dummy encode" a categorical feature?

- OneHotEncoder ✓
- ColumnTransformer ✗
- Pipeline ✗

4 - Quiz

20 sec

Does the output of a OneHotEncoder have the same shape as the input?

- Yes, the output has the same shape as the input ✗
- No, the output has additional rows ✗
- No, the output has additional columns ✓

5 - Quiz

30 sec

Which of the following procedures outputs the most reliable estimate of a model's out-of-sample performance?

- Train/test split ✗
- K-fold cross-validation ✗
- Repeated K-fold cross-validation ✓

6 - Quiz

30 sec

Do you have to fit a model with data before passing it to the cross_val_score function?

- Yes ✗
- No ✓

7 - Quiz

30 sec

If you think of GridSearchCV as a "for" loop, what is happening within that loop?

- Train/test split ✗
- K-fold cross-validation ✓

8 - Quiz

30 sec

Which data structure is used to store the entire parameter grid for GridSearchCV?

- Dictionary ✓
- List ✗

9 - Quiz

60 sec

Do GridSearchCV and cross_val_score both use a fit method?

- Yes, both of them use a fit method ✗
- GridSearchCV uses a fit method, but cross_val_score does not ✓
- cross_val_score uses a fit method, but GridSearchCV does not ✗
- No, both do not use a fit method ✗

10 - Quiz

120 sec

How many times will cross-validation run on GridSearchCV for a grid of two parameters with values 2 and 4 respectively?

- 10 ✗
- 14 ✗
- 40 ✓
- We can't tell if we don't know how many folds there will be. ✗

11 - Quiz

20 sec

Is a logistic regression model used for regression problems or for classification problems?

- Regression problems ✗
- Classification problems ✓

12 - Quiz

20 sec

What's the main strength of classification accuracy as an evaluation metric?

- It's easy to understand ✓
- It tells you the types of errors your model is making ✗
- It compares the model to a "dumb" one that predicts the most frequent class ✗

13 - Quiz

20 sec

What's the conventional name for the class that has been encoded as a 1?

- Positive class ✓
- Negative class ✗

14 - Quiz

60 sec

Which of the following statements describes a False Positive?

- The true value is Negative and the predicted value is Positive ✓
- The true value is Positive and the predicted value is Negative ✗

15 - Quiz

30 sec

Can you use a confusion matrix when there are more than two response classes?

- Yes ✓
- No ✗

16 - Quiz

60 sec

Can you use the area under the curve (AUC) metric when there are more than two response classes?

- Yes ✗
- No ✓

17 - Quiz

60 sec

Can you use a grid search to choose a model with the "best" confusion matrix?

- Yes ✗
- No ✓

18 - Quiz

60 sec

What metric can you use to determine the "best" confusion matrix?

- MCC ✓
- f1 ✗
- CM ✗
- ROC ✗

19 - Quiz

30 sec

Classification threshold is 0.8 instead of 0.5. Will the positive class be predicted more frequently or less frequently?

- More frequently ✗
- Less frequently ✓

20 - Quiz

60 sec

What value for the classification threshold increases both sensitivity and precision?

- Below 0.5 ✗
- Above 0.5 ✗
- There is no such value ✓
- Exactly 0.5 ✗

21 - Quiz

20 sec

What is the best possible value for Area Under the Curve (AUC)?

- 0 ✗
- 0.5 ✗
- 1 ✓

22 - Quiz

30 sec

In cases of class imbalance, is classification accuracy or AUC a more useful evaluation metric for two classes?

- Accuracy is more useful ✗
- AUC is more useful ✓
- They are equally useful ✗
- The AUC is not a classification metric ✗

23 - Quiz

60 sec

Which of the following hyperparameter(s) when increased might cause Random forest to overfit the data?

- Number of trees ✗
- Depth of tree ✓
- Learning rate ✗
- All of them ✗

24 - Quiz

30 sec

Your ML team decides to use KMeans to separate the following points into 2 groups. Do you agree?

- Yes ✗
- No ✓

25 - Quiz

30 sec

When a model underfits what is its bias?

- High ✓
- Low ✗
- In the middle ✗

26 - Quiz

120 sec

A high recall and low precision model corresponds to the case of having what variance and bias?

- high variance and low bias ✓
- high variance and high bias ✗
- low variance and low bias ✗
- low variance and high bias ✗