Validation

Practice Quiz, 4 questions

3/4 points (75.00%)

Congratulations! You passed!

Next Item

1 / 1 point

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?

- 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.

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

- Leave-one-out because the data is not homogeneous.

---

1 / 1 point

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?

- Holdout
- LOO
- KFold

Validation

Practice Quiz, 4 questions

3/4 points (75.00%)

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.

---

0 / 1 point

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

- True
- False

This should not be selected
Incorrect. For an explanation check out the third video in the module about choosing a train/test split.

---

1 / 1 point

4.
What of these can indicate an expected leaderboard shuffle in a competition?

- [x] Different public/private data or target distributions

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

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

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

- [x] Most of the competitors have very similar scores

Correct

In this case randomness can shuffle scores on the private leaderboard

Validation

Practice Quiz, 4 questions

3/4 points (75.00%)