



Nested

Cross-Validation

Uses of Cross-Validation

• Estimate the generalization error of a given model

- Select best performing model from a group of models
 - Different algorithms
 - Different feature subsets

Select hyperparameters



Uses of Cross-Validation

• Estimate the generalization error of a given model

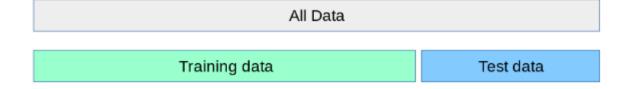
- Select best performing model from a group of models
 - Different algorithms
 - Different feature subsets

Select hyperparameters



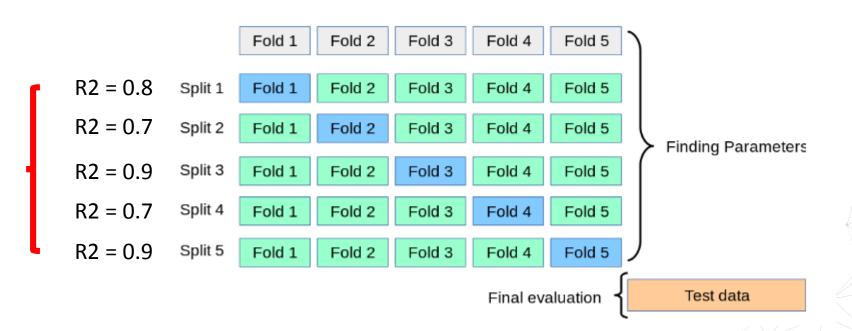
Cross-Validation & Generalization Error

When we have 1 model:



Generalization Error:

 $R2 = 0.8 \pm 0.1$

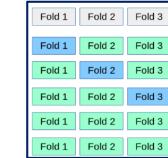


https://scikit-learn.org/stable/modules/cross_validation.html



Generalization Error:

 $R2 = 0.8 \pm 0.1$



Fold 1

Fold 1

Fold 1

Fold 1

Fold 1

Fold 1

Fold 2

Fold 2

Fold 2

Fold 2

Fold 2

Fold 2

Hyperparams 1

Generalization Error:

 $R2 = 0.84 \pm 0.06$

R2 = 0.8

R2 = 0.8

R2 = 0.8

R2 = 0.8

R2 = 0.7

R2 = 0.9

R2 = 0.7

R2 = 0.9

R2 = 0.9

R2 = 0.9

Generalization Error:

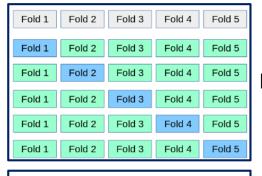
 $R2 = 0.78 \pm 0.11$

R2 = 0.7R2 = 0.7

R2 = 0.9

R2 = 0.7

R2 = 0.9



Fold 3

Fold 3

Fold 3

Fold 3

Fold 3

Fold 3

Fold 4

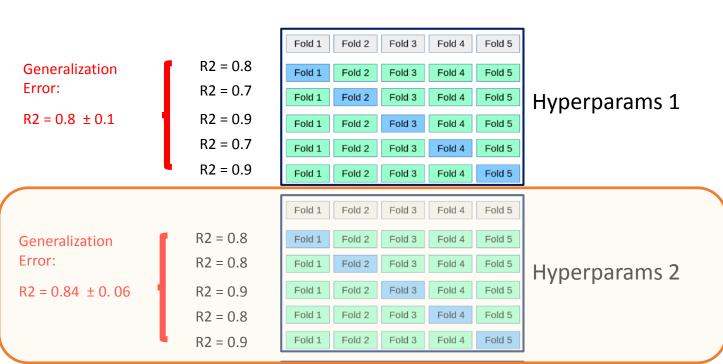
Fold 5

Hyperparams 2

Fold 5 Fold 5

Hyperparams 3





Generalization Error:

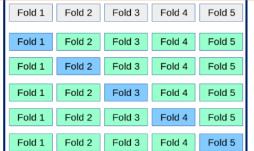
 $R2 = 0.78 \pm 0.11$

R2 = 0.7

R2 = 0.7

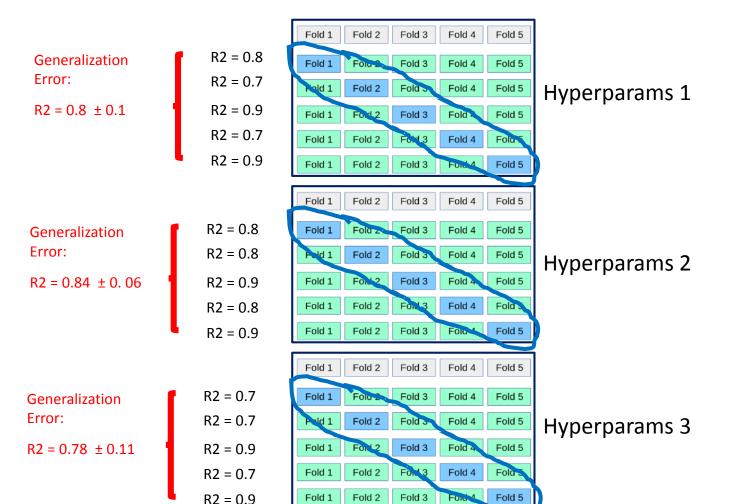
R2 = 0.9R2 = 0.7

R2 = 0.9



Hyperparams 3





Validation tests are the
 same for all models

- Validation set "leaks" information to model "selection" procedure
- Generalization error is optimistically biased



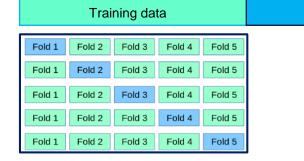
We need a *different* test set to get an **unbiased** evaluation of the generalization error of selected model (or selected hyperparameters for the purpose of this course)

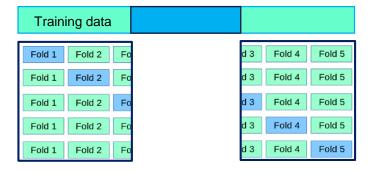
Particularly important for data science competitions.



Outer LoopEstimate
Generalization Error

All Data Training data Test data





Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 5 Fold 1 Fold 2 Fold 3 Fold 4 Fold 3 Fold 1 Fold 2 Fold 4 Fold 5 Fold 3 Fold 1 Fold 2 Fold 4 Fold 5 Fold 3 Fold 4 Fold 5 Fold 1 Fold 2

Training data

Inner Loop
Select Model –
Select hyperparameters



Final evaluation

Test data

ainindata.com| 9

Outer LoopEstimate
Generalization Error

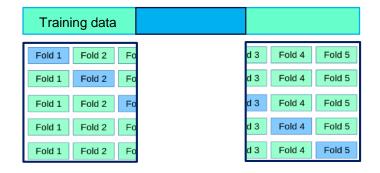
All Data

Training data

Test data

Training data Fold 2 Fold 1 Fold 3 Fold 4 Fold 5 Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 4 Fold 2 Fold 3 Fold 5 Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 1 Fold 4 Fold 1 Fold 2 Fold 3 Fold 5

- Train all models
- Select Hyperparams
 - Asses error on outer loop



Training data

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5

Inner Loop

Select Model – Select hyperparameters



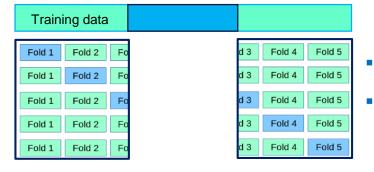
Final evaluation

Test data

inindata.com | 10

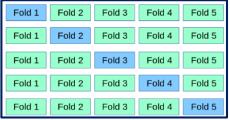
Outer Loop
Estimate
Generalization Error

All Data Training data Test data Training data Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 4 Fold 2 Fold 5 Fold 1 Fold 3 Fold 2 Fold 3 Fold 4 Fold 5 Fold 1 Fold 4 Fold 1 Fold 2 Fold 3 Fold 5



- Train all models
- Select Hyperparams
 - Asses error on outer loop

Training data



Inner Loop

Select Model – Select hyperparameters



Final evaluation

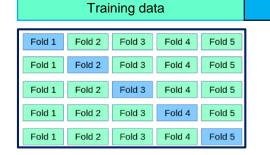
Test data

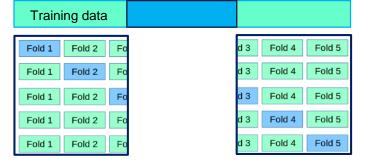
Training data

All Data

Test data

Outer Loop
Estimate
Generalization Error





Training data

- Train all models
- Select Hyperparams
- Asses error on outer loop

Fold 1 Fold 2 Fold 3 Fold 4 Fold 5 Fold 5 Fold 1 Fold 2 Fold 3 Fold 4 Fold 3 Fold 1 Fold 2 Fold 4 Fold 5 Fold 3 Fold 1 Fold 2 Fold 4 Fold 5 Fold 3 Fold 4 Fold 1 Fold 2 Fold 5

Inner Loop
Select Model –
Select hyperparameters



Final evaluation

Test data

Nested CV, considerations

Computationally expensive

- Useful when we need a good estimation of the generalization error
- Different inner models may have different hyperparameters, although I would expect to be among the top performing hyperparameters.





THANK YOU

www.trainindata.com