

Validation
TOTAL POINTS 4

1.Question 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.

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

1 point

2.Question 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

☒ KFold

☐ LOO

1 point

3.Question 3

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

☐ False

☒ True

1 point

4.Question 4

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

(X) Little amount of training or/and testing data

(X) Different public/private data or target distributions

(X) Most of the competitors have very similar scores

1 point