Cross-validation

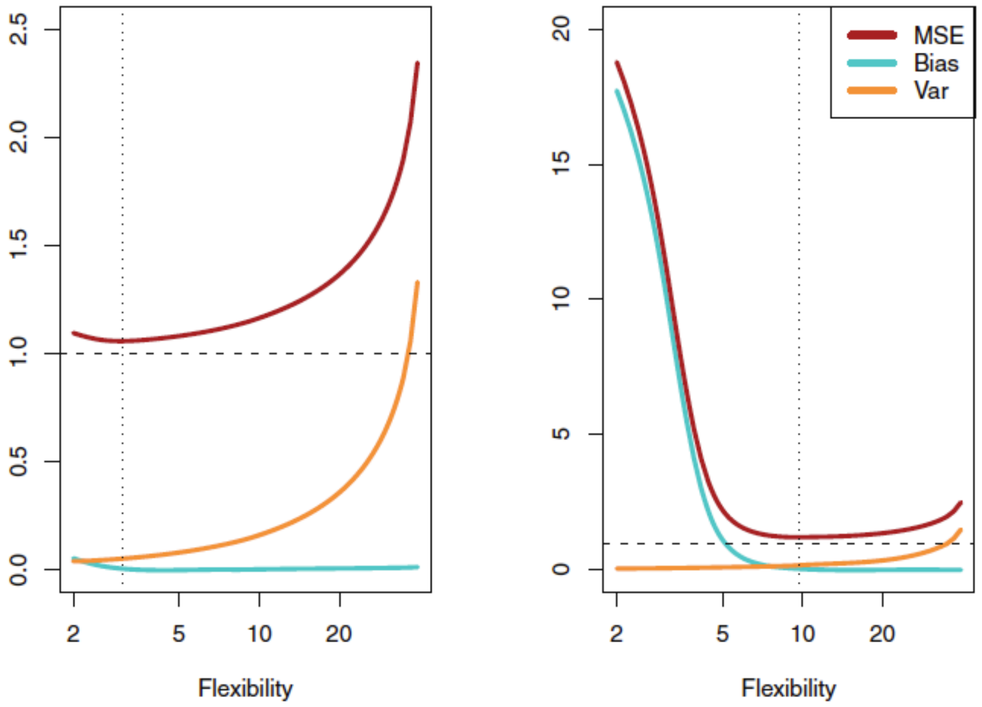
STAT 4710

September 22, 2022

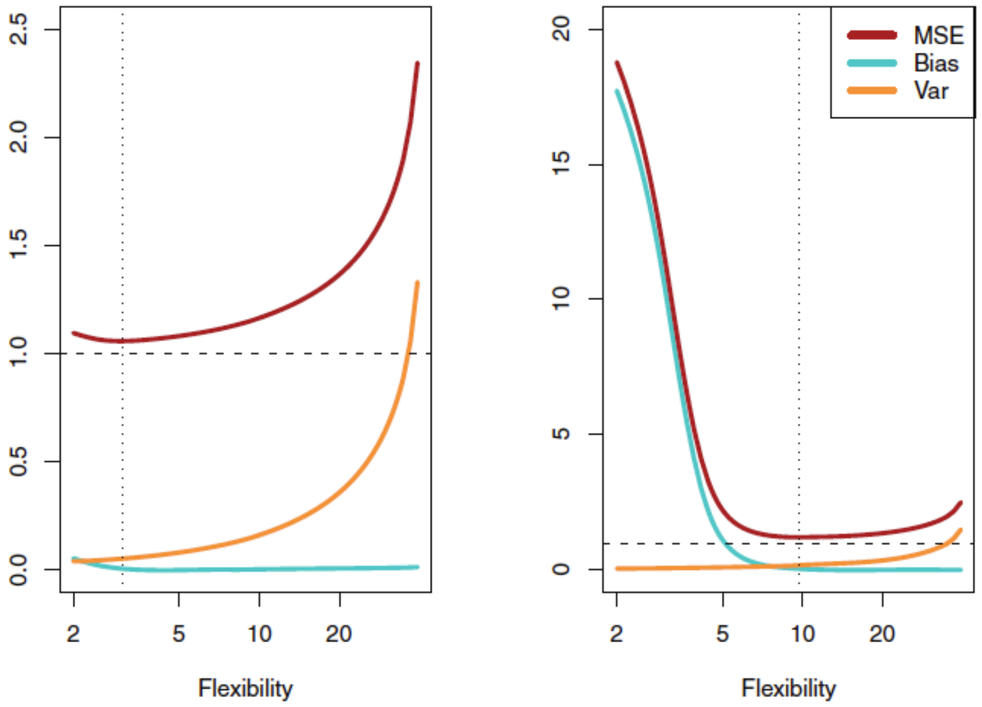
Where we are

- ✓ **Unit 1:** R for data mining
- **Unit 2:** Prediction fundamentals
  - **Lecture 1:** Model complexity
  - **Lecture 2:** Bias-variance trade-off
  - **Lecture 3:** Cross-validation
  - **Lecture 4:** Classification
  - **Lecture 5:** Unit review and quiz in class
- **Unit 3:** Regression-based methods
- **Unit 4:** Tree-based methods
- **Unit 5:** Deep learning

Navigating the bias-variance tradeoff in practice

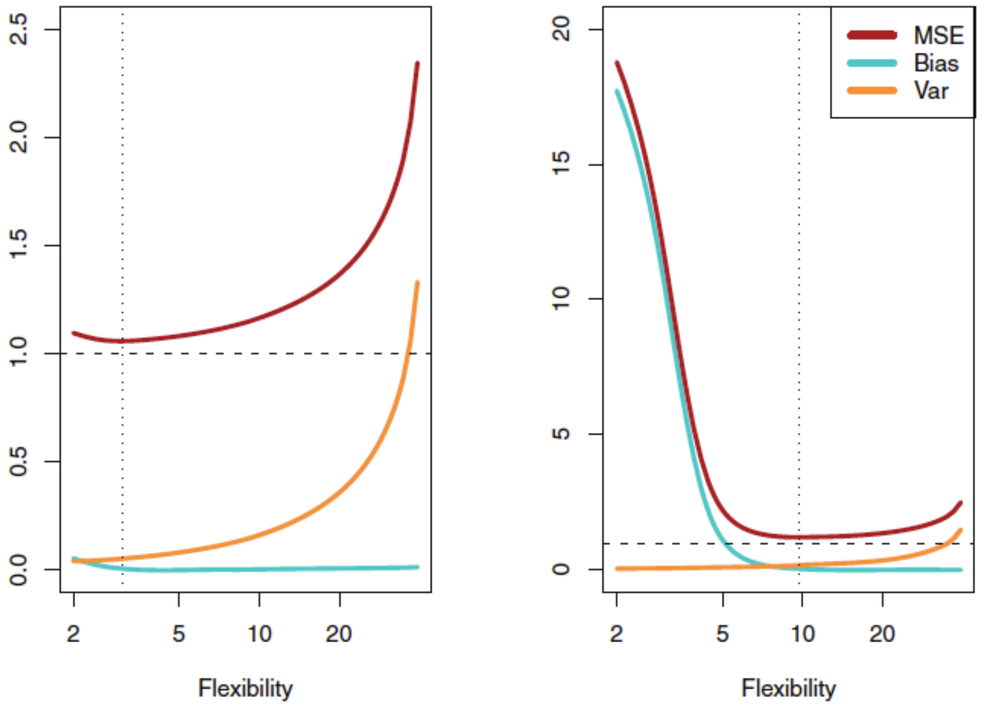


Navigating the bias-variance tradeoff in practice



Estimate test error for each model complexity

Navigating the bias-variance tradeoff in practice

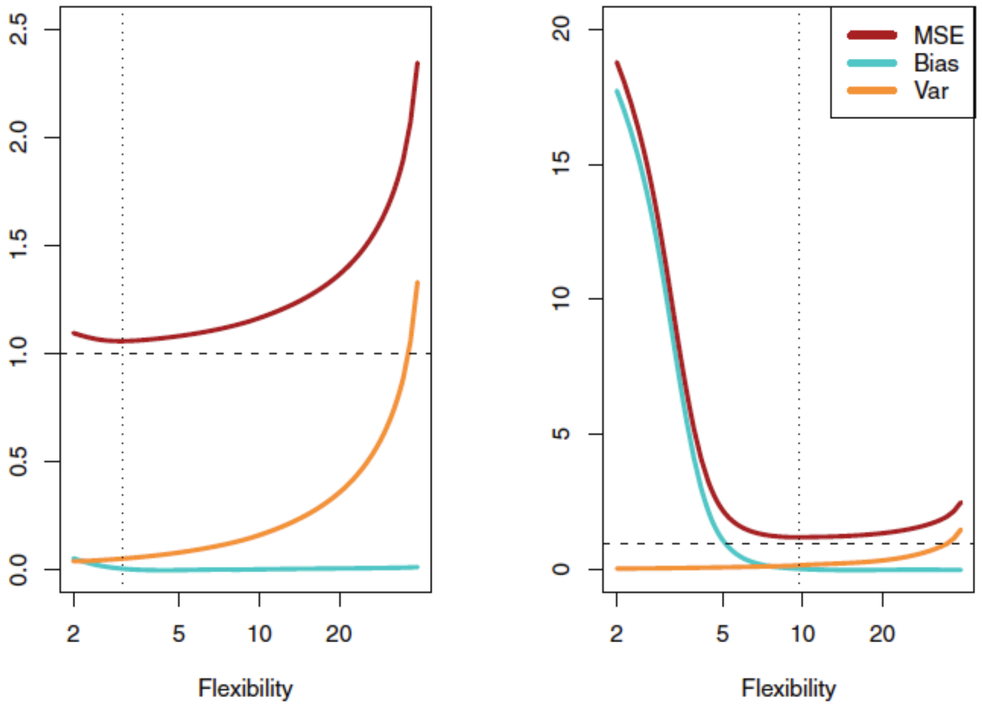


Estimate test error for each model complexity

↓

Estimate the best model complexity

Navigating the bias-variance tradeoff in practice



Estimate test error for each model complexity

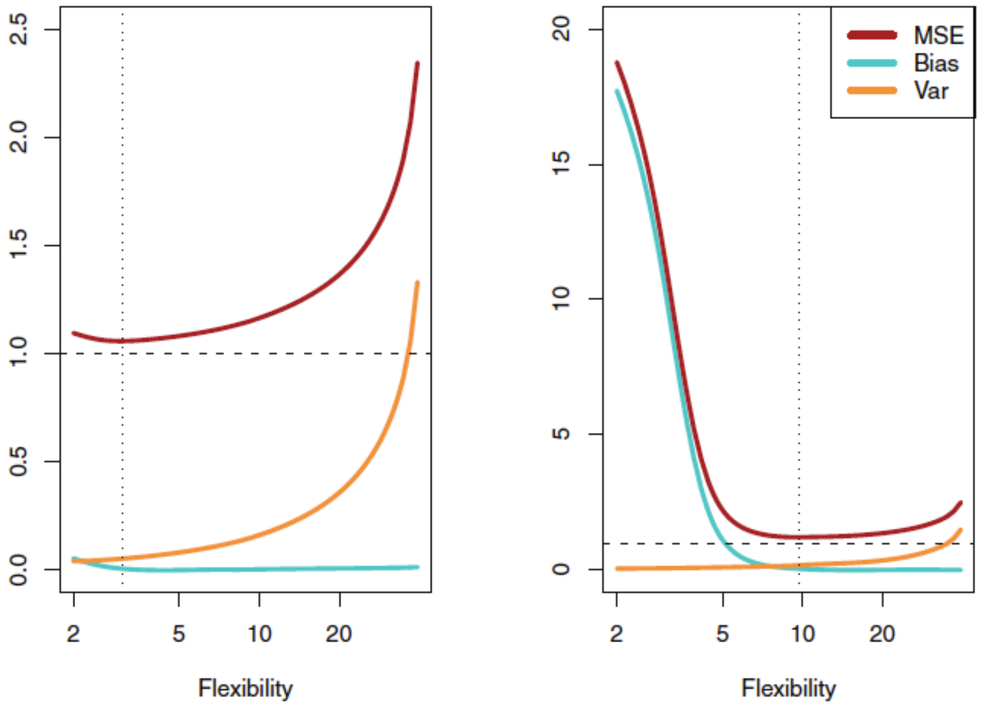
↓

Estimate the best model complexity

↓

Fit final predictive model using chosen df

Navigating the bias-variance tradeoff in practice



Estimate test error for each model complexity

↓

Estimate the best model complexity

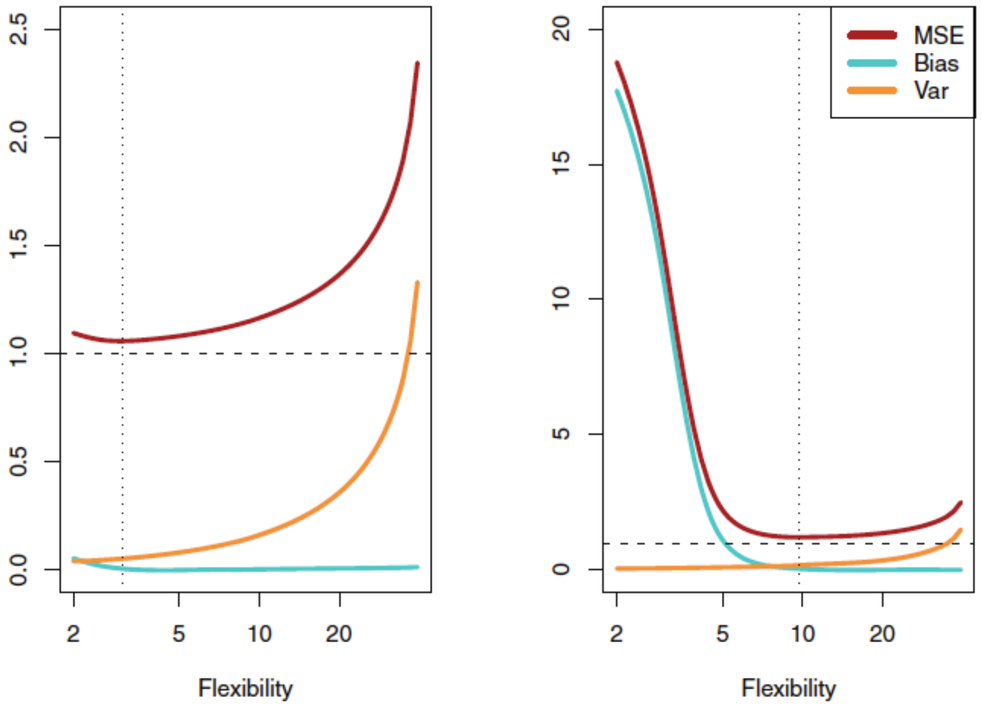
↓

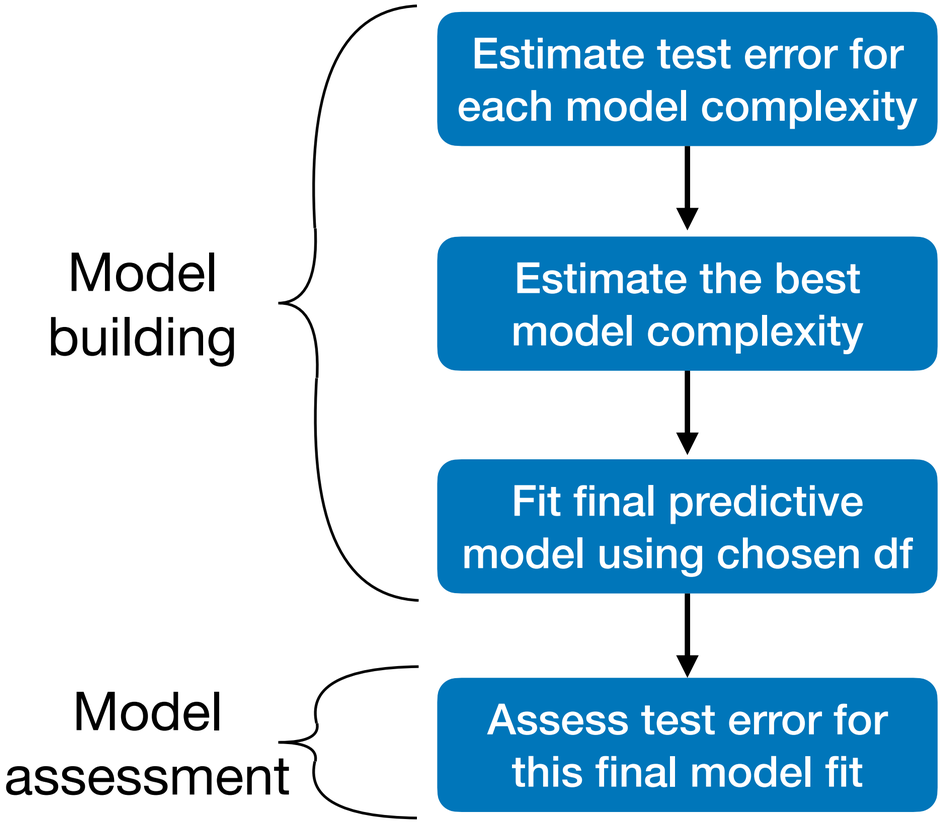
Fit final predictive model using chosen df

↓

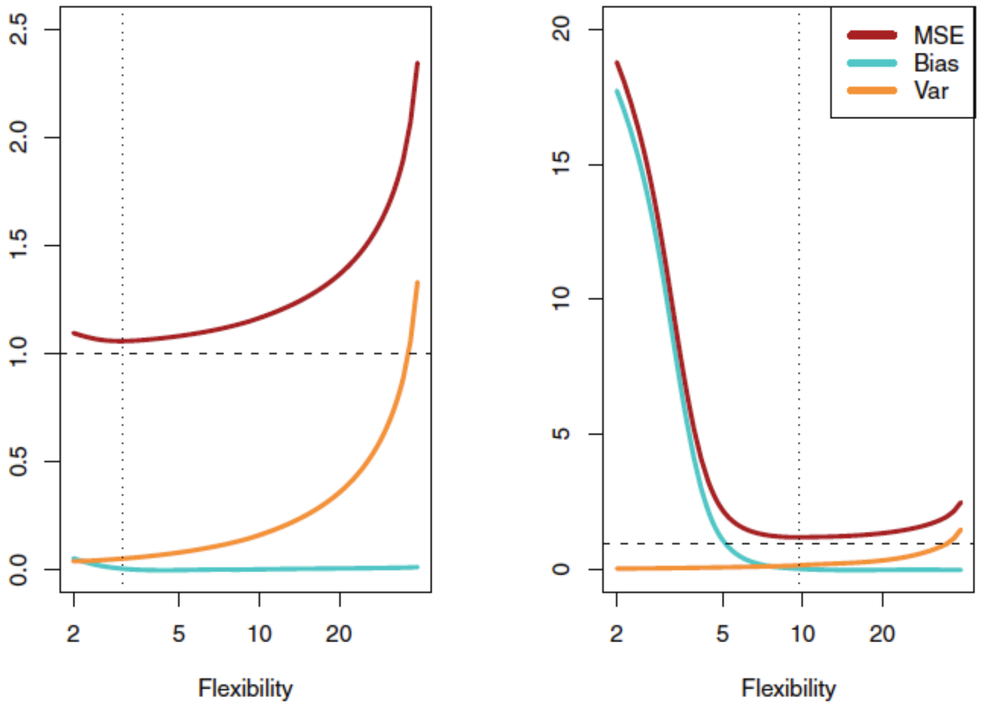
Assess test error for this final model fit

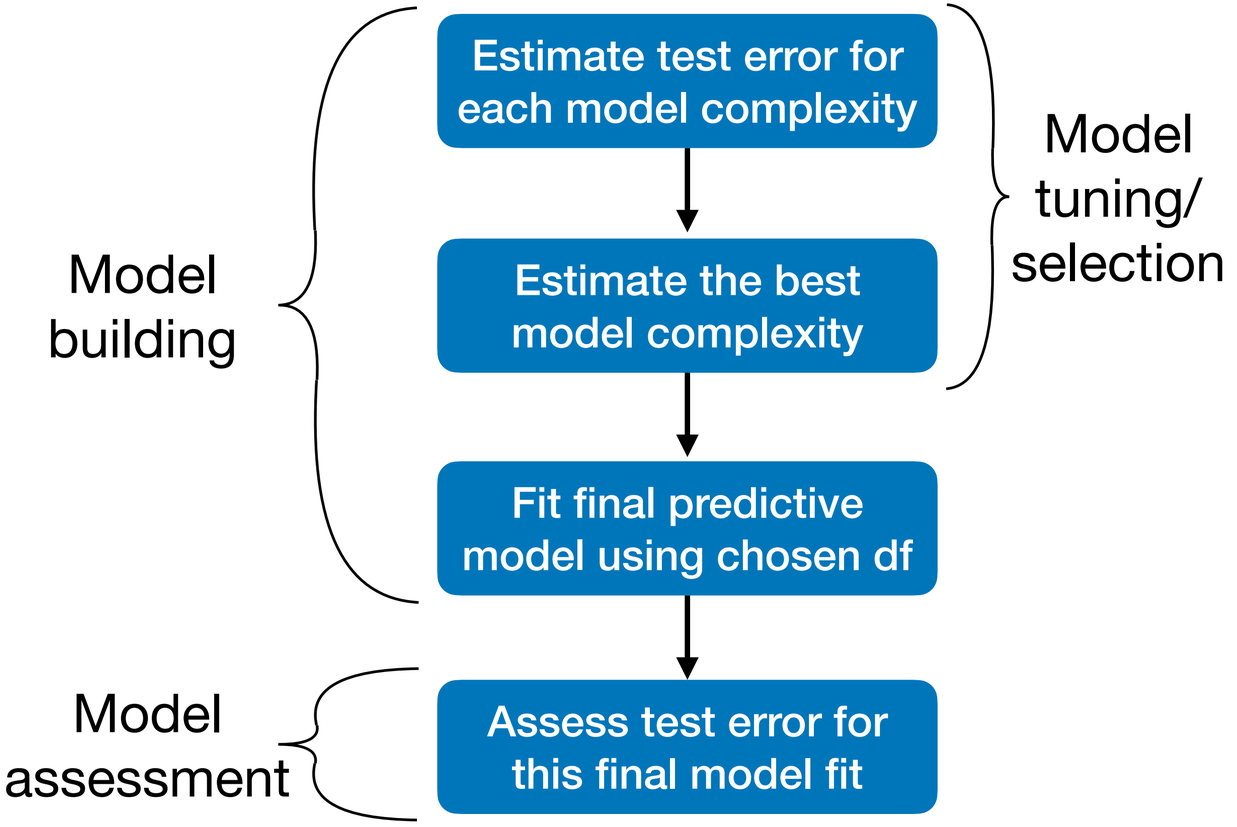
Navigating the bias-variance tradeoff in practice



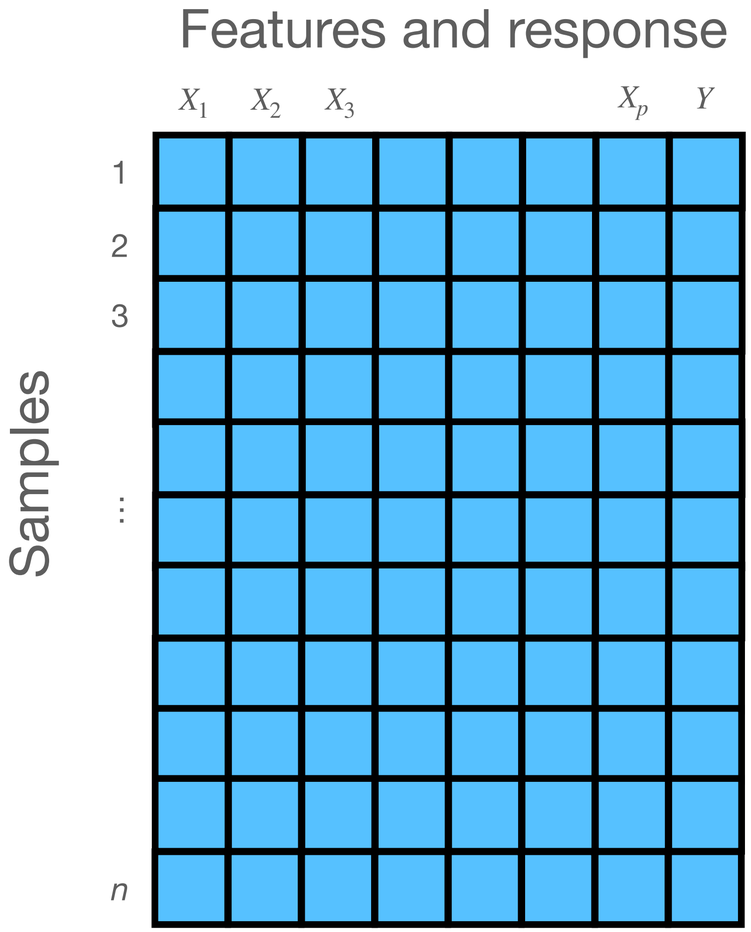


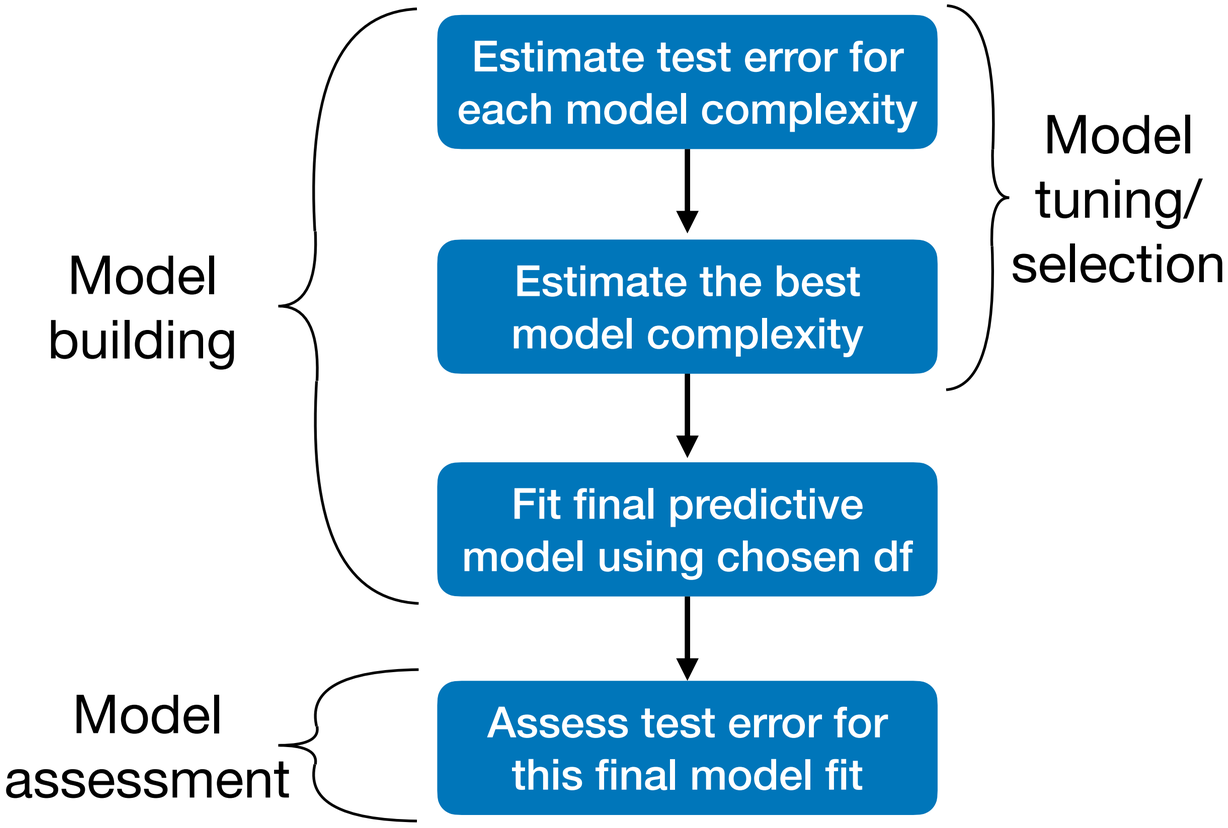
Navigating the bias-variance tradeoff in practice



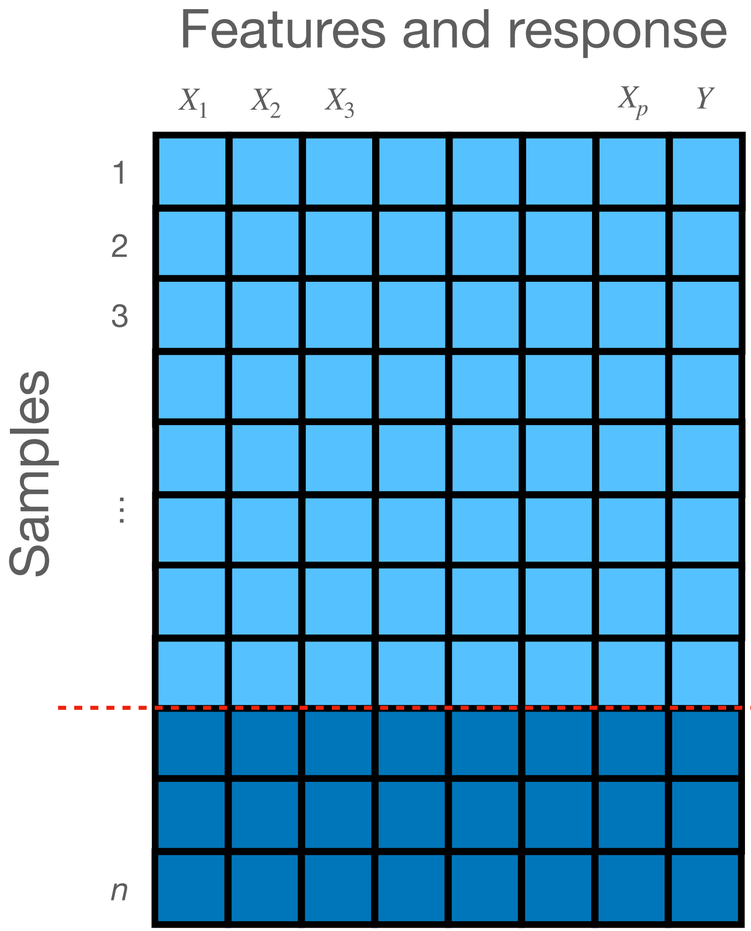


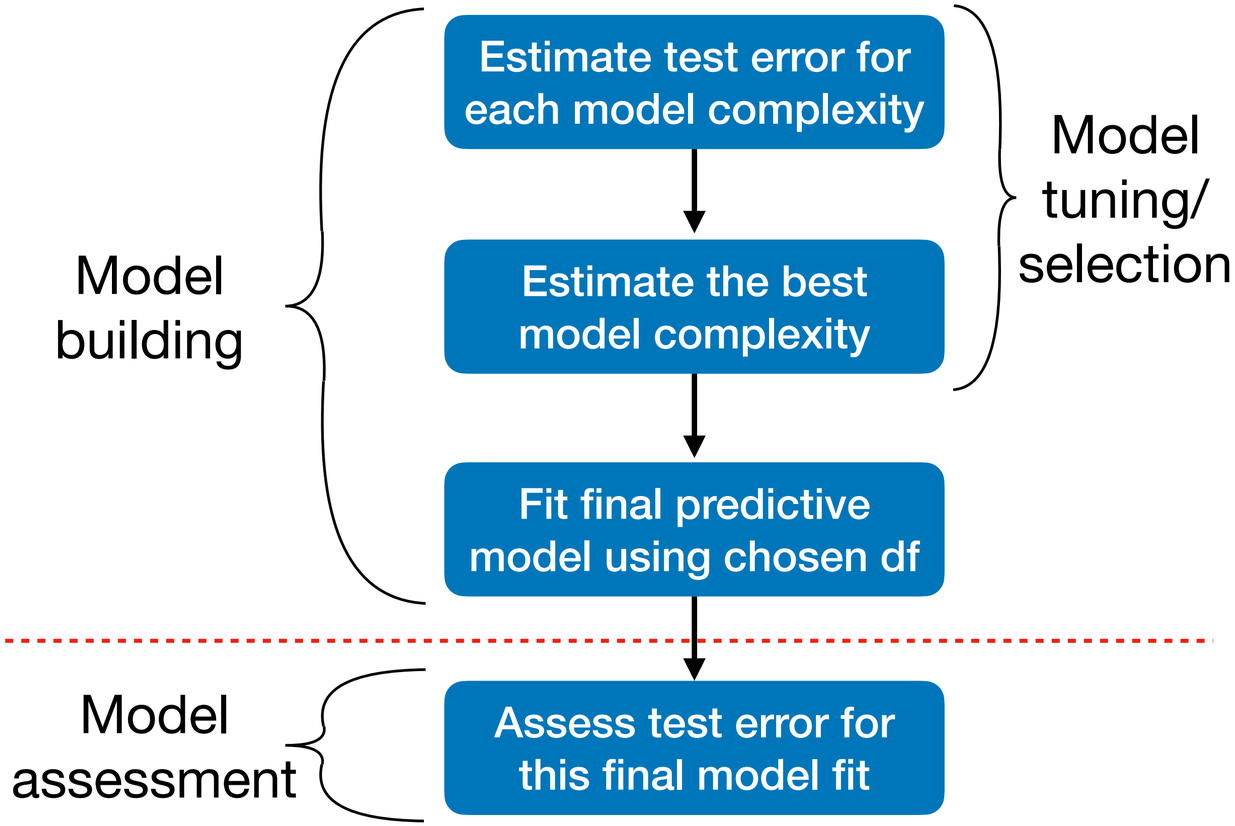
Separating data for model building and assessment



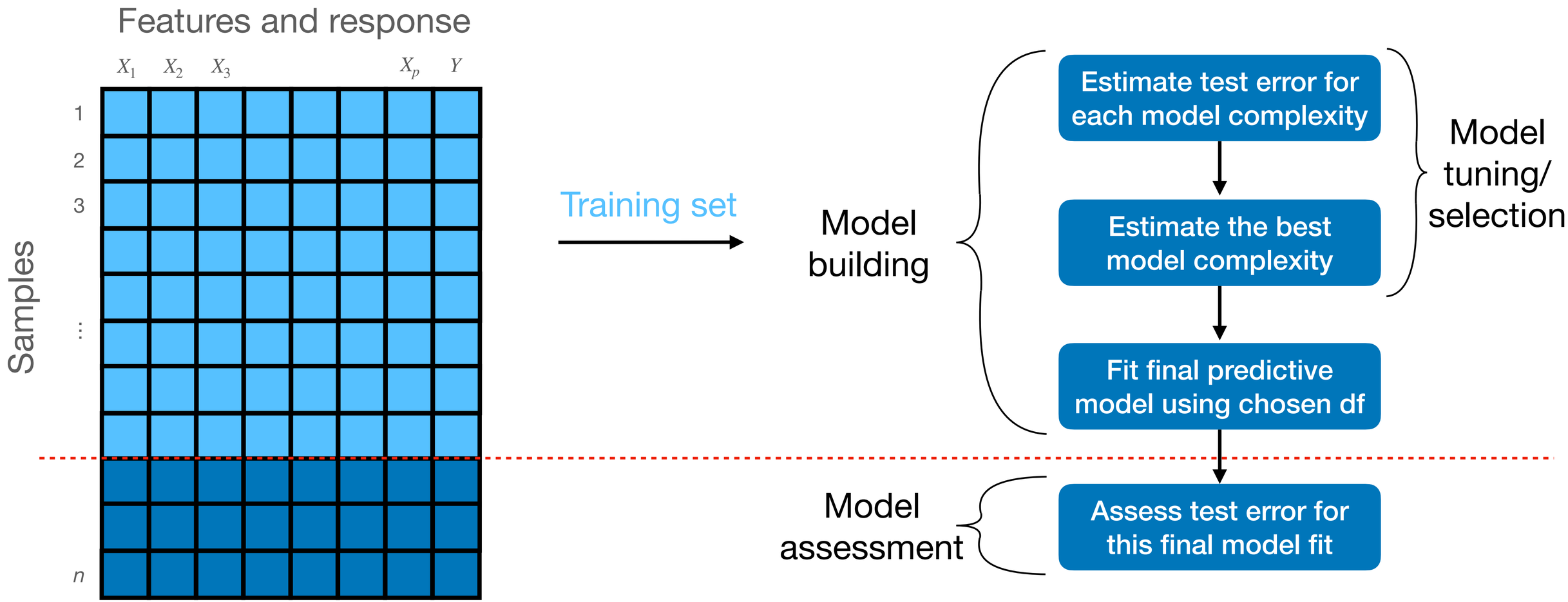


Separating data for model building and assessment

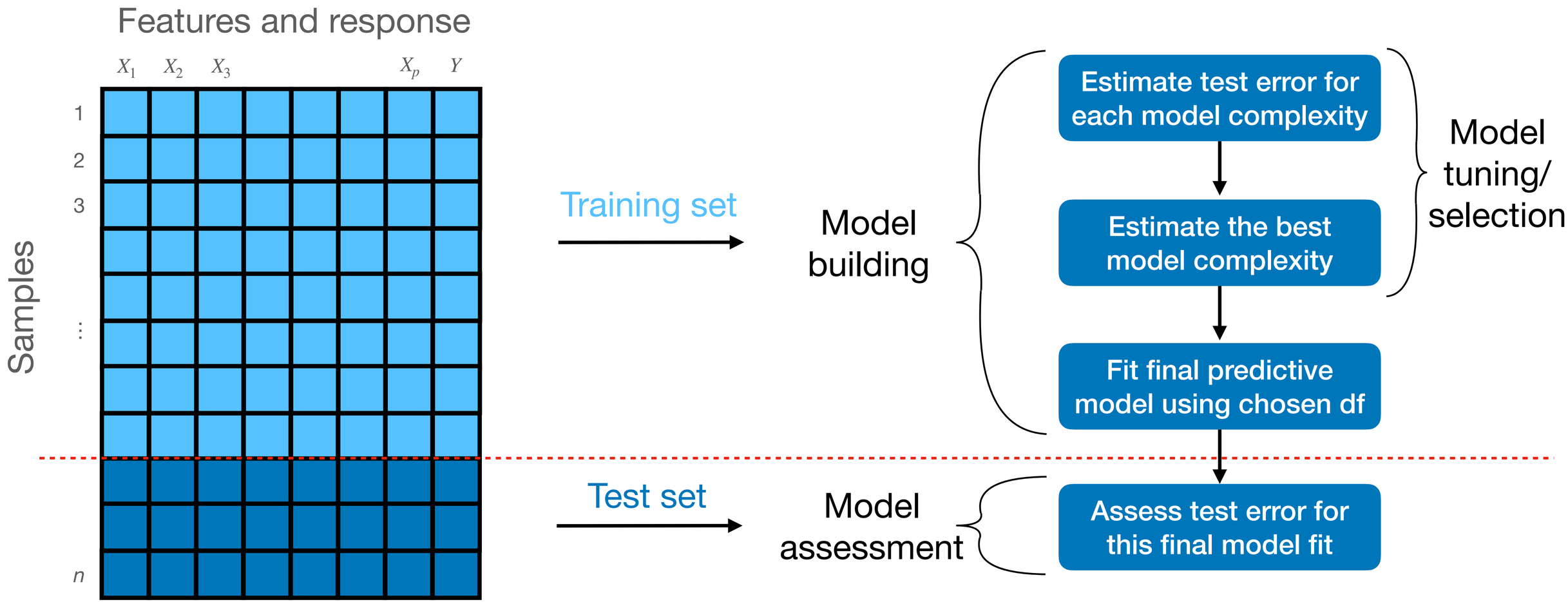




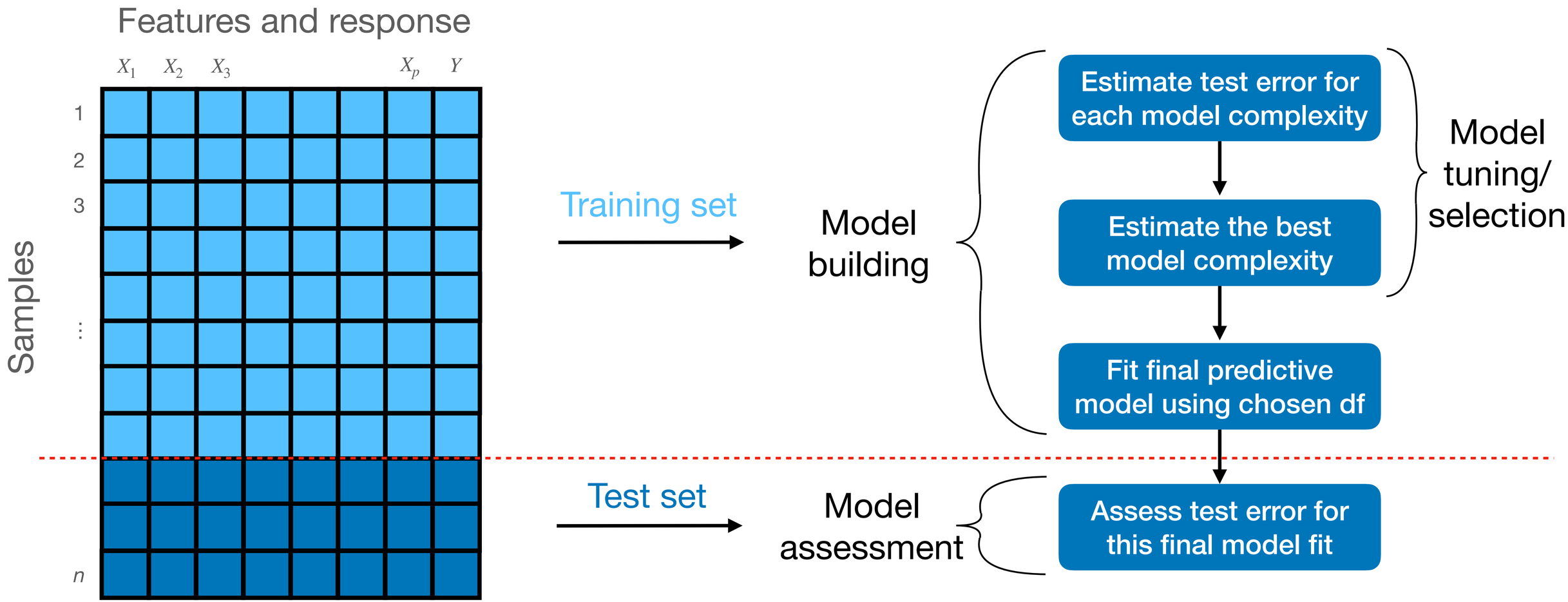
Separating data for model building and assessment



Separating data for model building and assessment



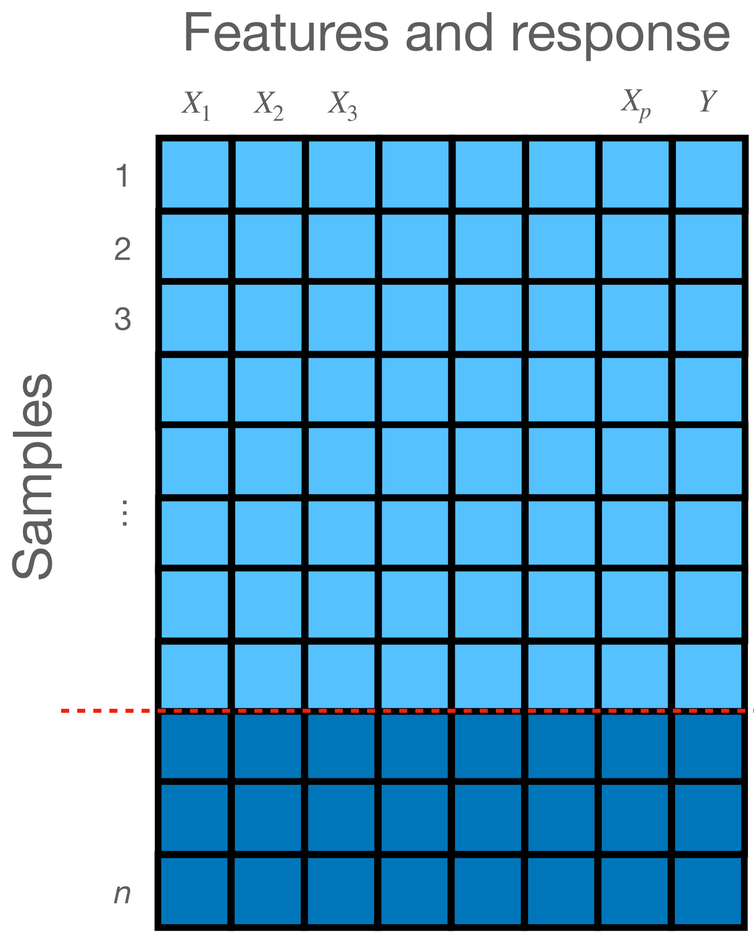
Separating data for model building and assessment



More samples for training → better fitted model;

More samples for testing → better estimate of test error.

Separating data for model building and assessment



Training set

(50%-80% of data)

Model building

Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Model tuning/ selection

Test set

(20%-50% of data)

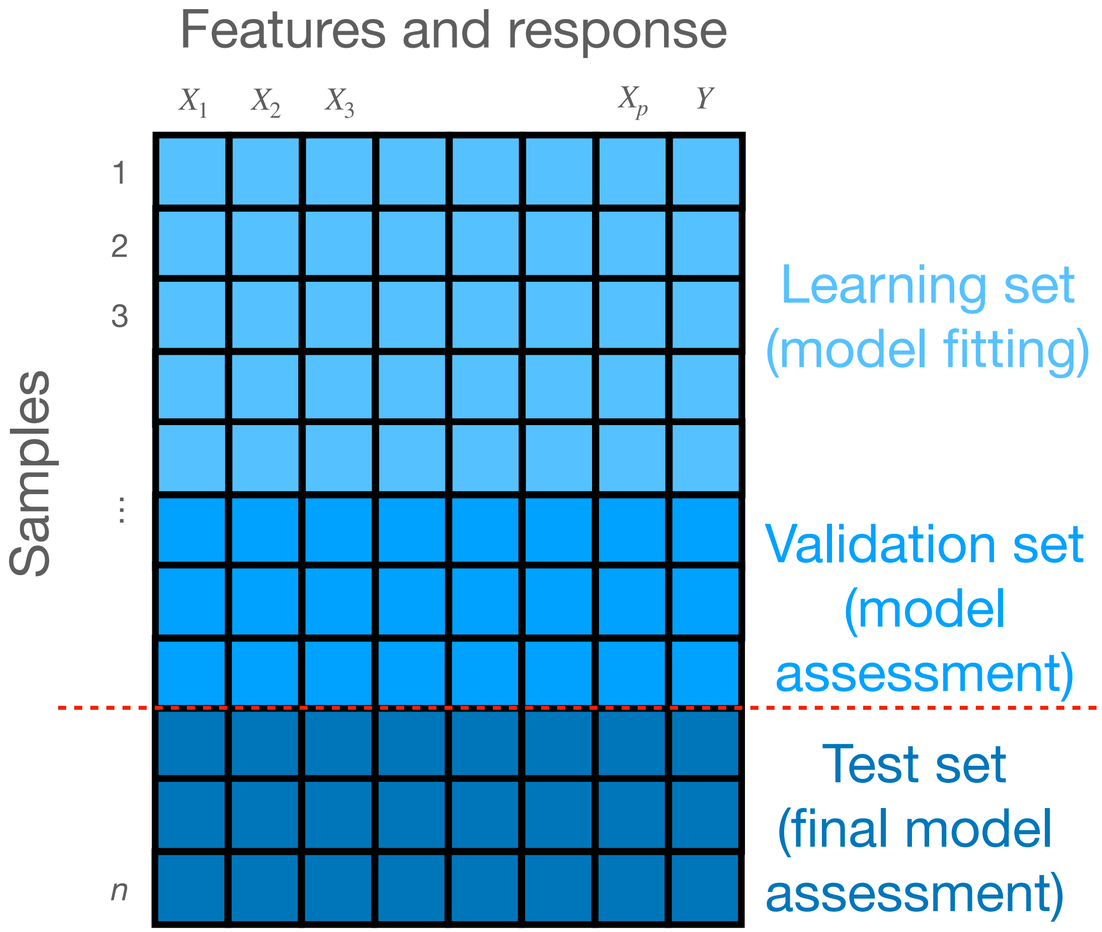
Model assessment

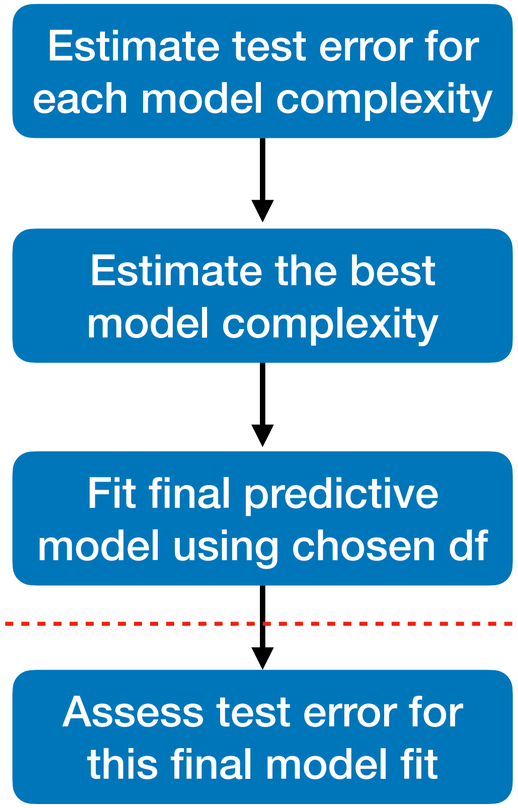
Assess test error for this final model fit

More samples for training → better fitted model;

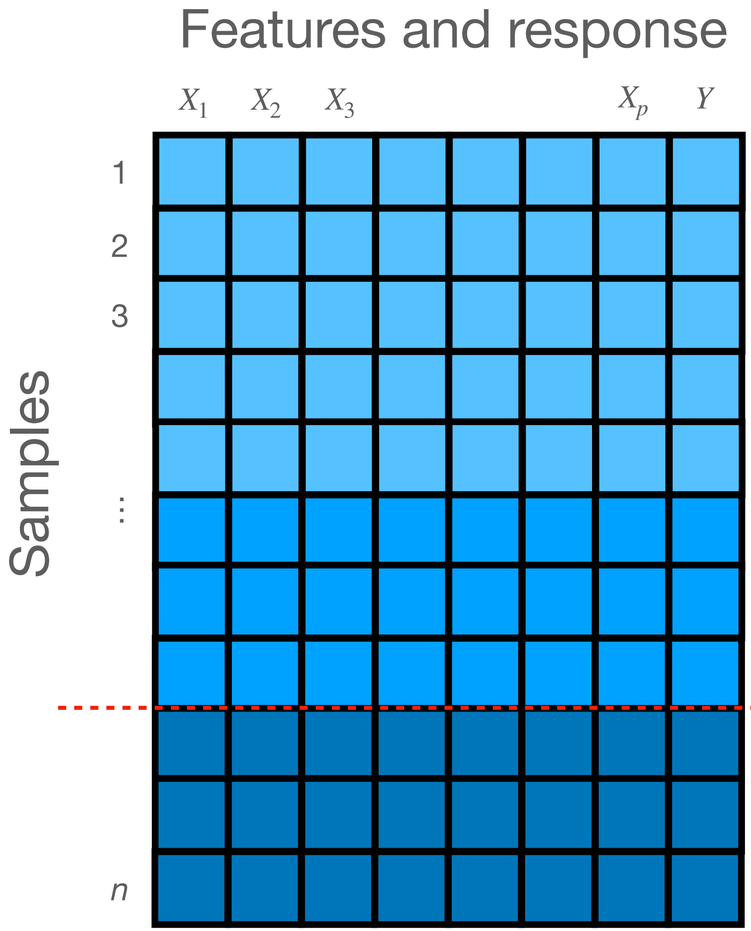
More samples for testing → better estimate of test error.

Validation set approach for model selection





Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set

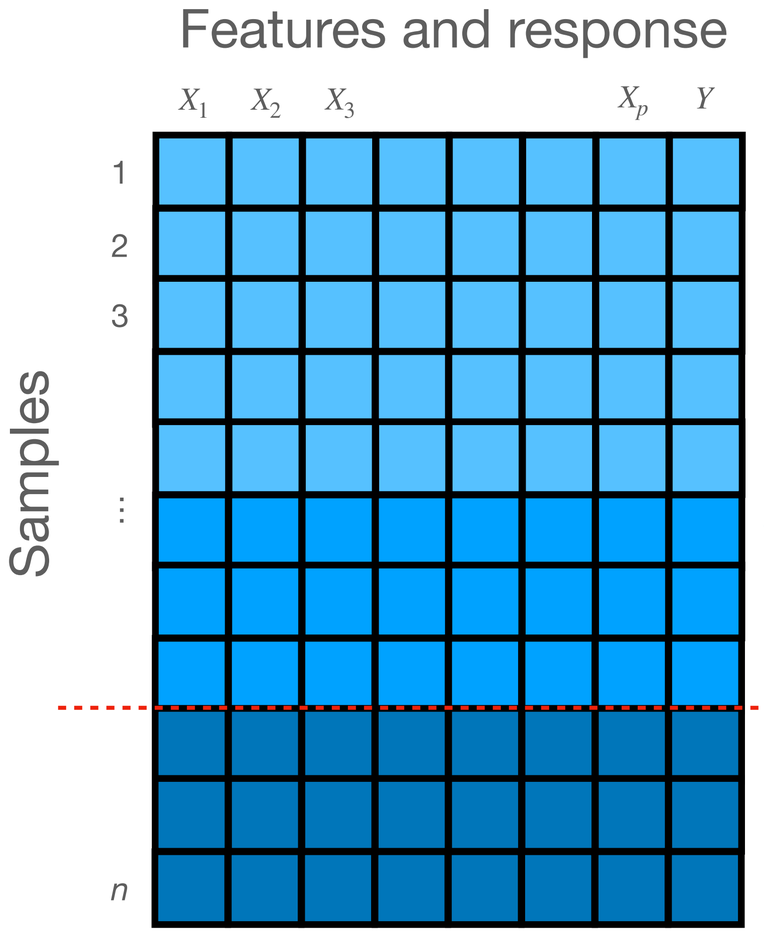
Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Assess test error for this final model fit

Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set
2. Estimate test error for each model on validation set

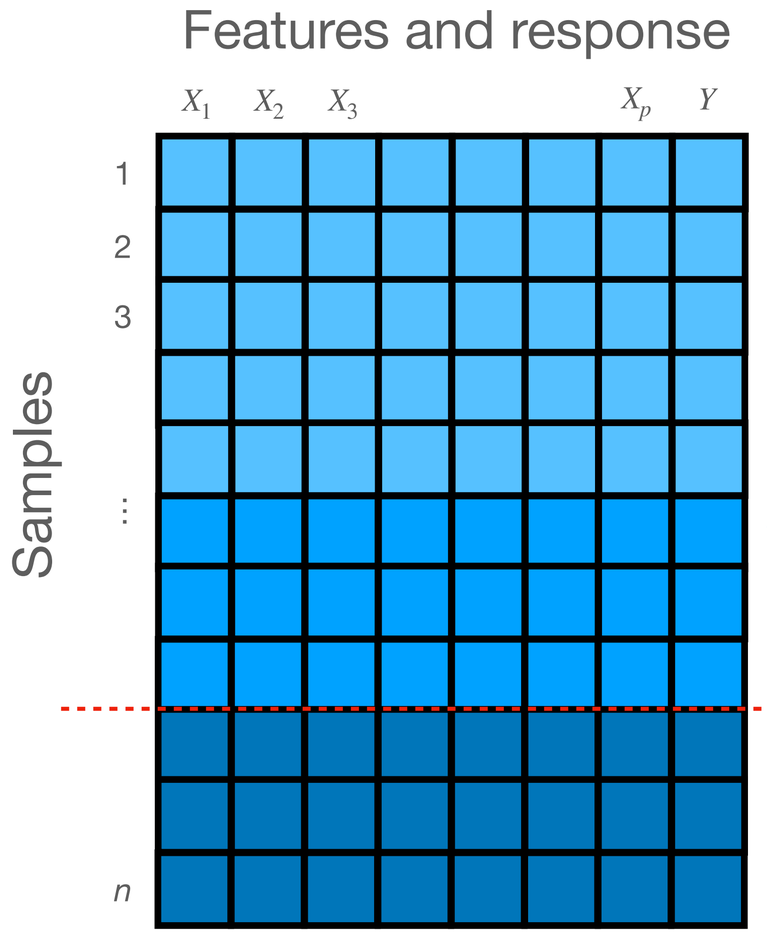
Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Assess test error for this final model fit

Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set
2. Estimate test error for each model on validation set
3. Choose model complexity to minimize validation error

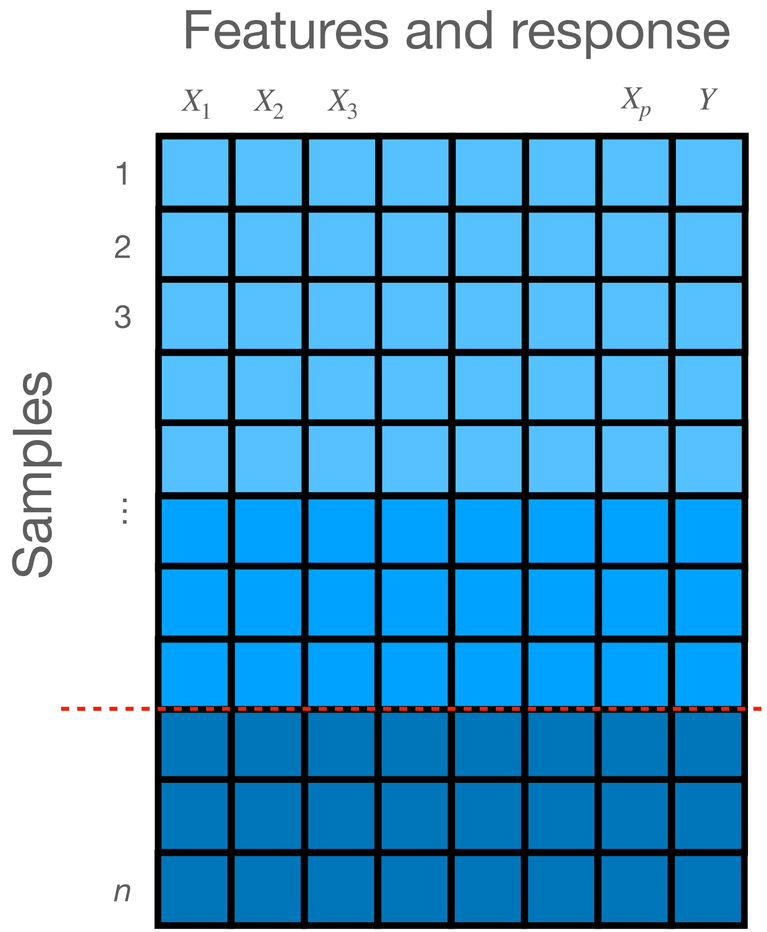
Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Assess test error for this final model fit

Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set
2. Estimate test error for each model on validation set
3. Choose model complexity to minimize validation error
4. Refit this model on combined learning and validation sets

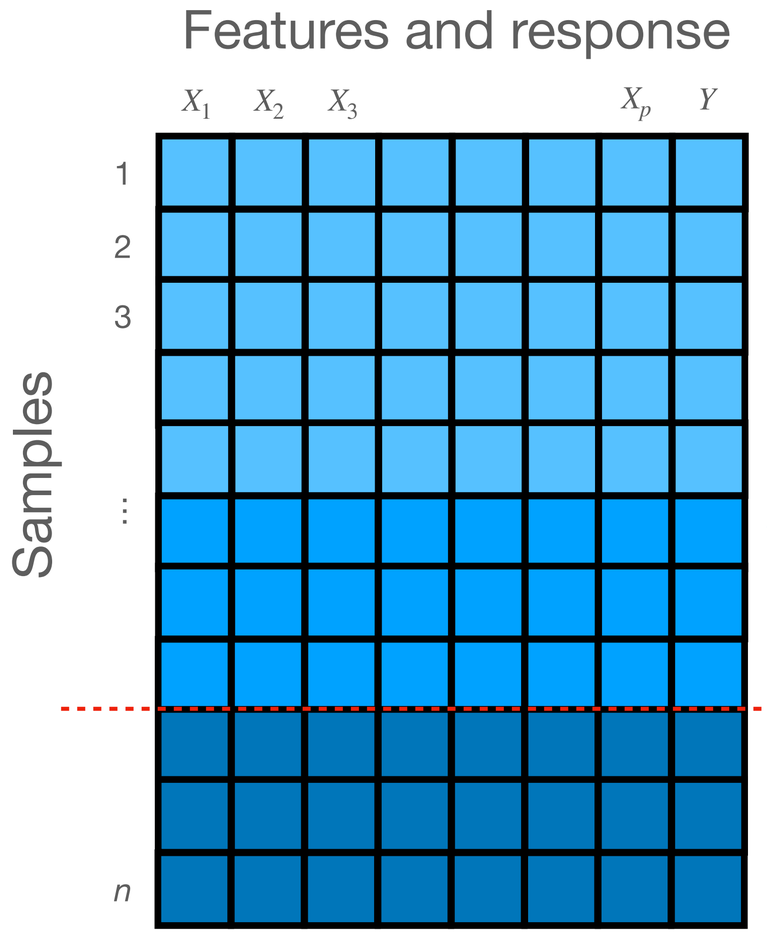
Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Assess test error for this final model fit

Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set
2. Estimate test error for each model on validation set
3. Choose model complexity to minimize validation error
4. Refit this model on combined learning and validation sets
5. Evaluate the final model on the test set

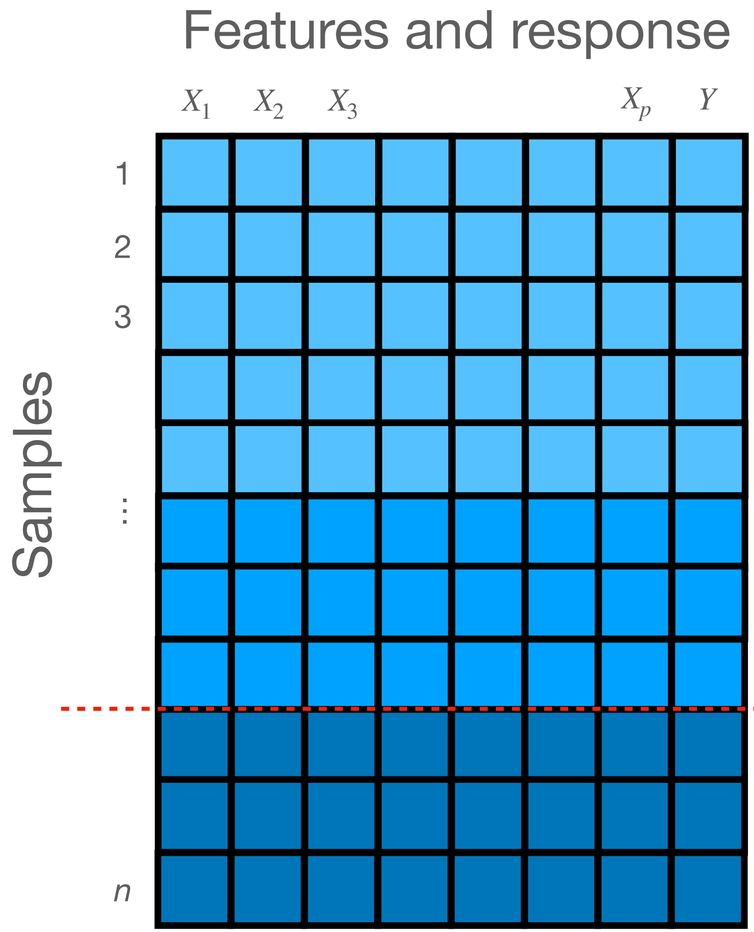
Estimate test error for each model complexity

Estimate the best model complexity

Fit final predictive model using chosen df

Assess test error for this final model fit

Validation set approach for model selection



Learning set (model fitting)

Validation set (model assessment)

Test set (final model assessment)

1. Fit models of varying complexity to learning set
2. Estimate test error for each model on validation set
3. Choose model complexity to minimize validation error
4. Refit this model on combined learning and validation sets
5. Evaluate the final model on the test set

Estimate test error for each model complexity

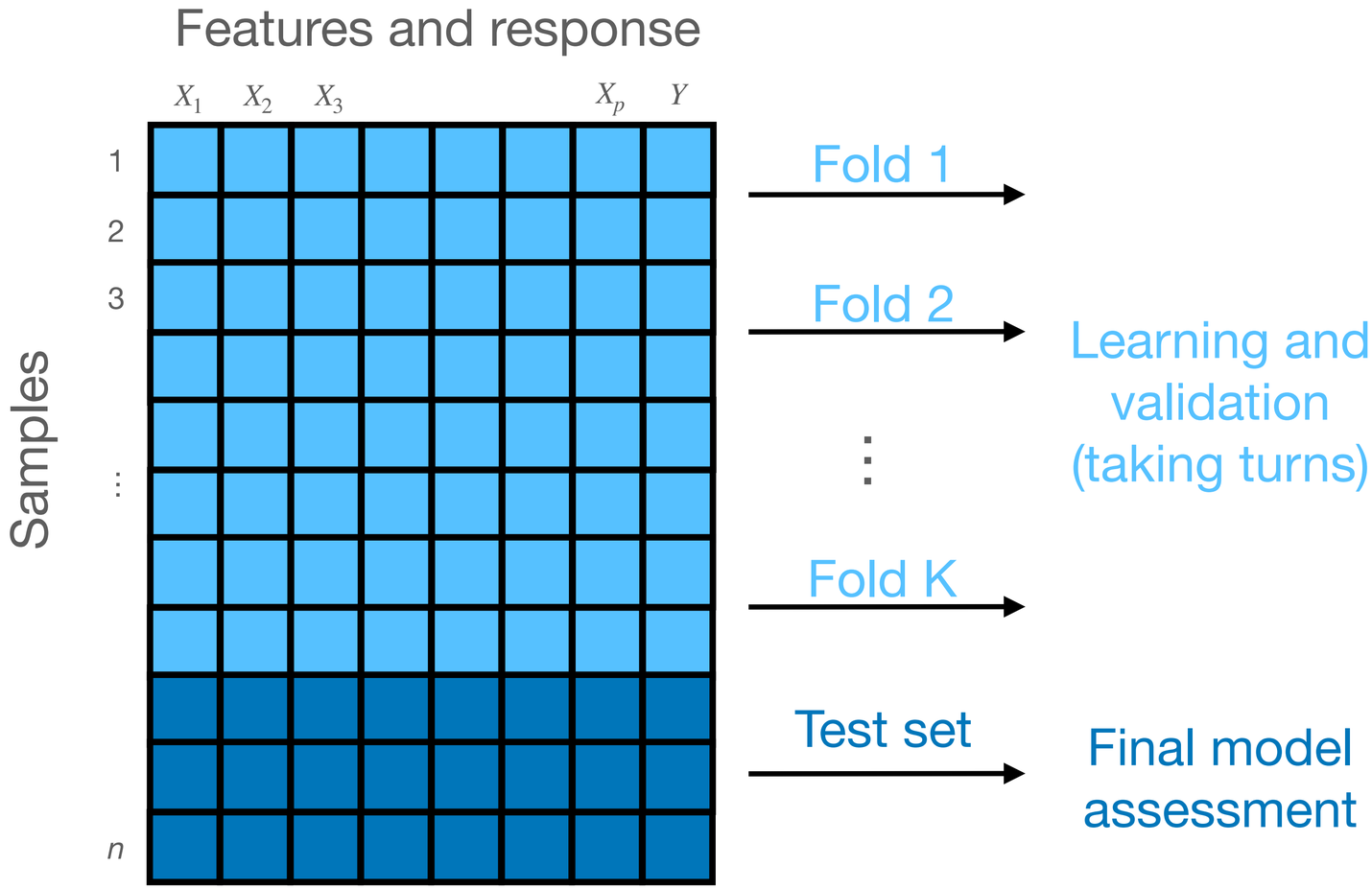
Estimate the best model complexity

Fit final predictive model using chosen df

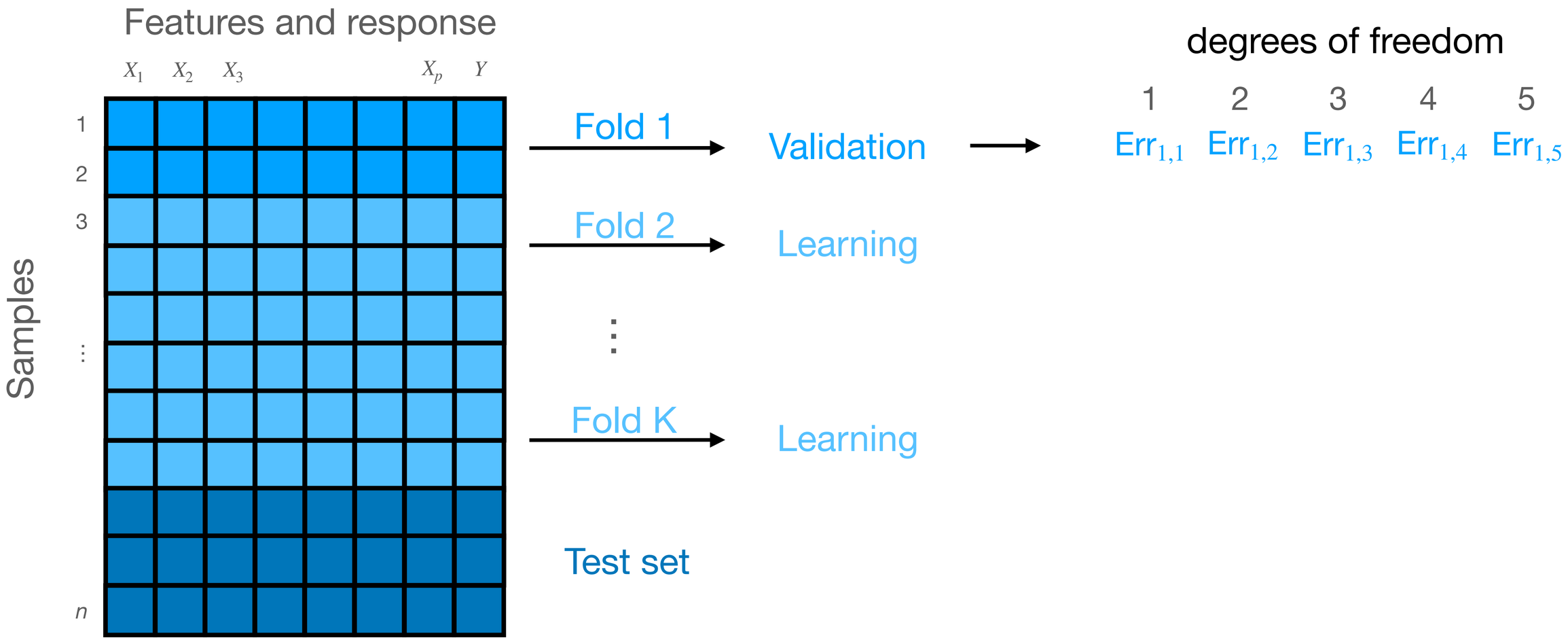
Assess test error for this final model fit

Drawback: Inefficient use of training samples,
e.g. small validation set may lead to poor model selection.

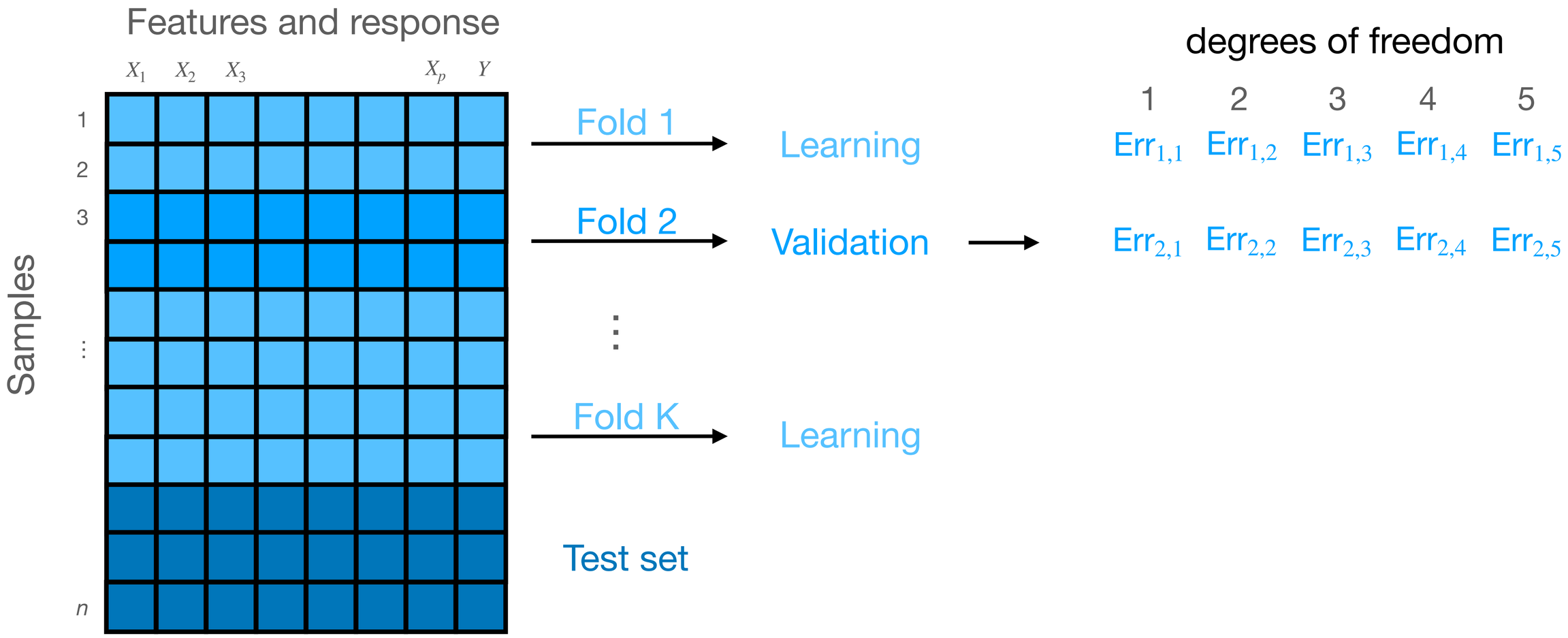
Cross-validation for model selection



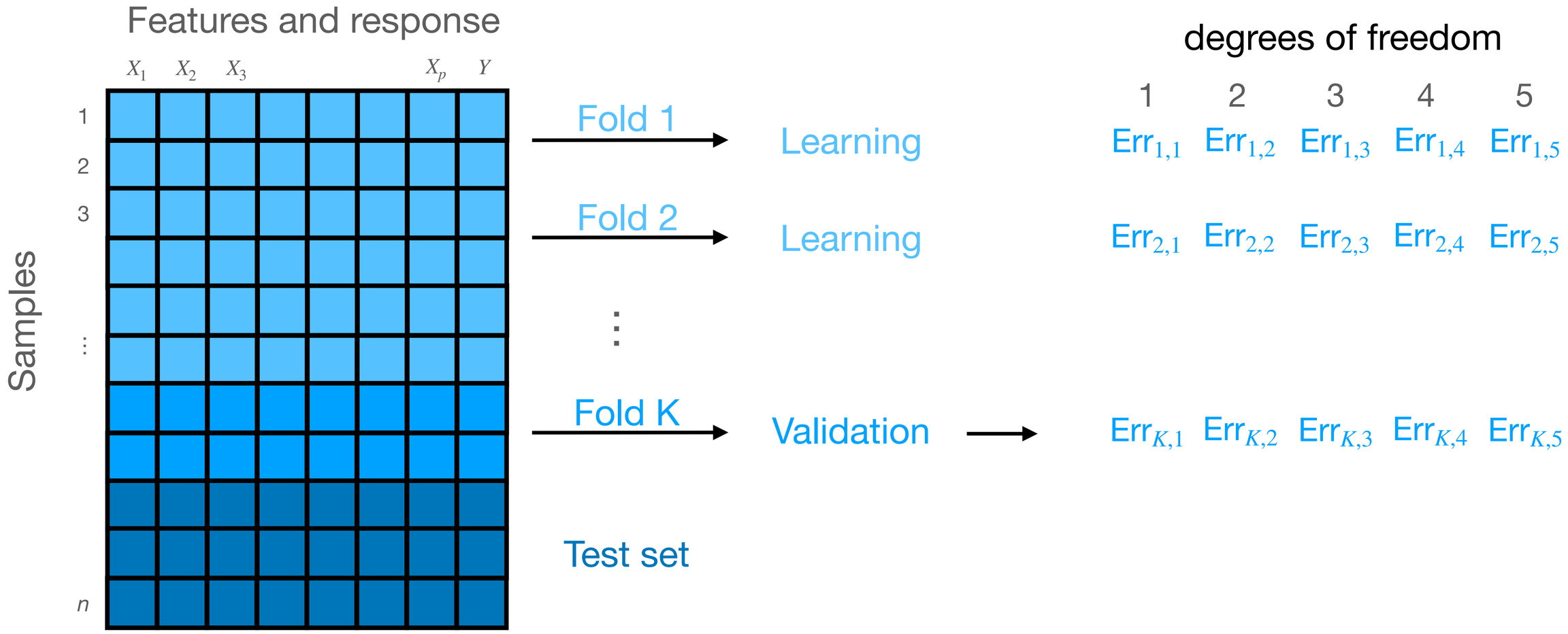
Cross-validation for model selection



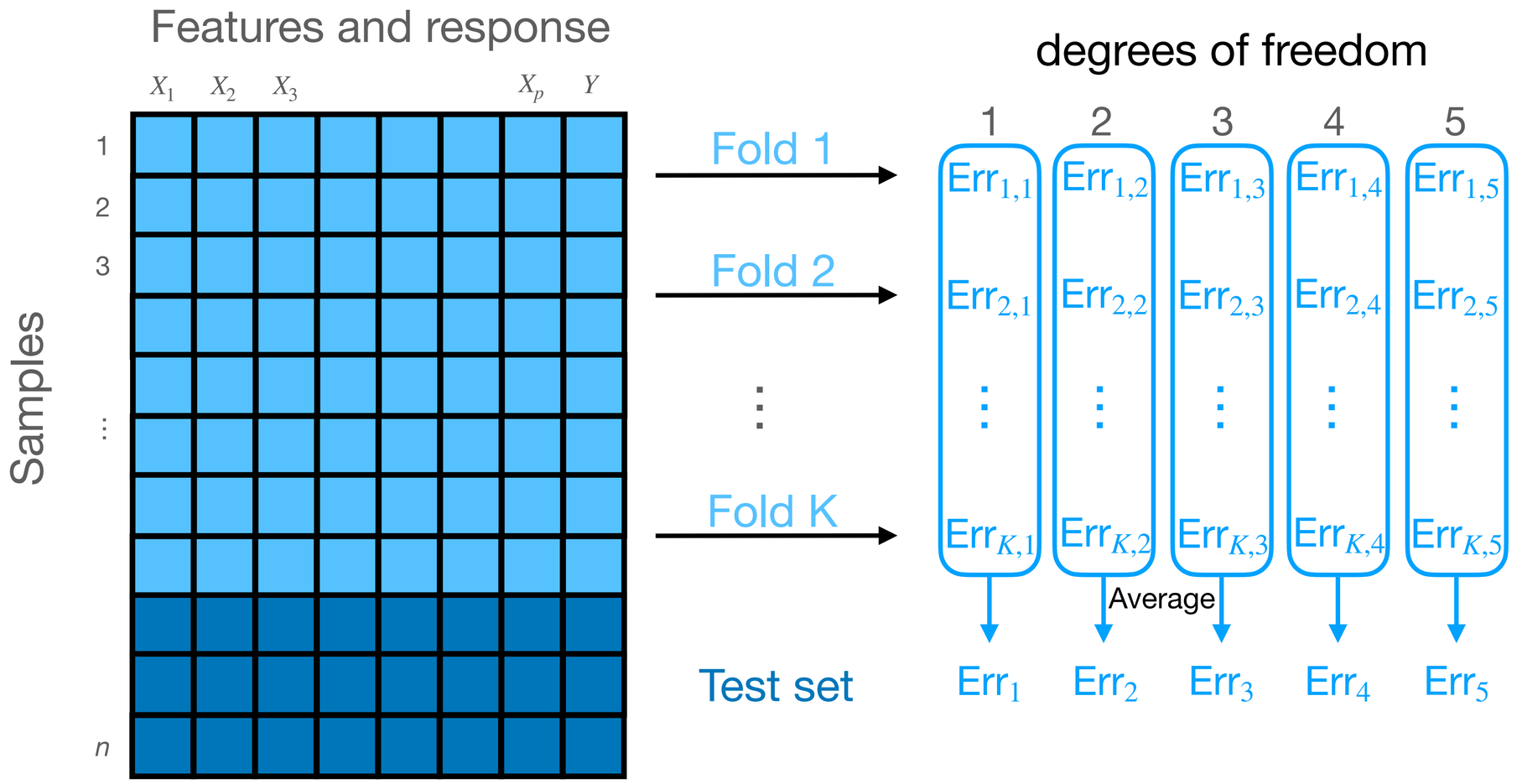
Cross-validation for model selection



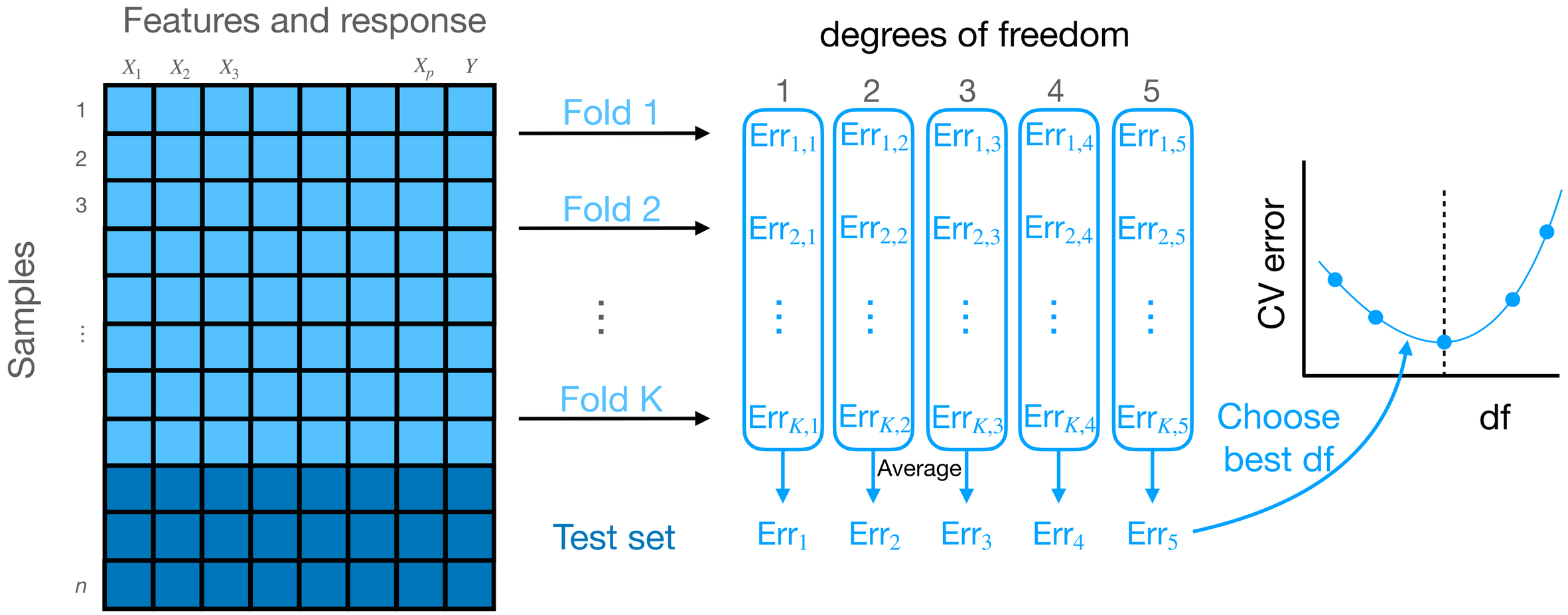
Cross-validation for model selection



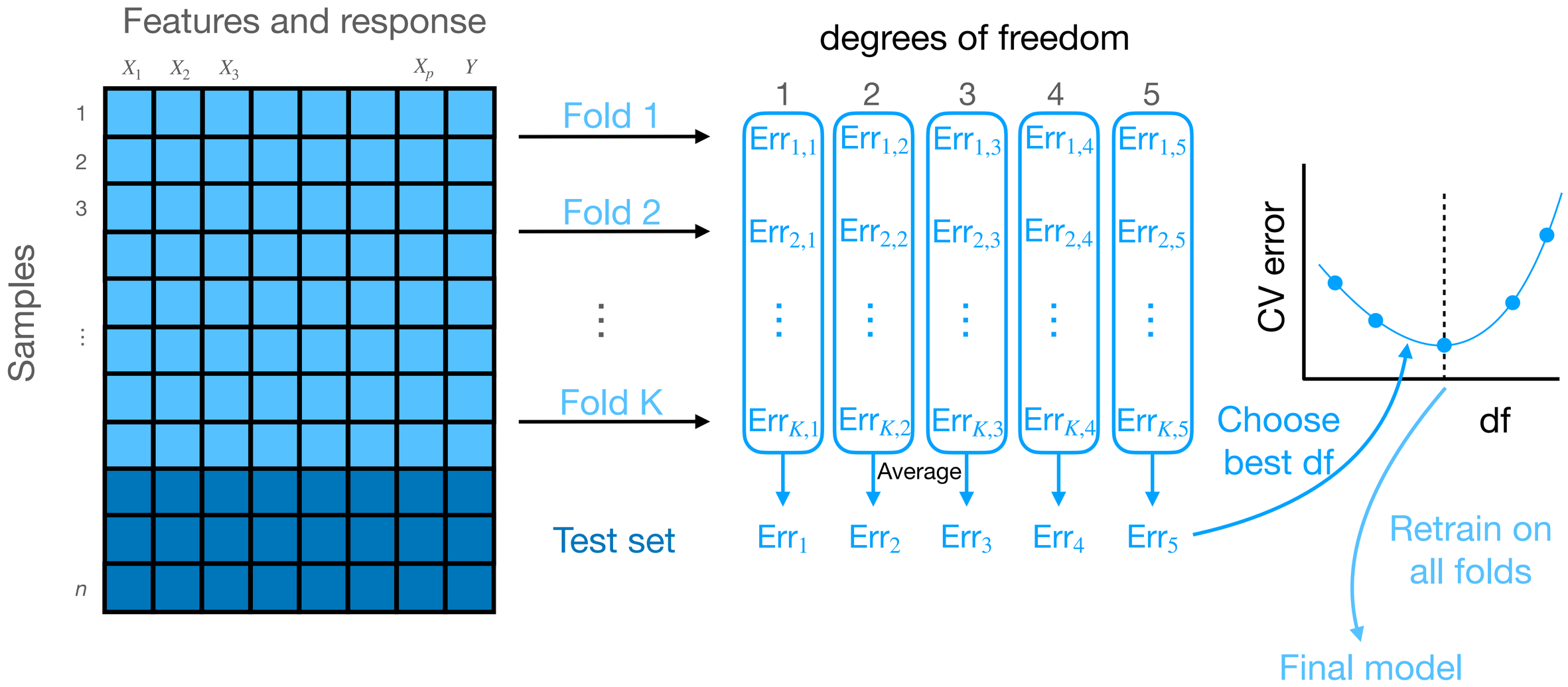
Cross-validation for model selection



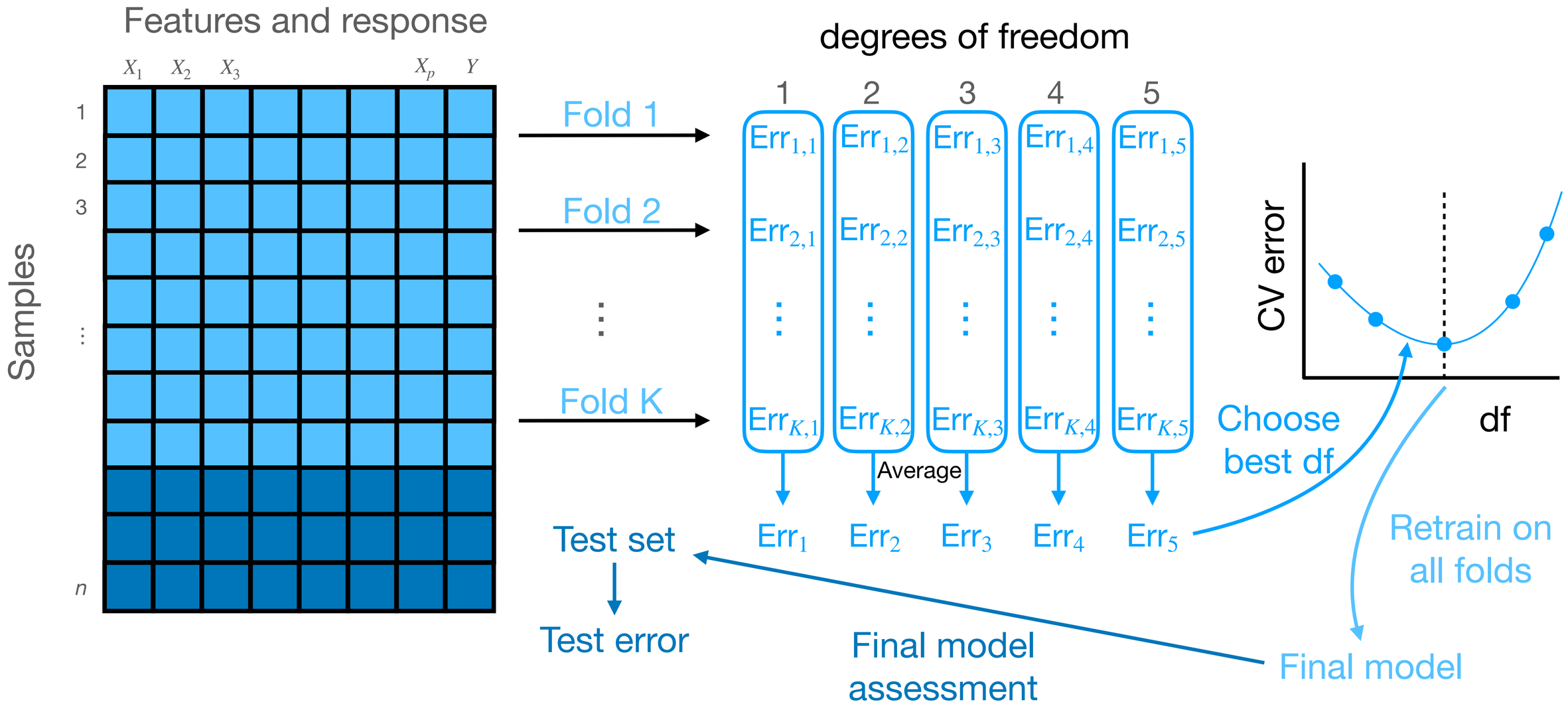
Cross-validation for model selection



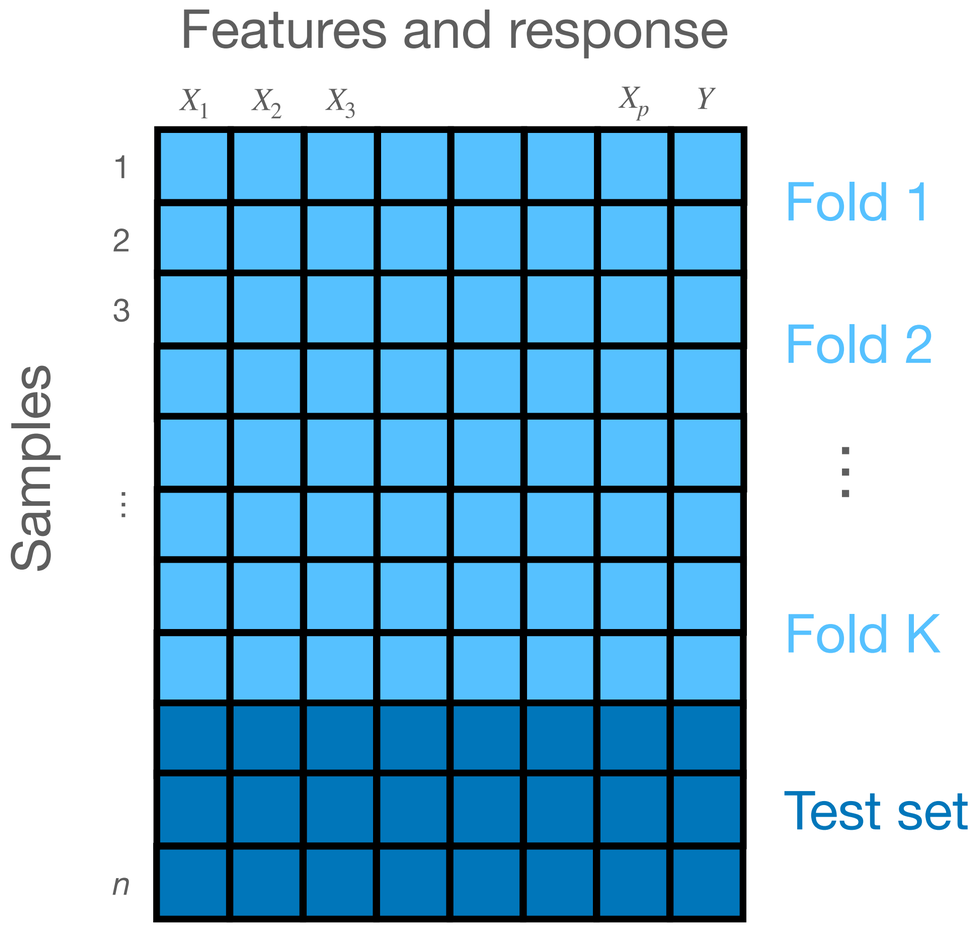
Cross-validation for model selection



Cross-validation for model selection



Cross-validation (summary)



1. Split data into K folds
2. For each fold k,
   - Fit models of varying complexity to training data, holding out fold k
   - Evaluate validation error for each model on fold k
3. Average across folds to get CV error for each model complexity

4. Choose model complexity to minimize CV error

5. Refit this model on all folds

6. Evaluate final model on the test set

Estimate test error for each model complexity

Estimate the best model complexity

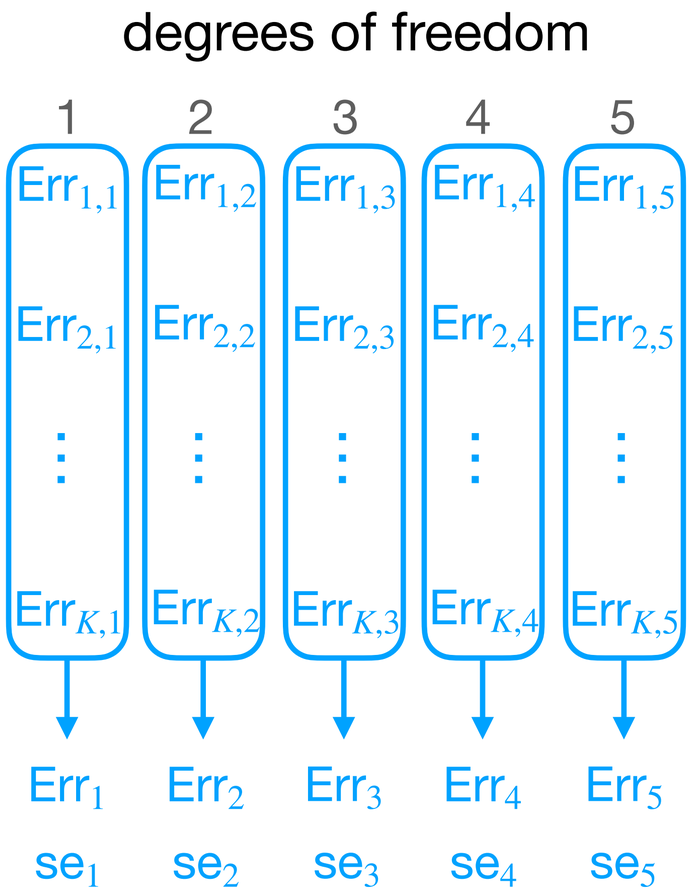
Fit final predictive model using chosen df

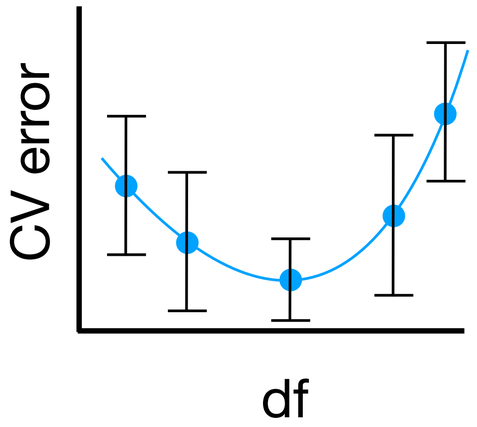
Assess test error for this final model fit

Choosing the number of folds

- More folds means more computation
- Fewer folds means the training sets used for model selection are much smaller than the actual training set
- In practice, $K = 5$ or $K = 10$ are common choices

Cross-validation standard error



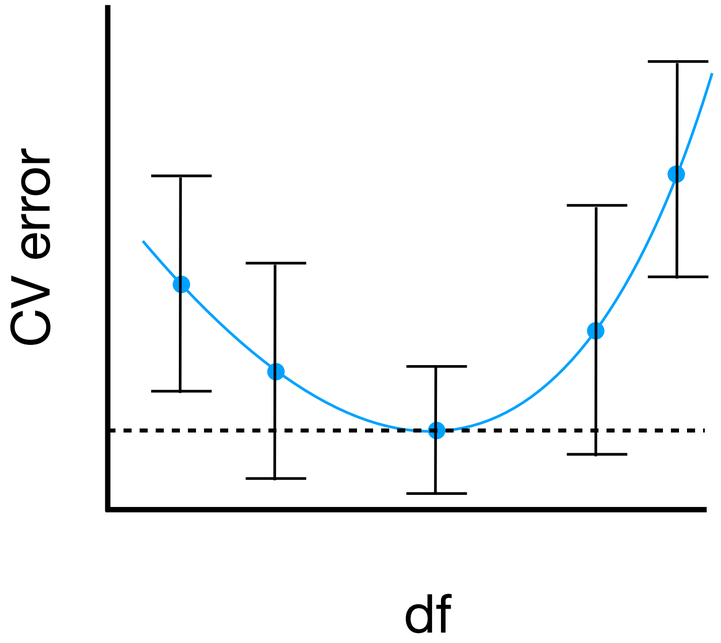


$$se_{df} = \frac{1}{\sqrt{K}} \times \text{s.d.}(Err_{1,df}, \ldots, Err_{K,df})$$

One standard error rule

Occam's razor:

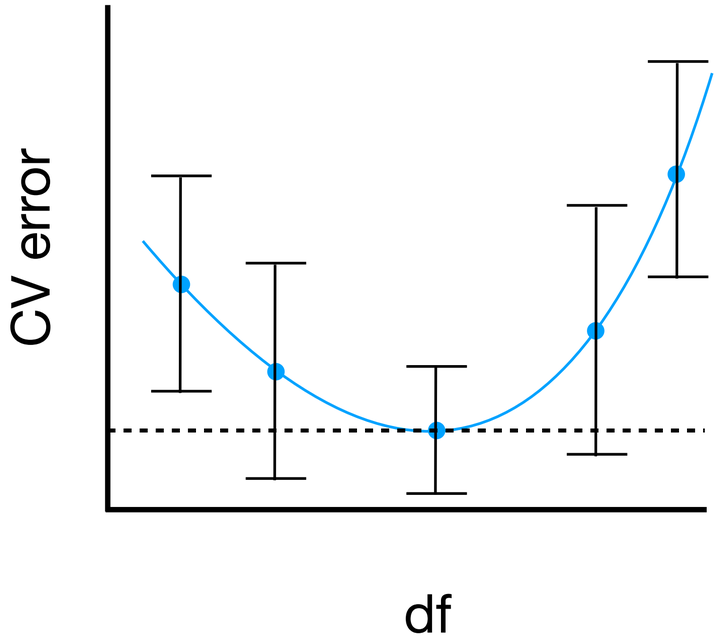
Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.



One standard error rule

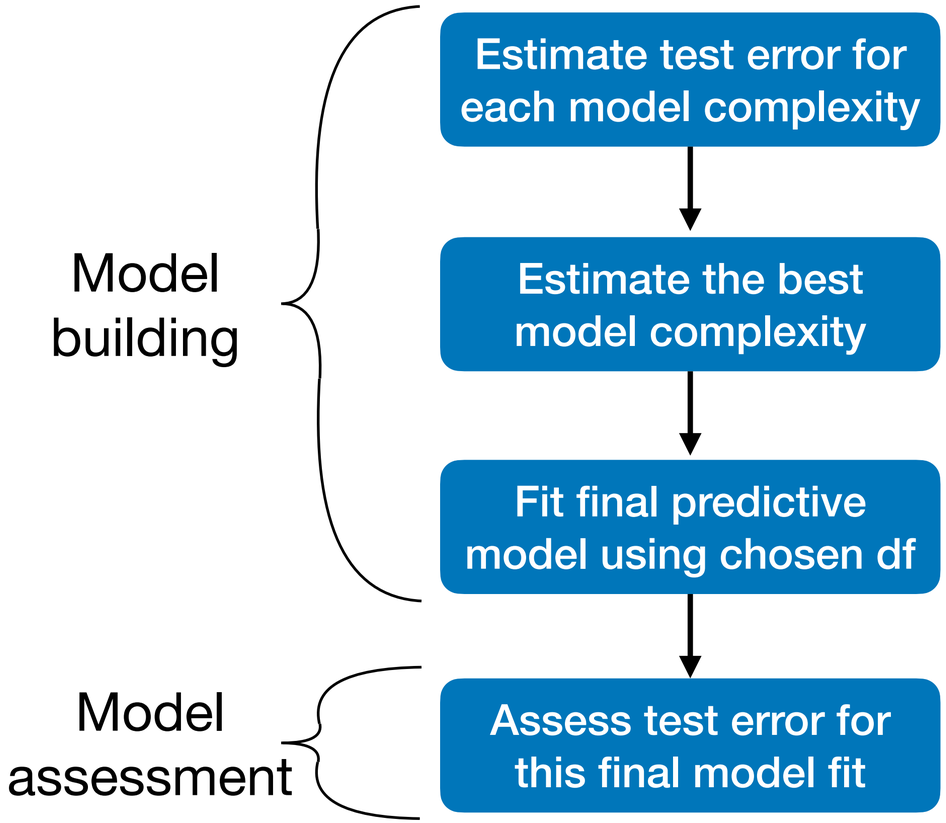
Occam's razor:

Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.

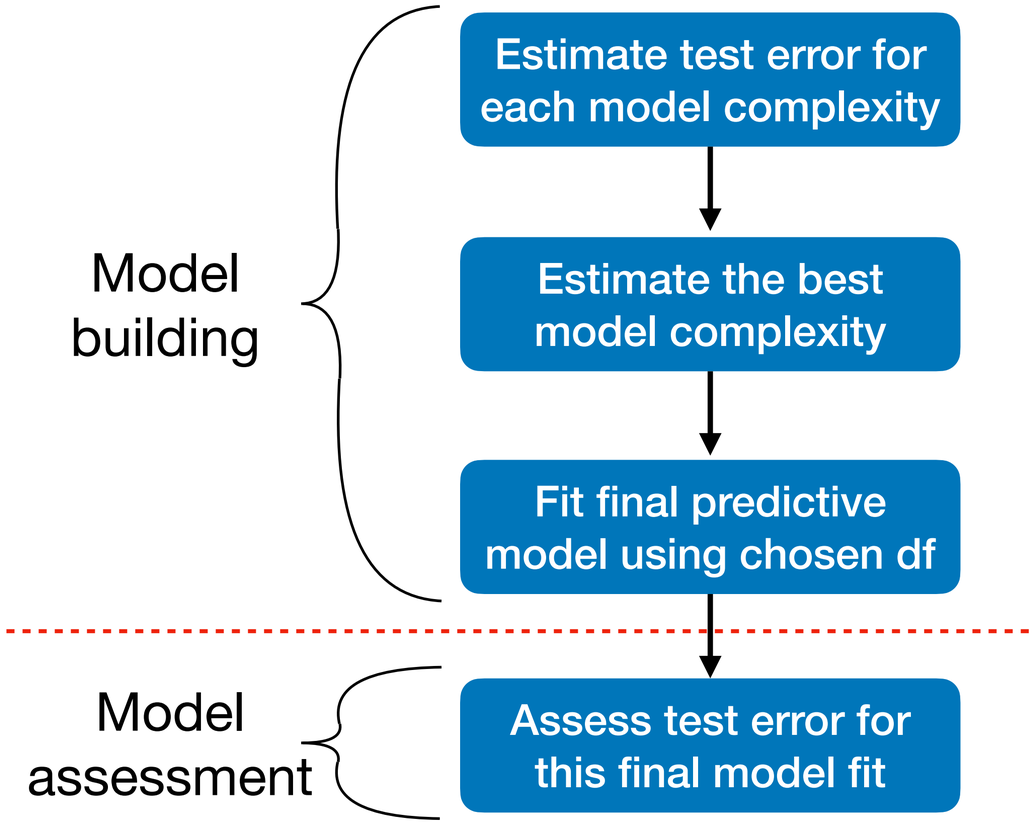


Often used instead of choosing the minimum of CV curve.

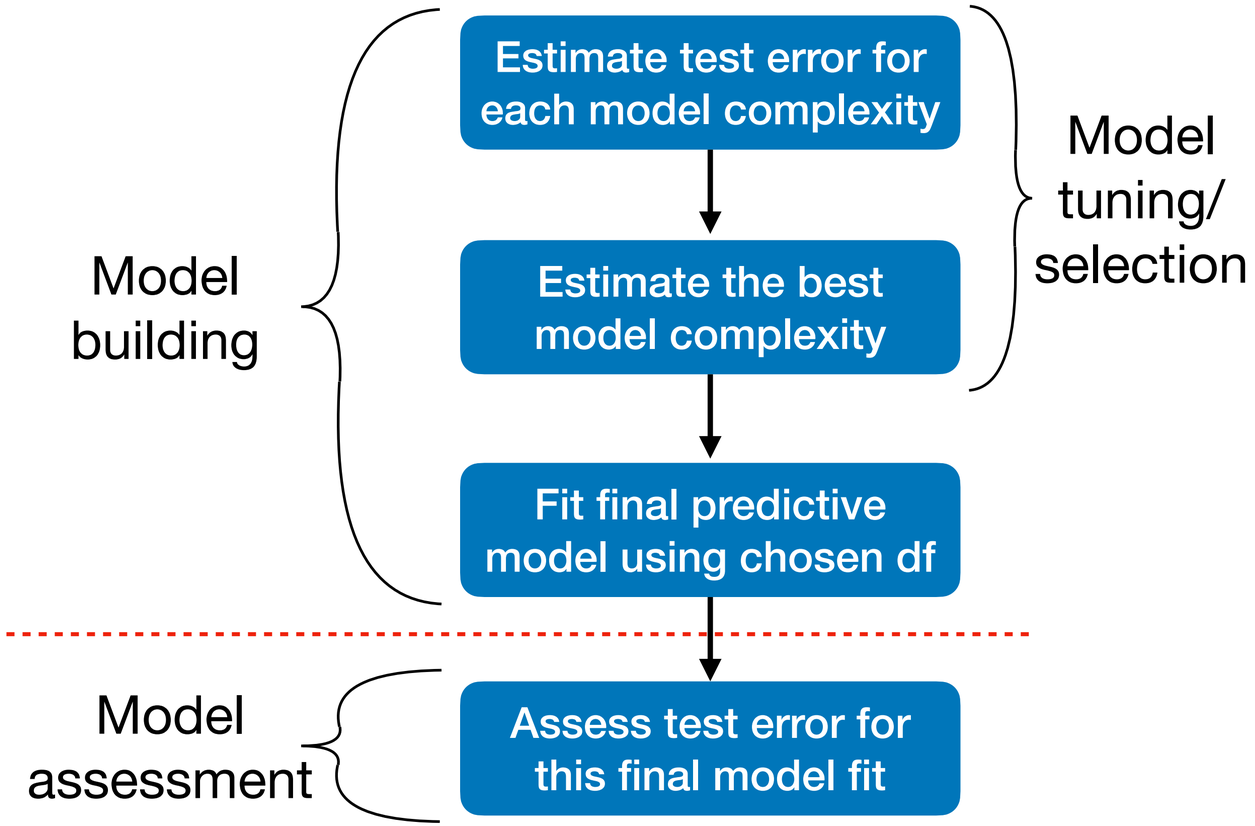
Summary


Model building
Model assessment
Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Assess test error for this final model fit

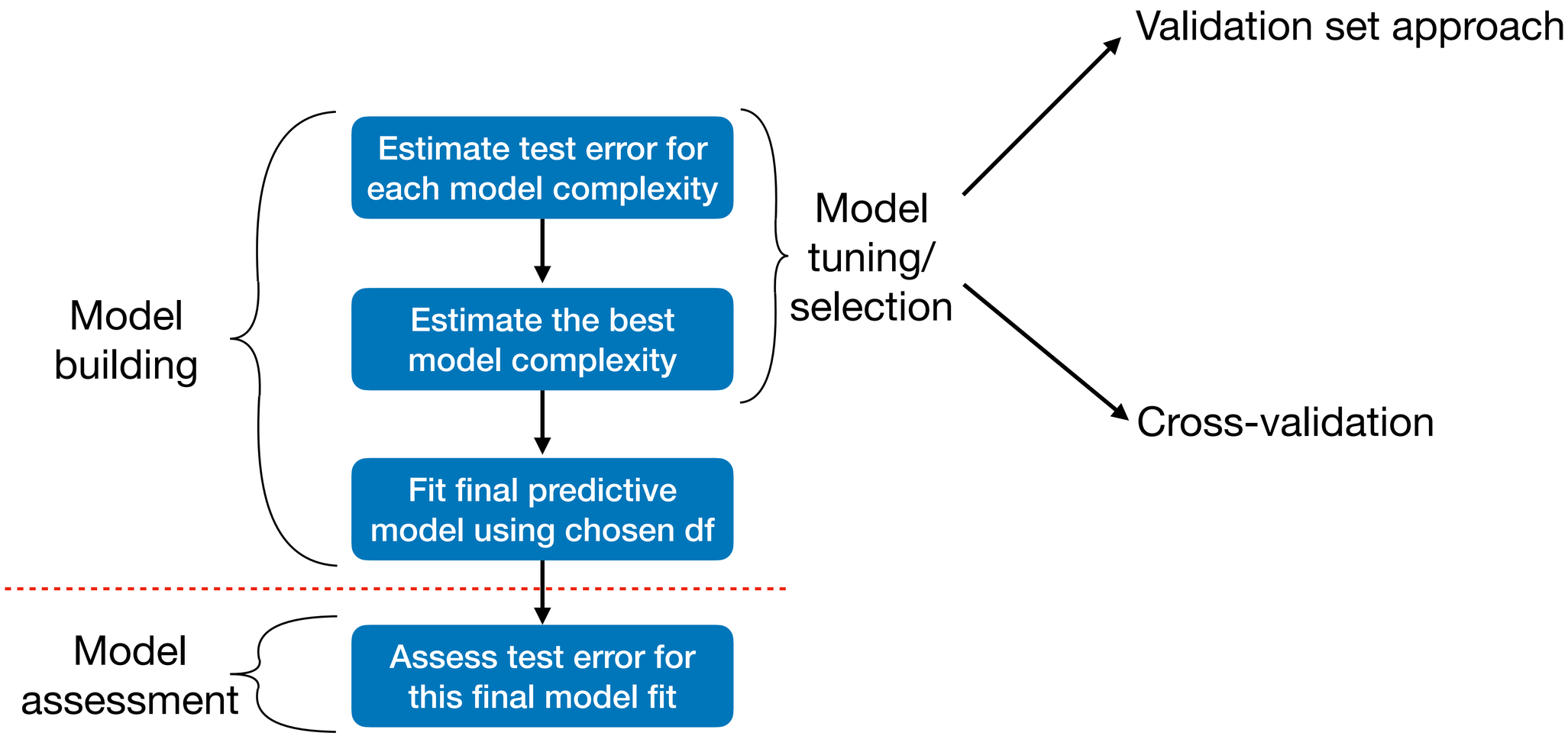
Summary


Model building
Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Model assessment
Assess test error for this final model fit

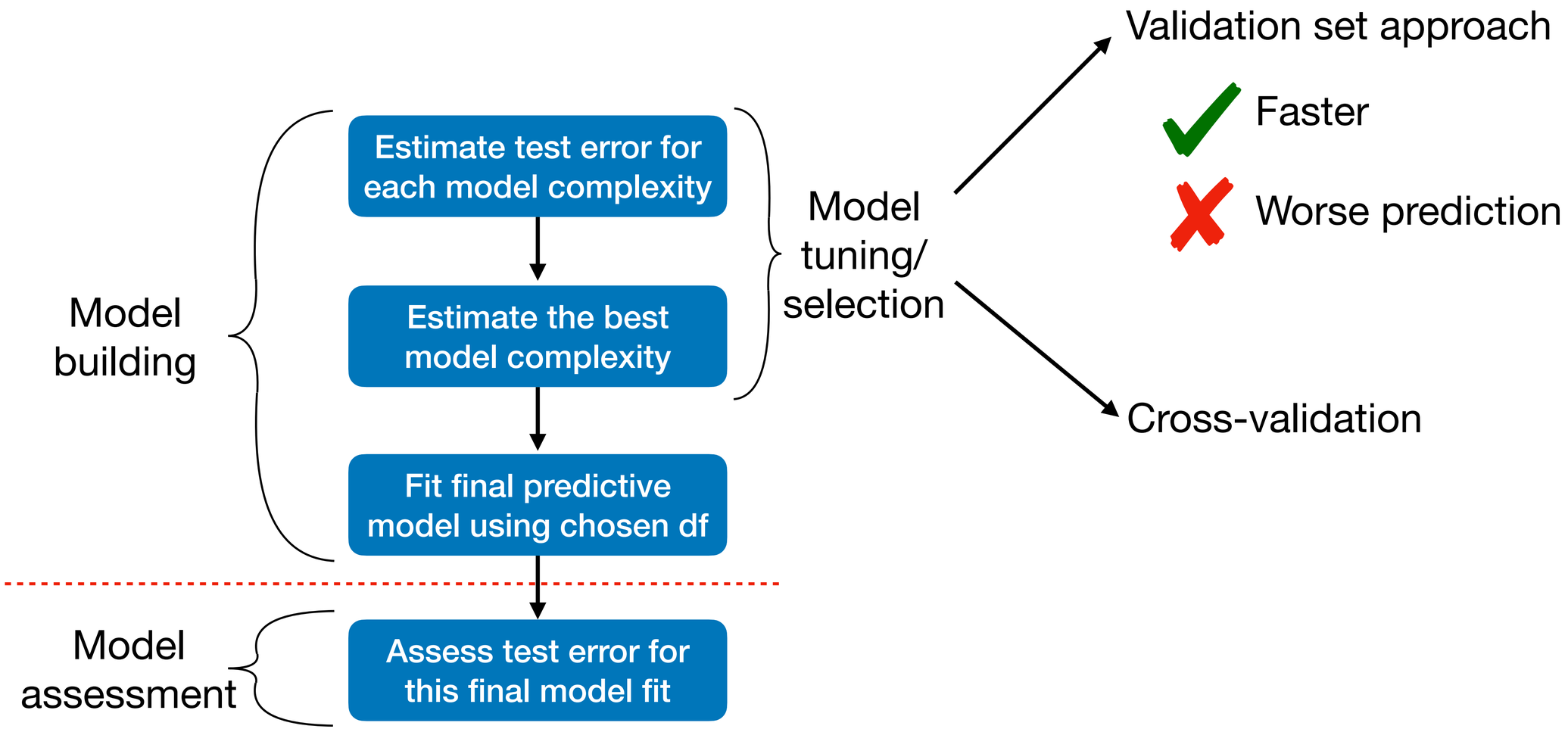
Summary


Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Assess test error for this final model fit
Model building
Model tuning/ selection
Model assessment

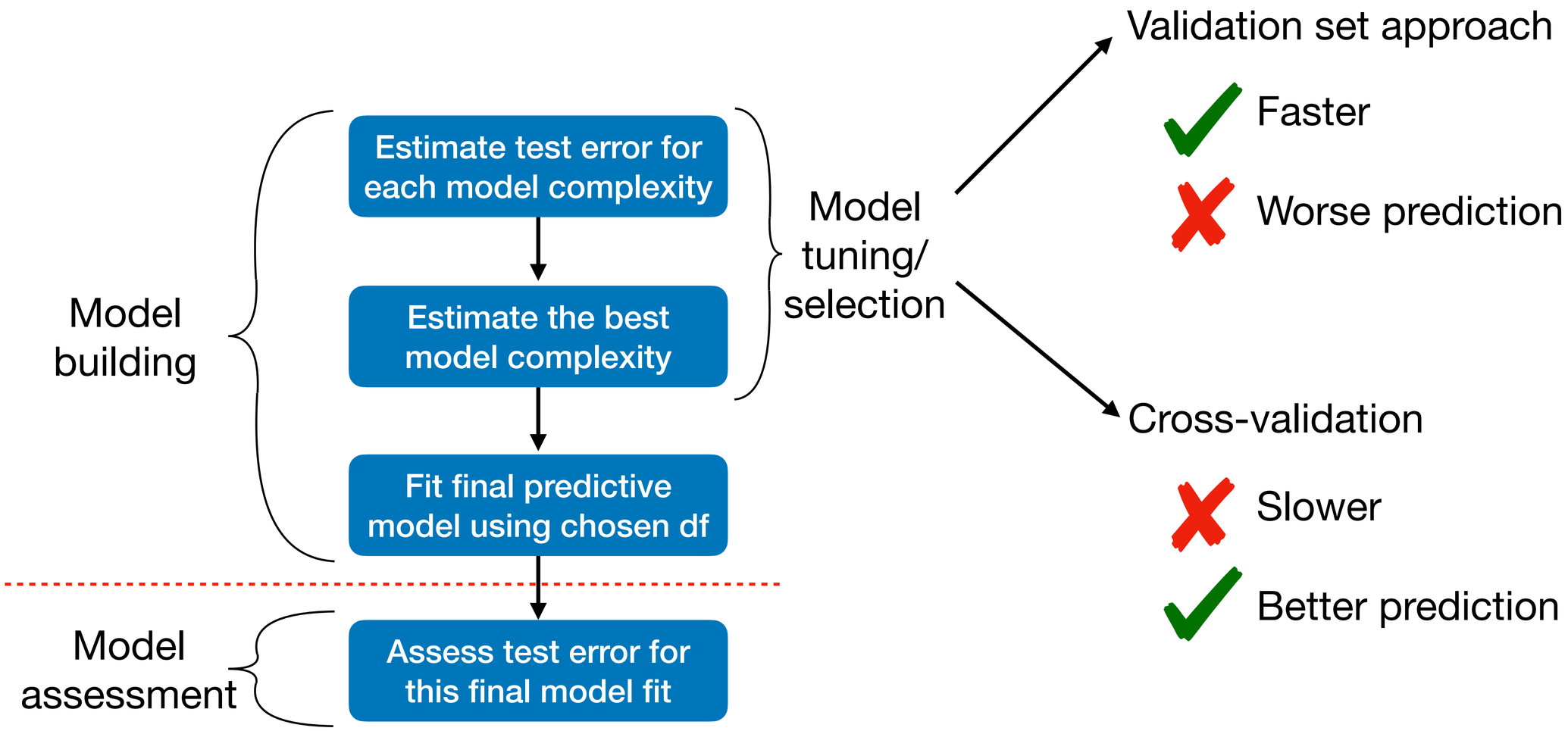
Summary


Validation set approach
Estimate test error for each model complexity
Model tuning/ selection
Model building
Estimate the best model complexity
Fit final predictive model using chosen df
Cross-validation
Model assessment
Assess test error for this final model fit

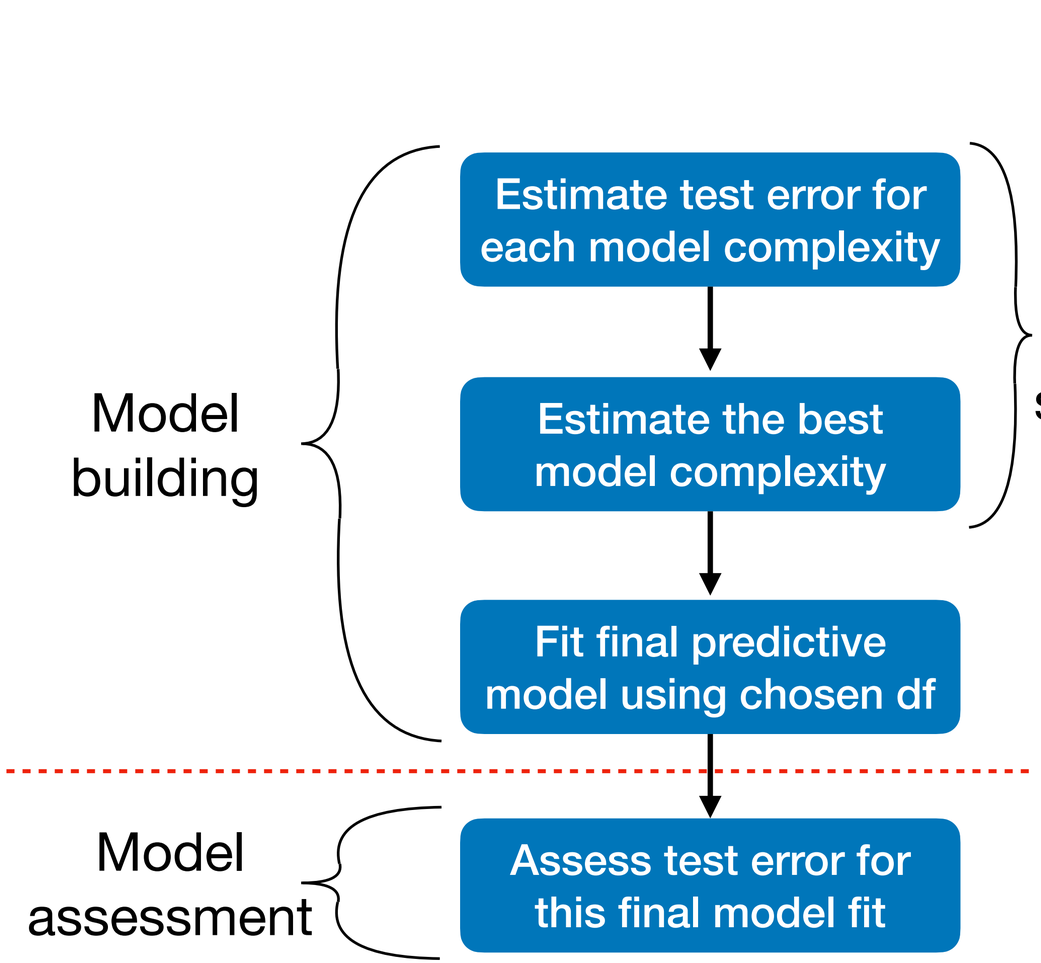
Summary


Validation set approach
Faster
Worse prediction
Model tuning/ selection
Cross-validation
Model building
Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Model assessment
Assess test error for this final model fit

Summary


Model building
Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Model assessment
Assess test error for this final model fit
Model tuning/ selection
Validation set approach
Faster
Worse prediction
Cross-validation
Slower
Better prediction

Summary



Validation set approach

✓ Faster

✗ Worse prediction

Cross-validation

✗ Slower

✓ Better prediction

Both approaches to model tuning used in practice.