

Validation strategies

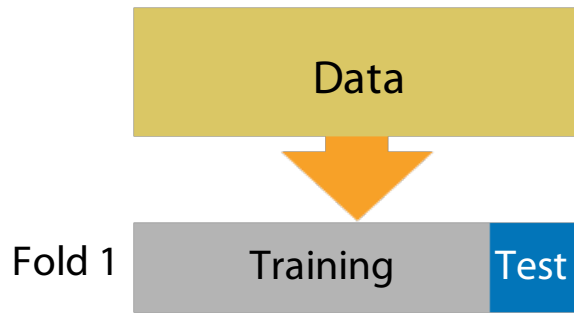
Validation types

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

Validation types

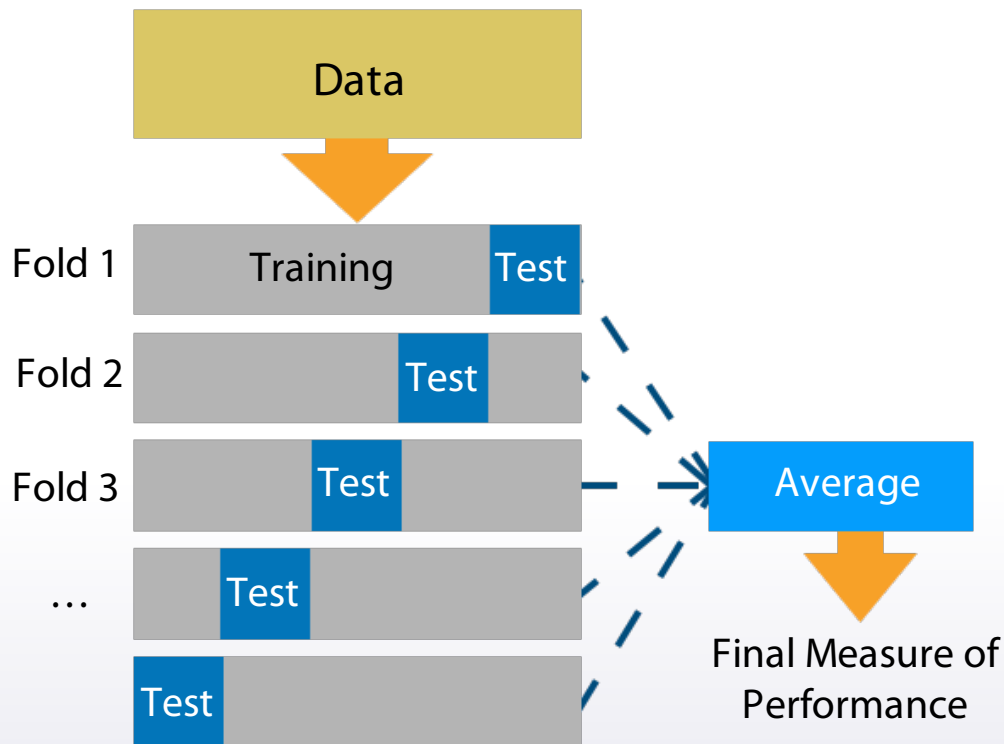
- Holdout: $n\text{groups} = 1$

`| sklearn.model_selection.ShuffleSplit`



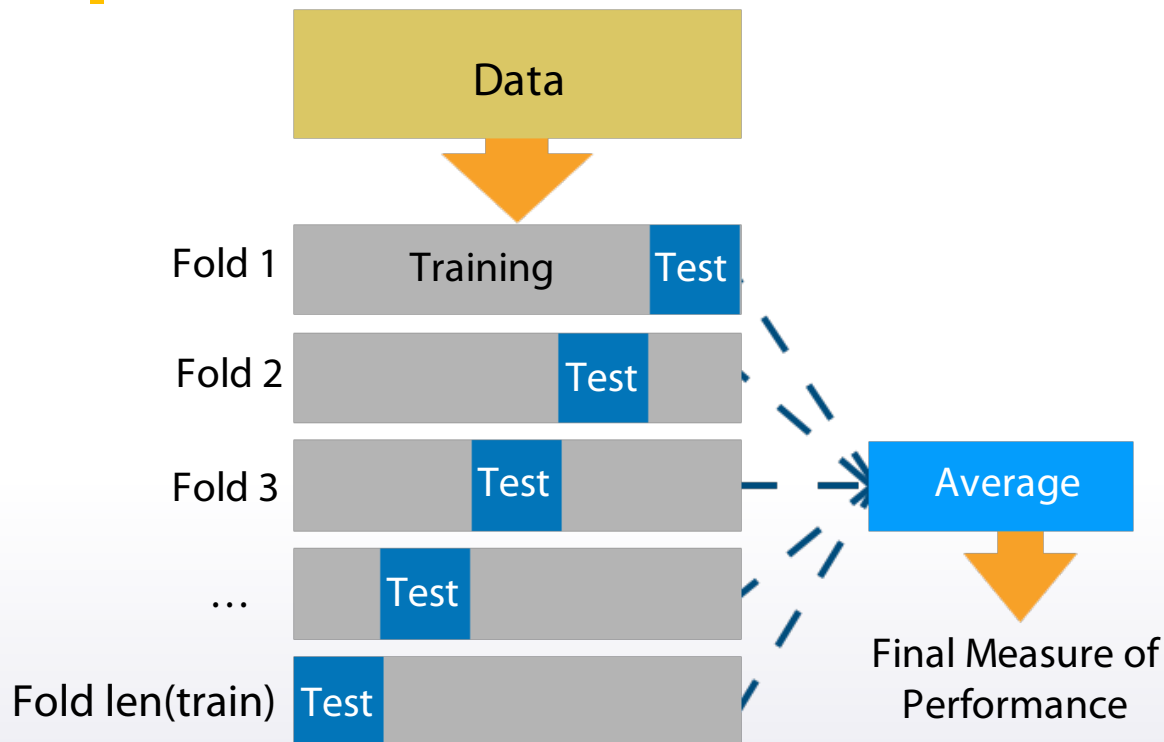
Validation types

- Holdout: $\text{ngroups} = 1$
| `sklearn.model_selection.ShuffleSplit`
- K-fold: $\text{ngroups} = k$
| `sklearn.model_selection.Kfold`



Validation types

- Holdout: $\text{ngroups} = 1$
| `sklearn.model_selection.ShuffleSplit`
- K-fold: $\text{ngroups} = k$
| `sklearn.model_selection.Kfold`
- Leave-one-out: $\text{ngroups} = \text{len}(\text{train})$
| `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

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

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

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