Validation strategies

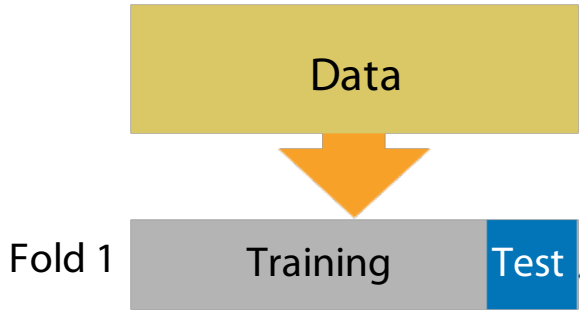
Validation types

- Holdout
- K-fold
- Leave-one-out

Validation types

- Holdout: ngroups = 1

  `sklearn.model_selection.ShuffleSplit`



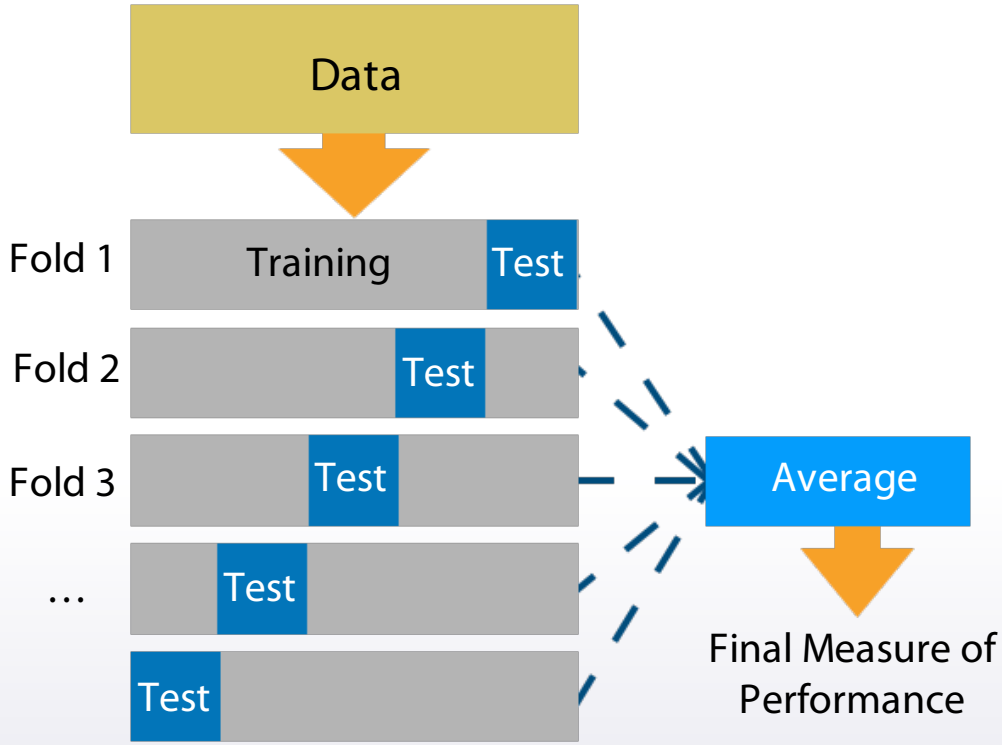
Validation types

- Holdout: ngroups = 1

  `sklearn.model_selection.ShuffleSplit`

- K-fold: ngroups = k

  `sklearn.model_selection.Kfold`



Validation types

- Holdout: ngroups = 1

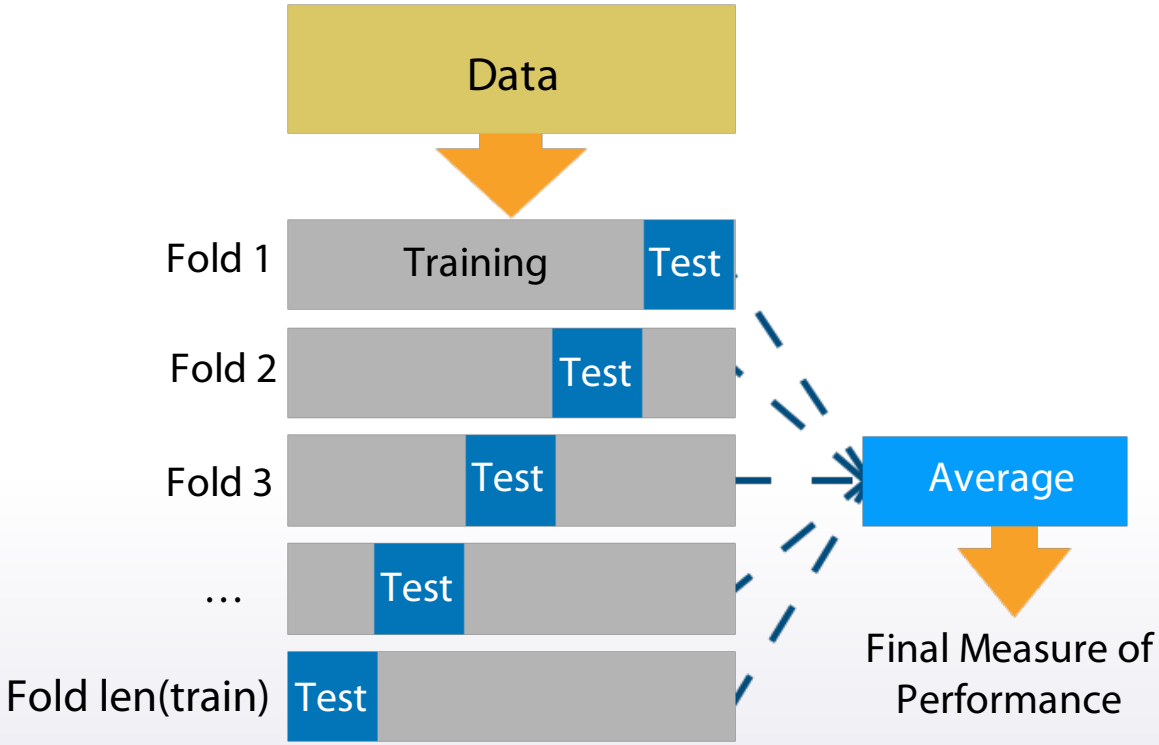
  `sklearn.model_selection.ShuffleSplit`

- K-fold: ngroups = k

  `sklearn.model_selection.Kfold`

- Leave-one-out: ngroups = len(train)

  `sklearn.model_selection.LeaveOneOut`



Stratification

Samples and their target values

0	1	0	0	1	1	1	0

Stratification

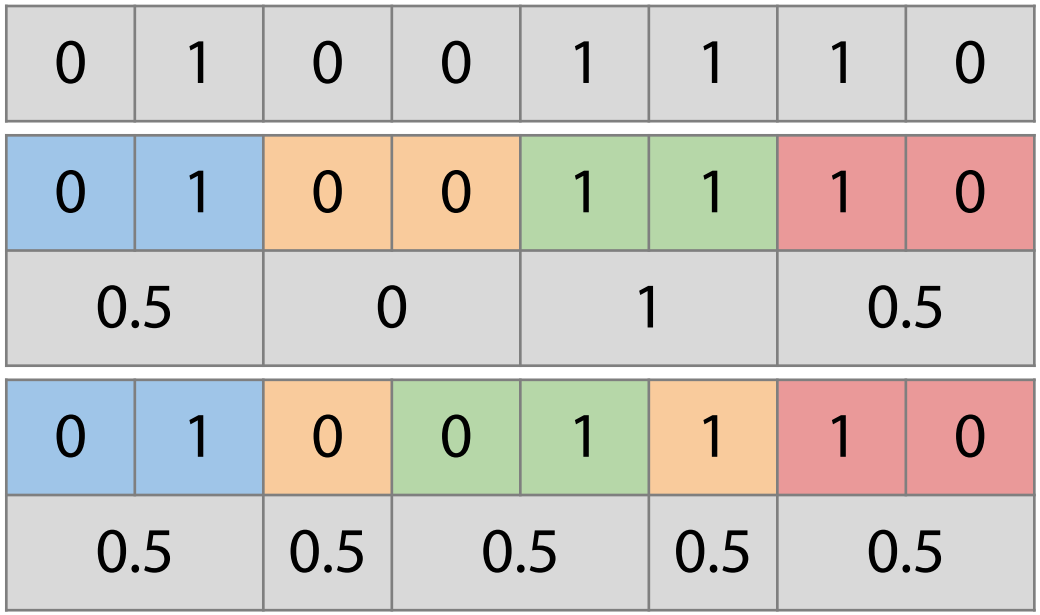
Samples and their target values

0	1	0	0	1	1	1	0

0	1	0	0	1	1	1	0
0.5		0		1		0.5	

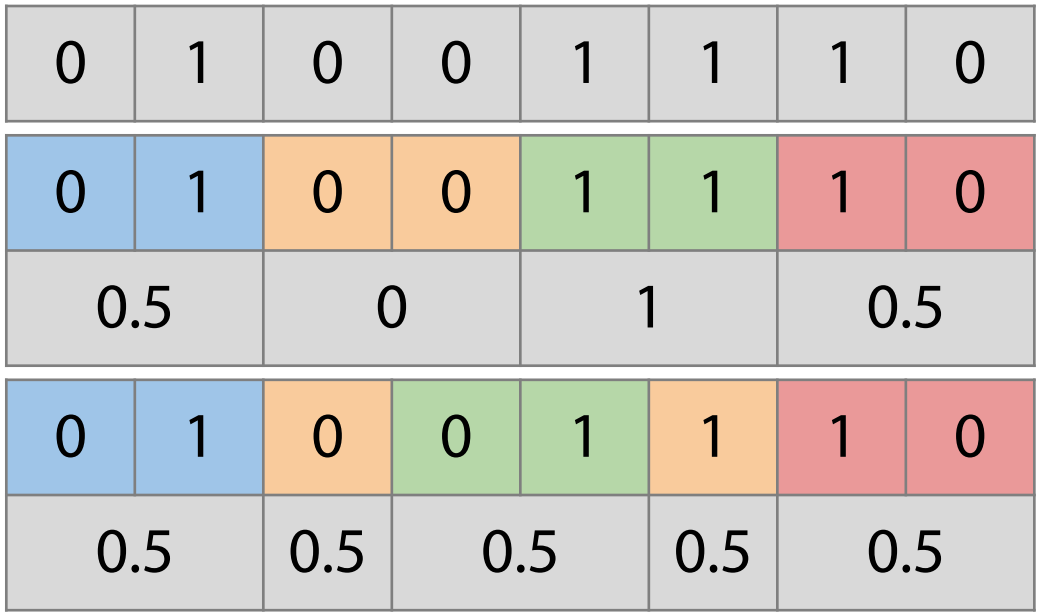
Stratification

Samples and their target values



Stratification

Samples and their target values



Stratification is useful for:

- Small datasets
- Unbalanced datasets
- Multiclass classification

Conclusion

There are three main validation strategies:

1. Holdout
2. KFold
3. LOO

Stratification preserve the same target distribution over different folds