Congratulations! You passed!
TO PASS 75% or higher

Keep Learning

GRADE
75%

Validation

TOTAL POINTS 4

1. Suppose we are given a huge dataset. We did a KFold validation once and noticed that scores on each fold are roughly the same. Which validation type is most practical to use? **1 / 1 point**

   - ( ) We should keep on using KFold scheme as the data is homogeneous and KFold is the most computationally efficient scheme.
   - (•) We can use a simple holdout validation scheme because the data is homogeneous.
   - ( ) Leave-one-out because the data is not homogeneous.

   Correct
   Correct! If scores on different folds are similar, we indeed can use holdout split. In fact, this is often the case.

2. Suppose we are given a medium-sized dataset and we did a KFold validation once. We noticed that scores on each fold differ noticeably. Which validation type is the most practical to use? **1 / 1 point**

   - (•) KFold
   - ( ) LOO
   - ( ) Holdout

   Correct
   Correct. This is the most frequent way to deal with this kind of situations. Also, scores deviation in KFold will help you to select statistically significant change in scores while tuning a model.

3. The features we generate depend on the train-test data splitting method. Is this true? **1 / 1 point**

   - ( ) False
   - (•) True

   Correct
   Correct. For an explanation check out the third video in the module about choosing a train/test split.

4. What of these can indicate an expected leaderboard shuffle in a competition? **0 / 1 point**

   - [x] Little amount of training or/and testing data

   Correct
   In this case randomness can shuffle scores on the private leaderboard

   - [ ] Most of the competitors have very similar scores
   - [x] Different public/private data or target distributions

   Correct
   In this case competitors can receive quite unexpected scores on private LB.

   You didn't select all the correct answers