Cross-validation

STAT 4710

September 21, 2023

Where we are

Unit 1: R for data mining

Unit 2: Prediction fundamentals

Unit 3: Regression-based methods

Unit 4: Tree-based methods

Unit 5: Deep learning

Lecture 1: Model complexity

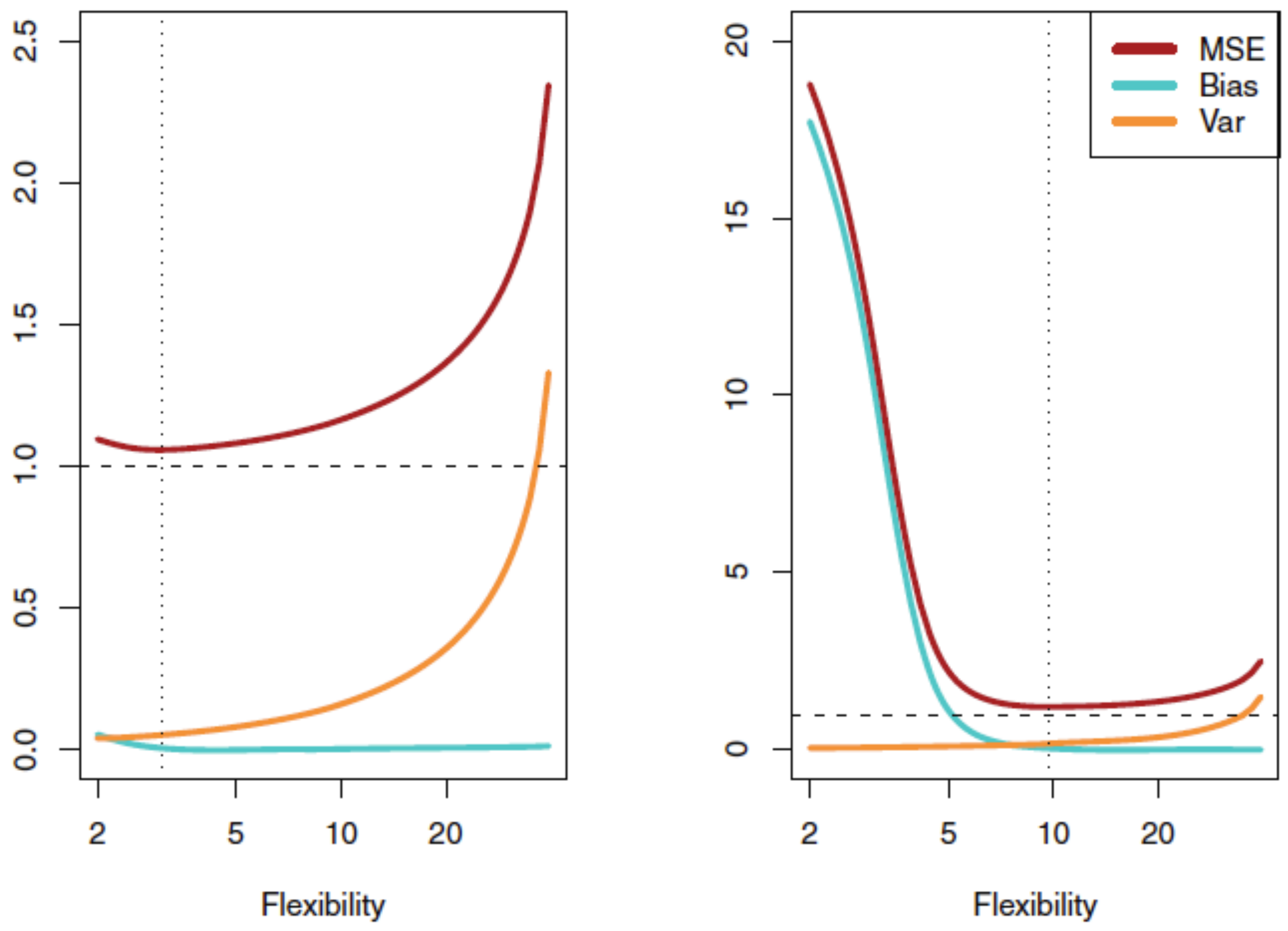
Lecture 2: Bias-variance trade-off

Lecture 3: Cross-validation

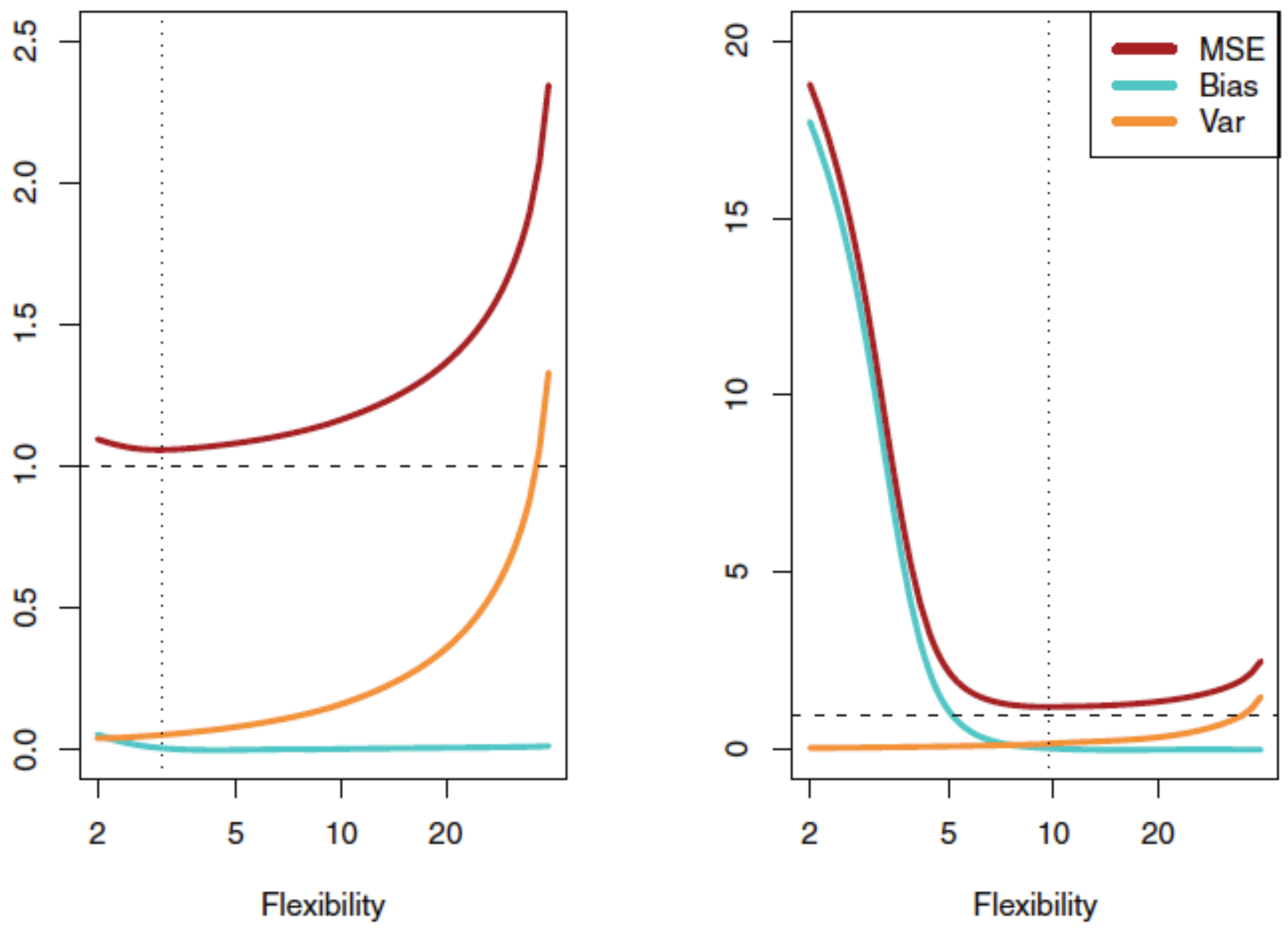
Lecture 4: Classification

Lecture 5: Unit review and quiz in class

Navigating the bias-variance tradeoff in practice

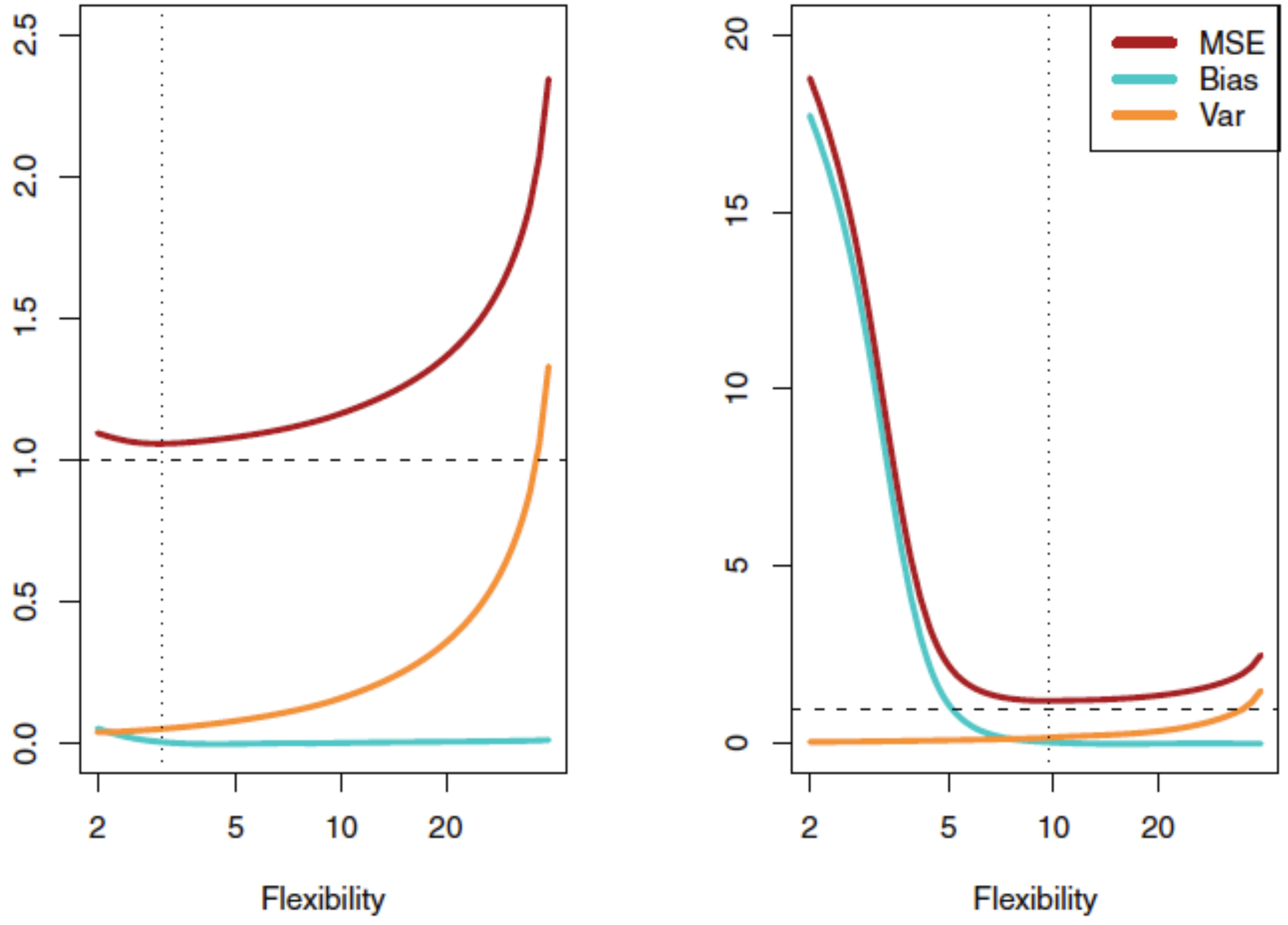


Navigating the bias-variance tradeoff in practice



Estimate test error for each model complexity

Navigating the bias-variance tradeoff in practice

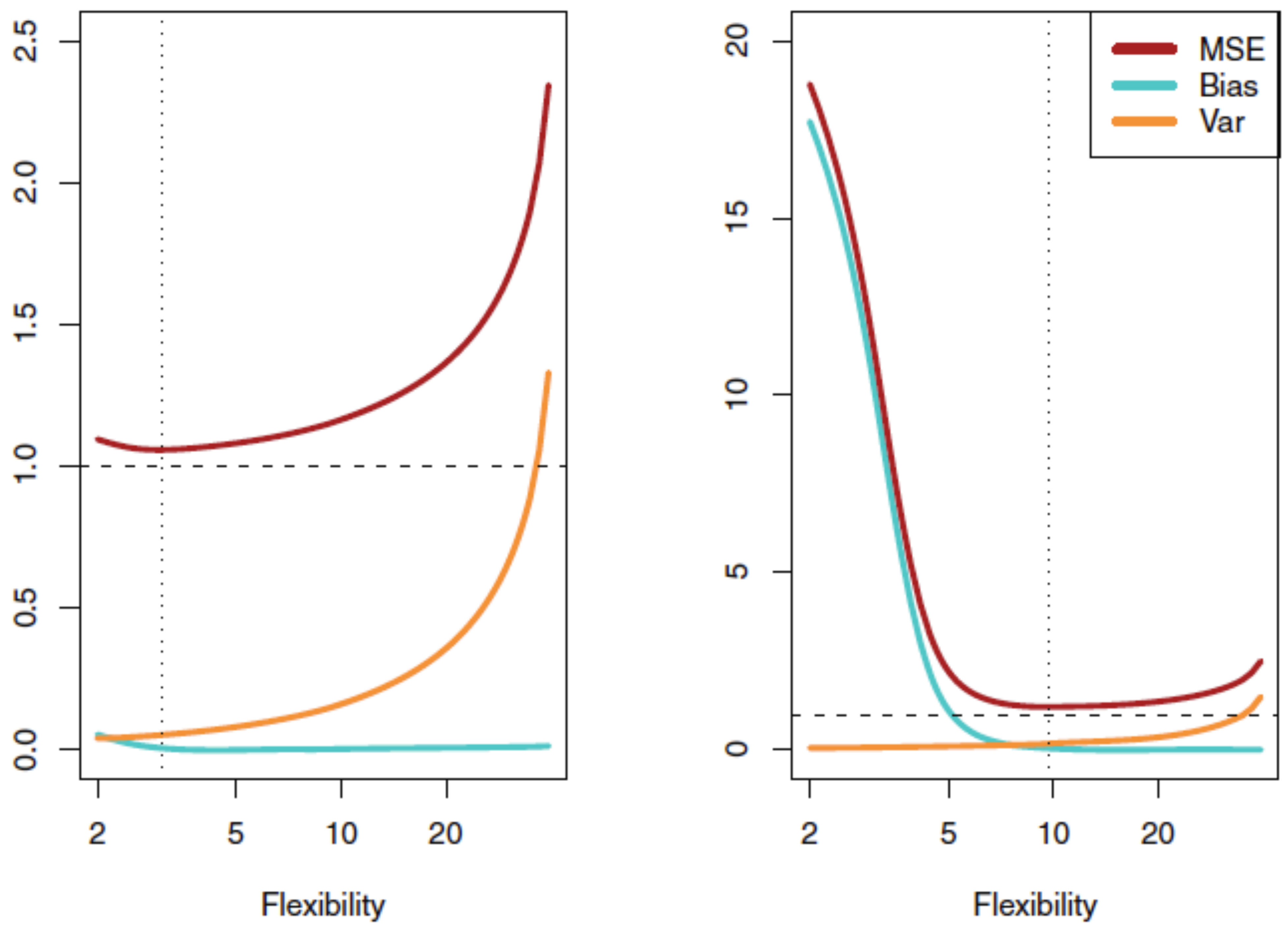


Estimate test error for each model complexity

↓

Estimate the best model complexity

Navigating the bias-variance tradeoff in practice

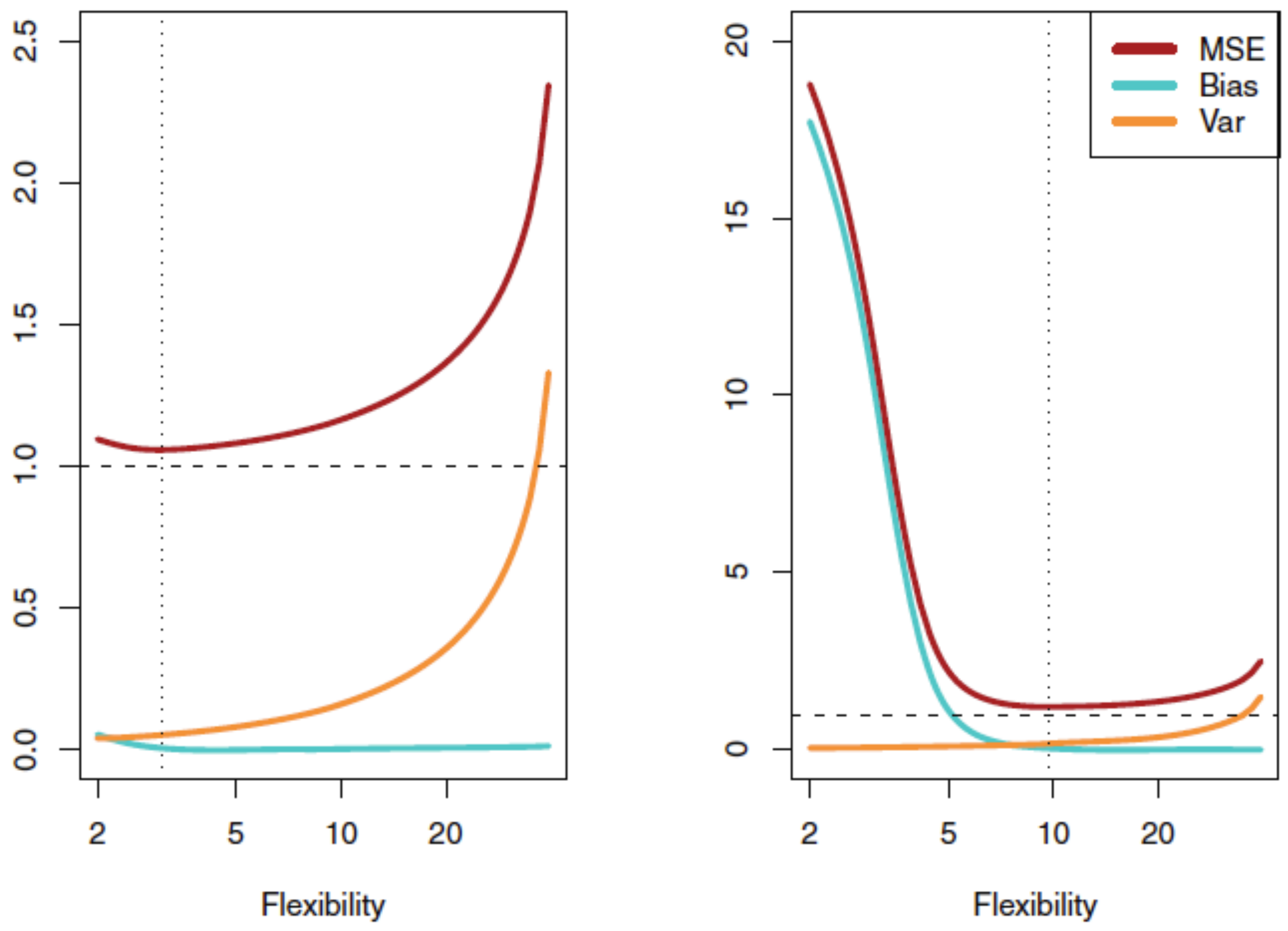


Estimate test error for each model complexity

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

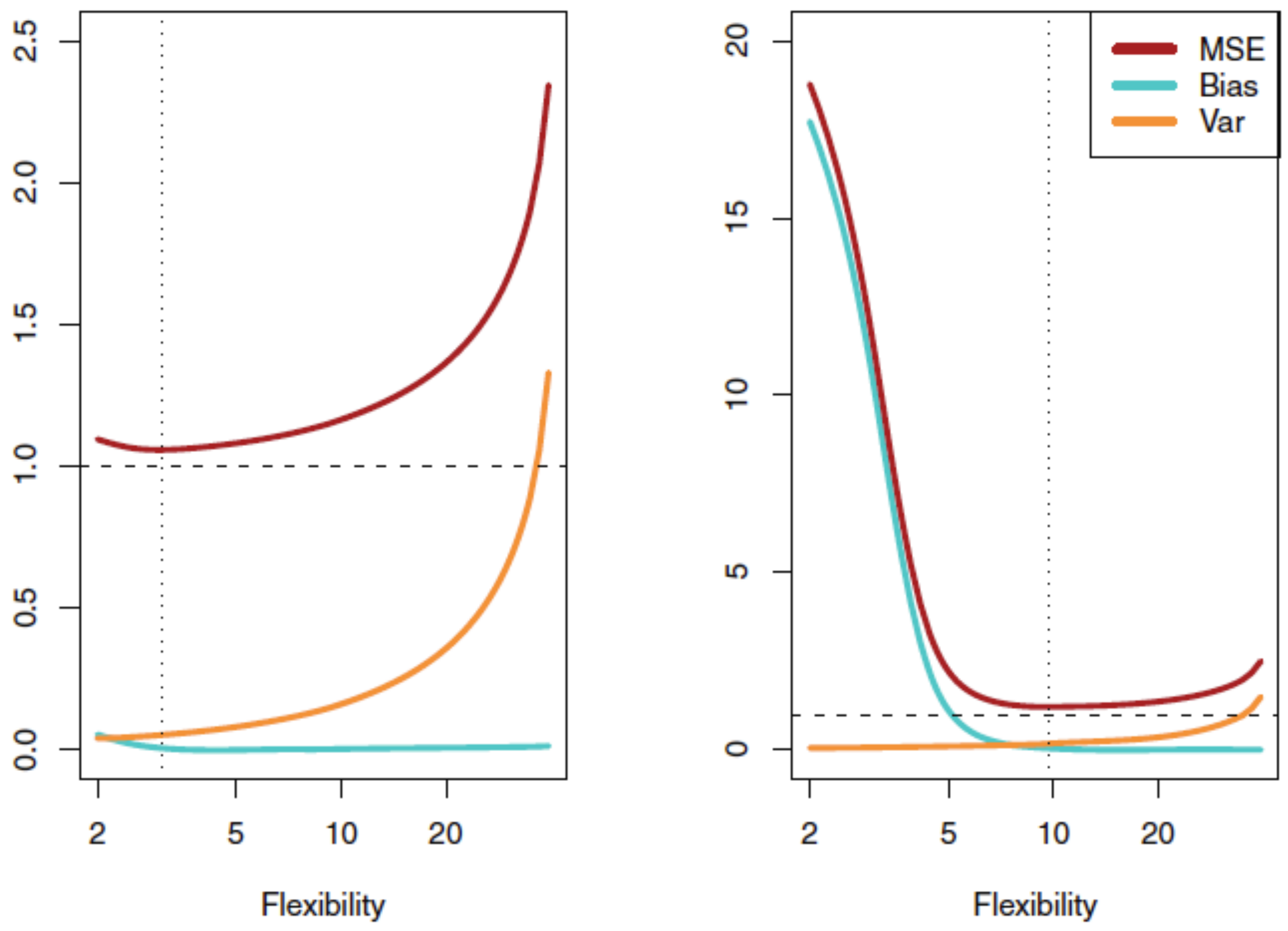
↓

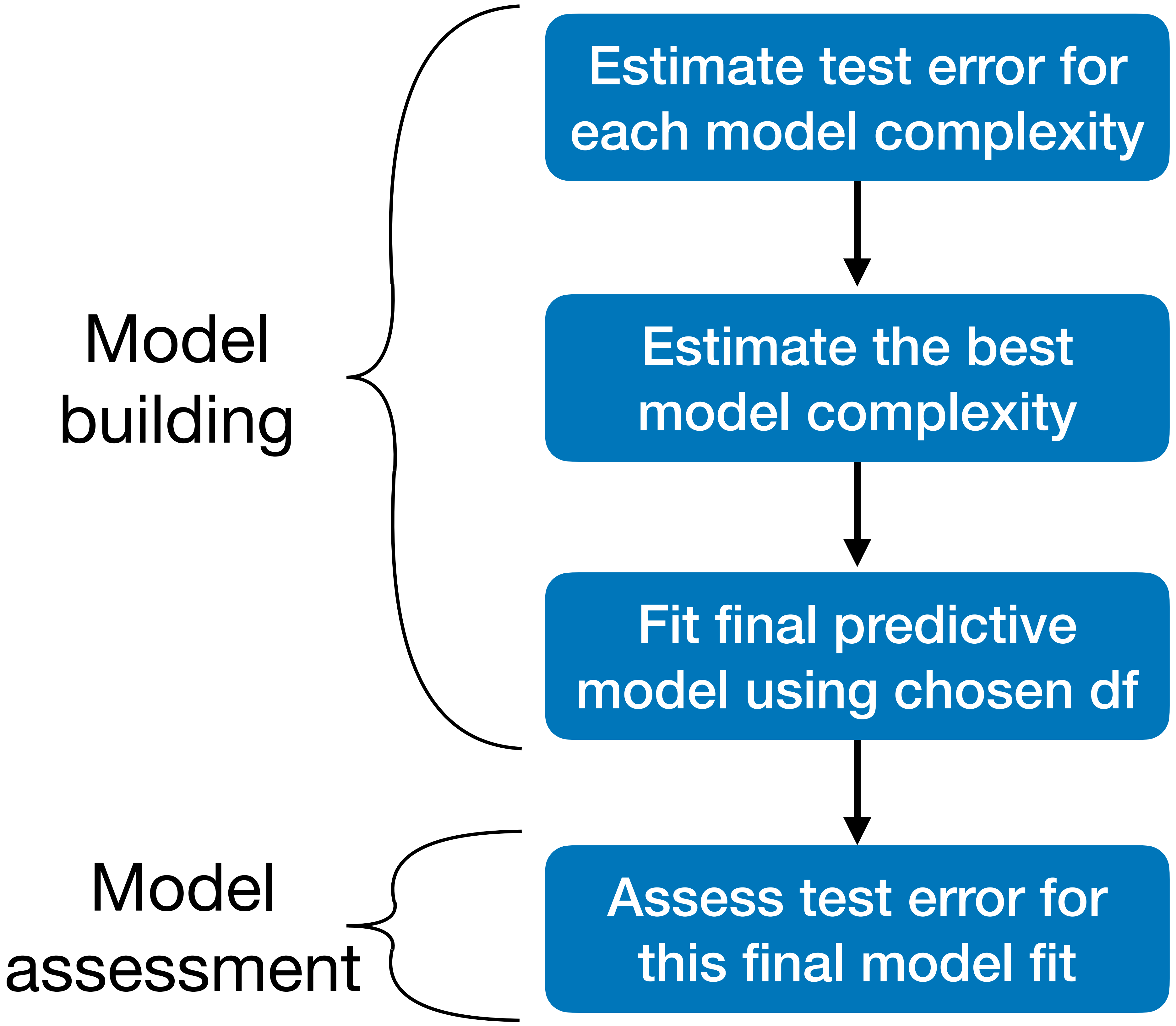
Fit final predictive model using chosen df

↓

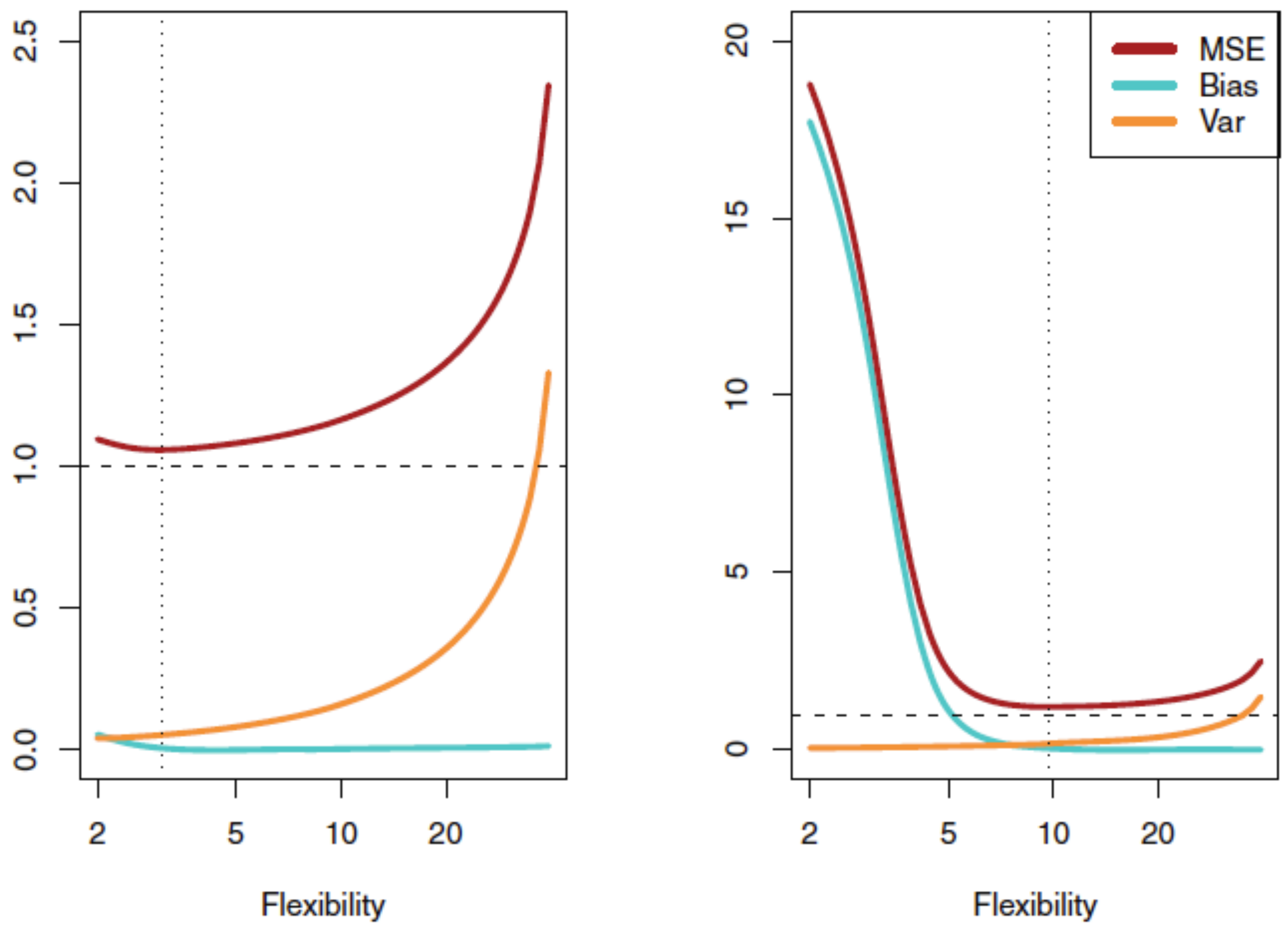
Assess test error for this final model fit

Navigating the bias-variance tradeoff in practice





Navigating the bias-variance tradeoff in practice



Model building

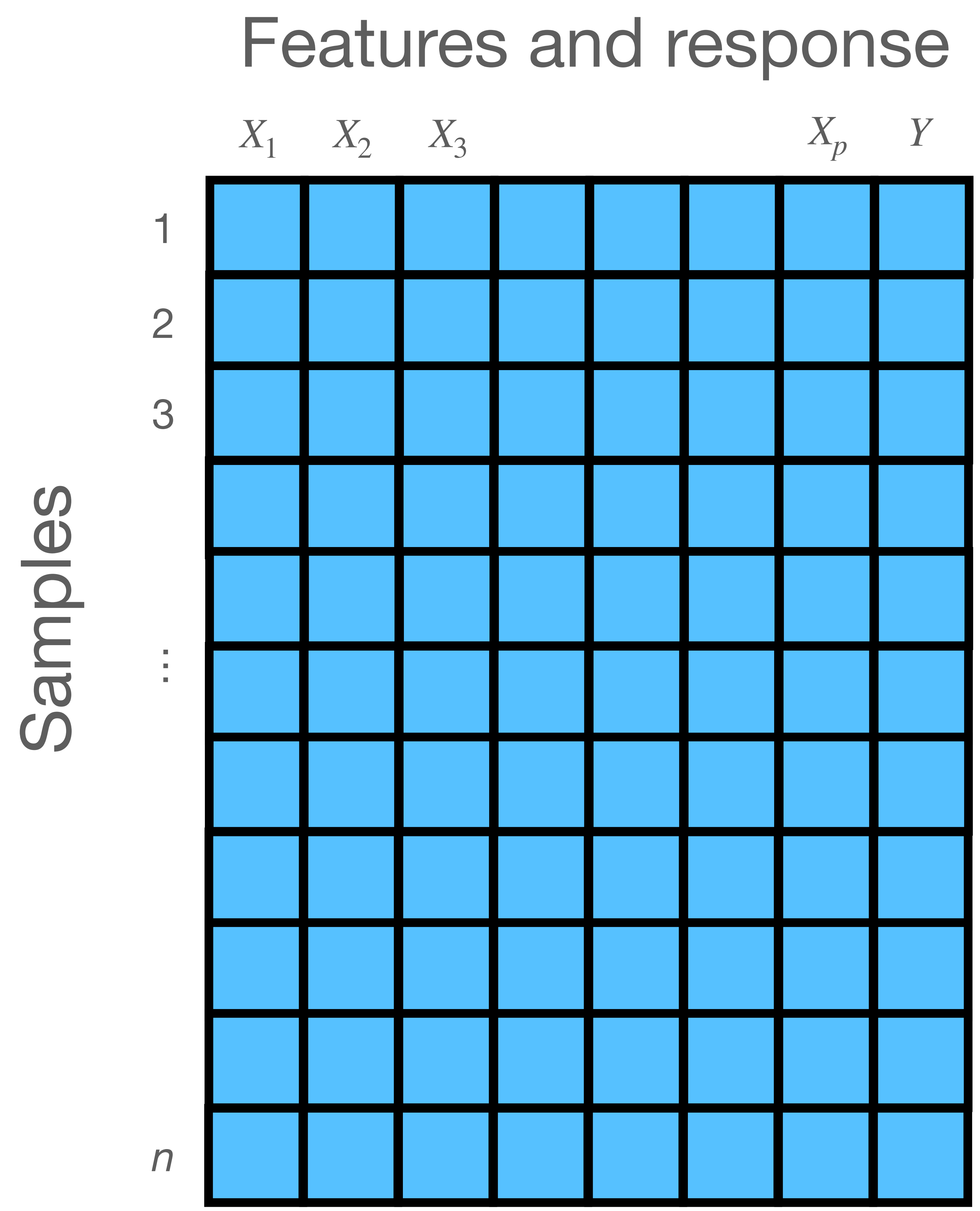
- **Model tuning/selection**
  - Estimate test error for each model complexity
  - ↓
  - Estimate the best model complexity
- ↓
- Fit final predictive model using chosen df

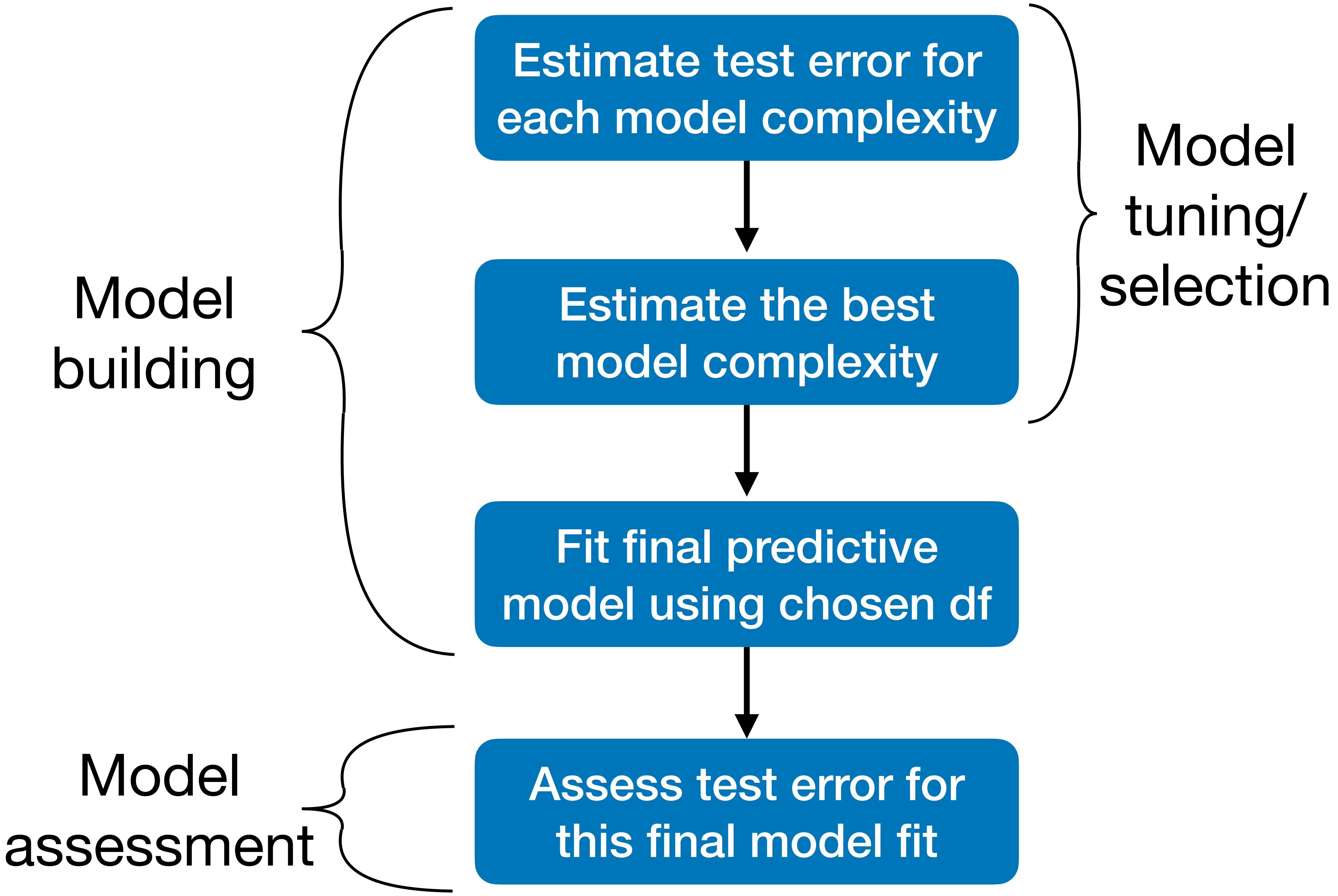
↓

Model assessment

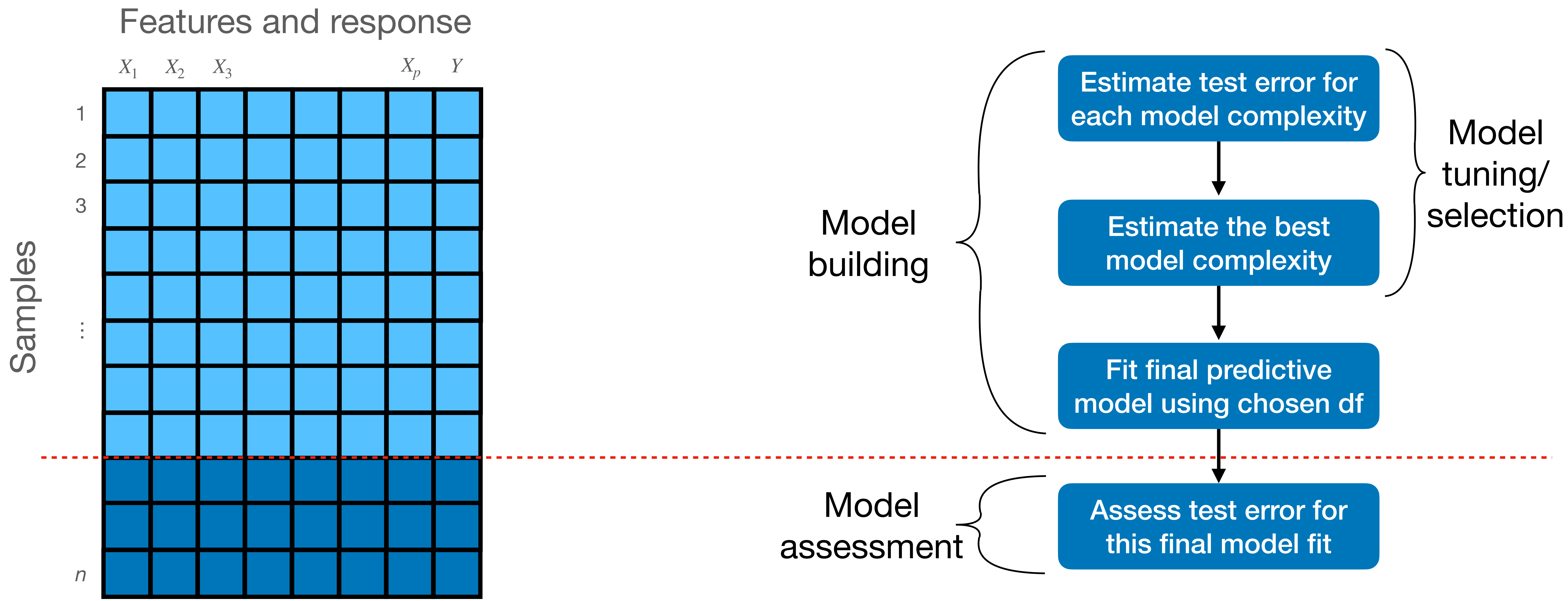
- Assess test error for this final model fit

Separating data for model building and assessment

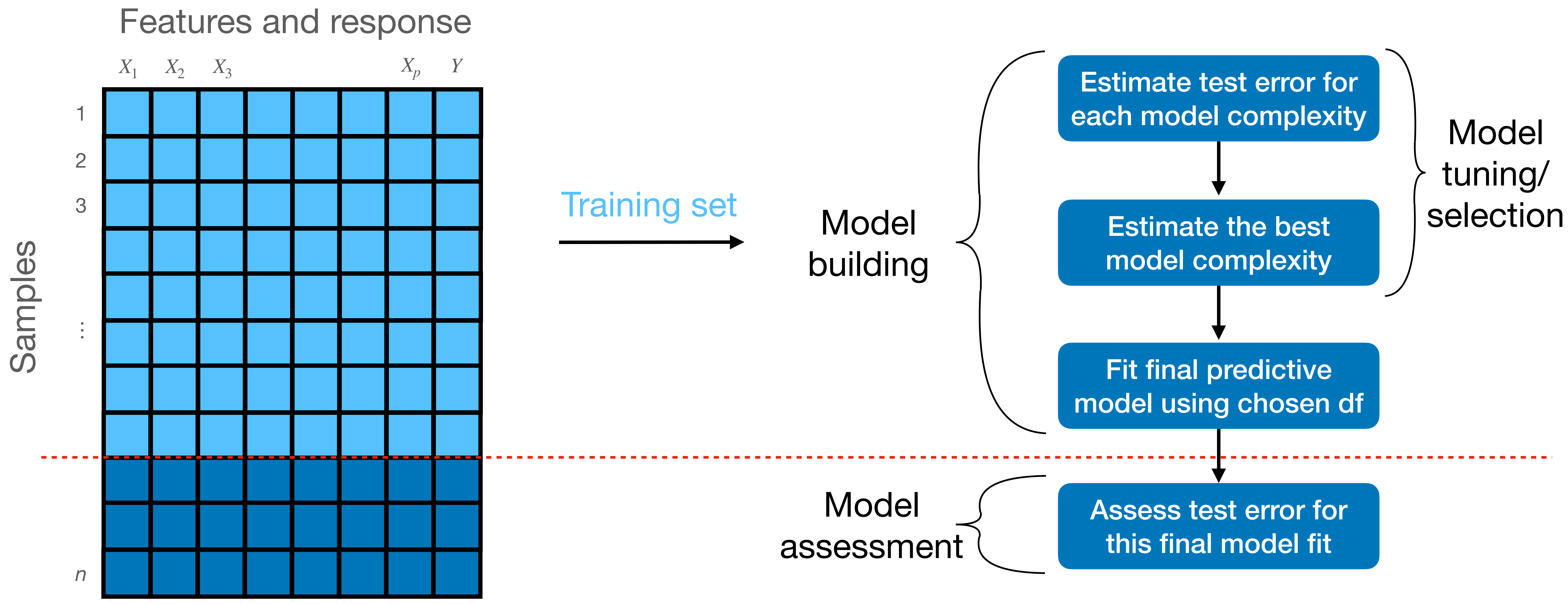




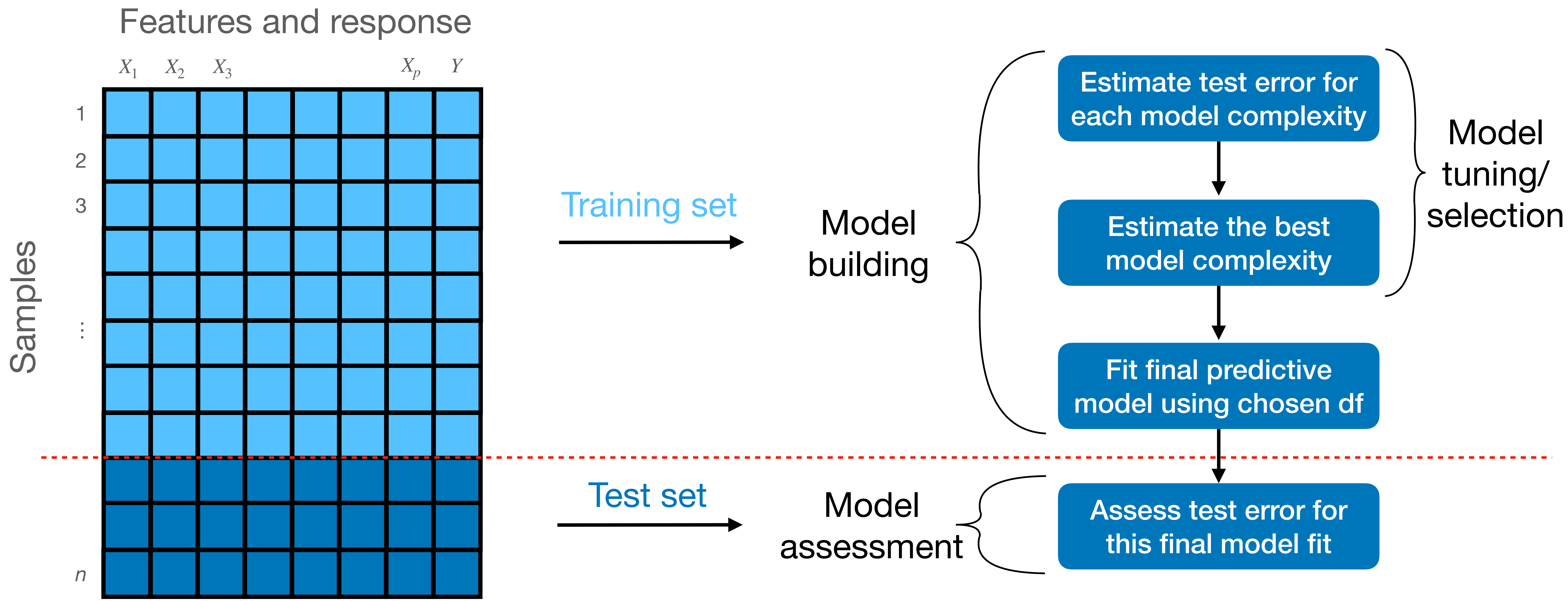
Separating data for model building and assessment



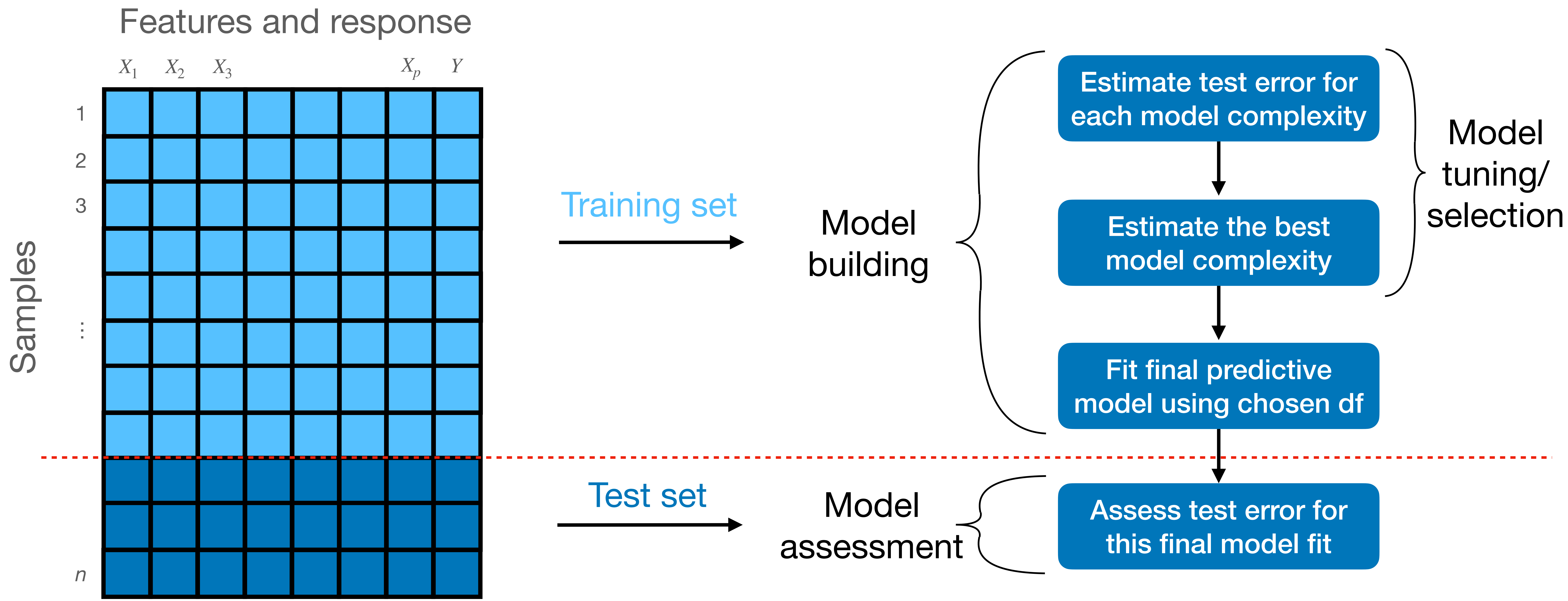
Separating data for model building and assessment



Separating data for model building and assessment



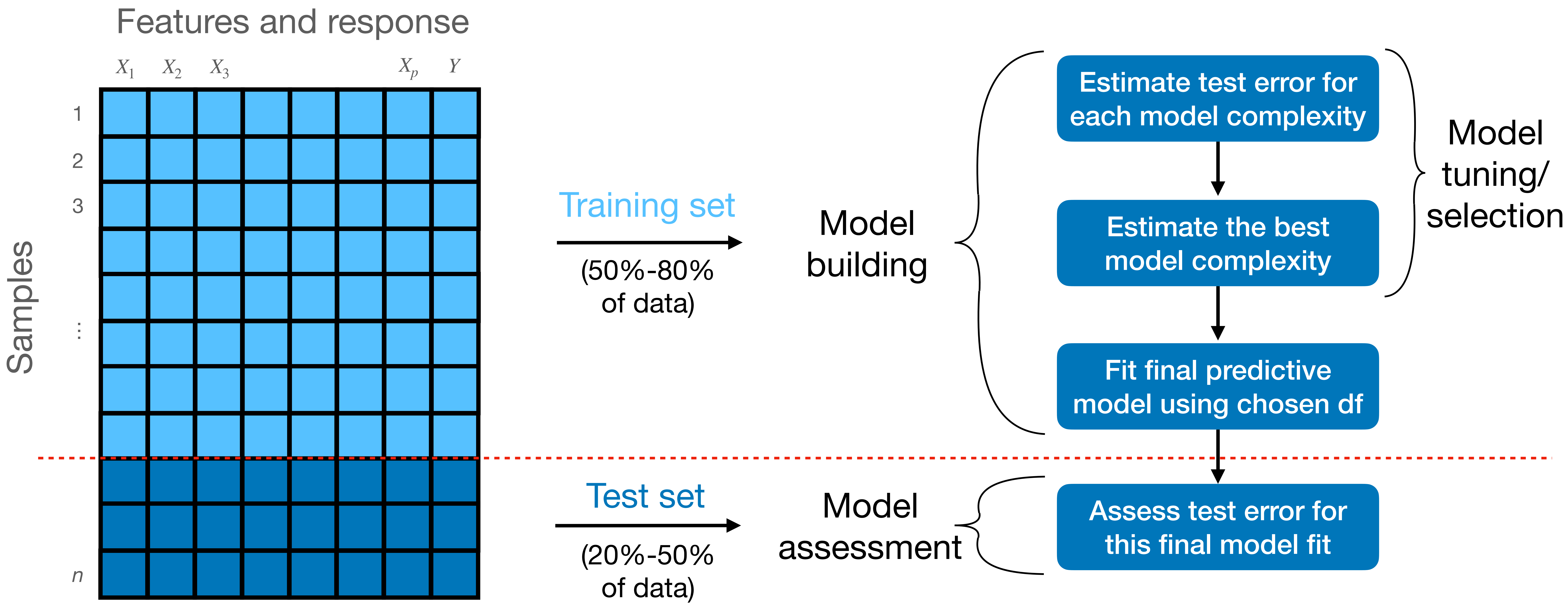
Separating data for model building and assessment



More samples for training → better fitted model;

More samples for testing → better estimate of test error.

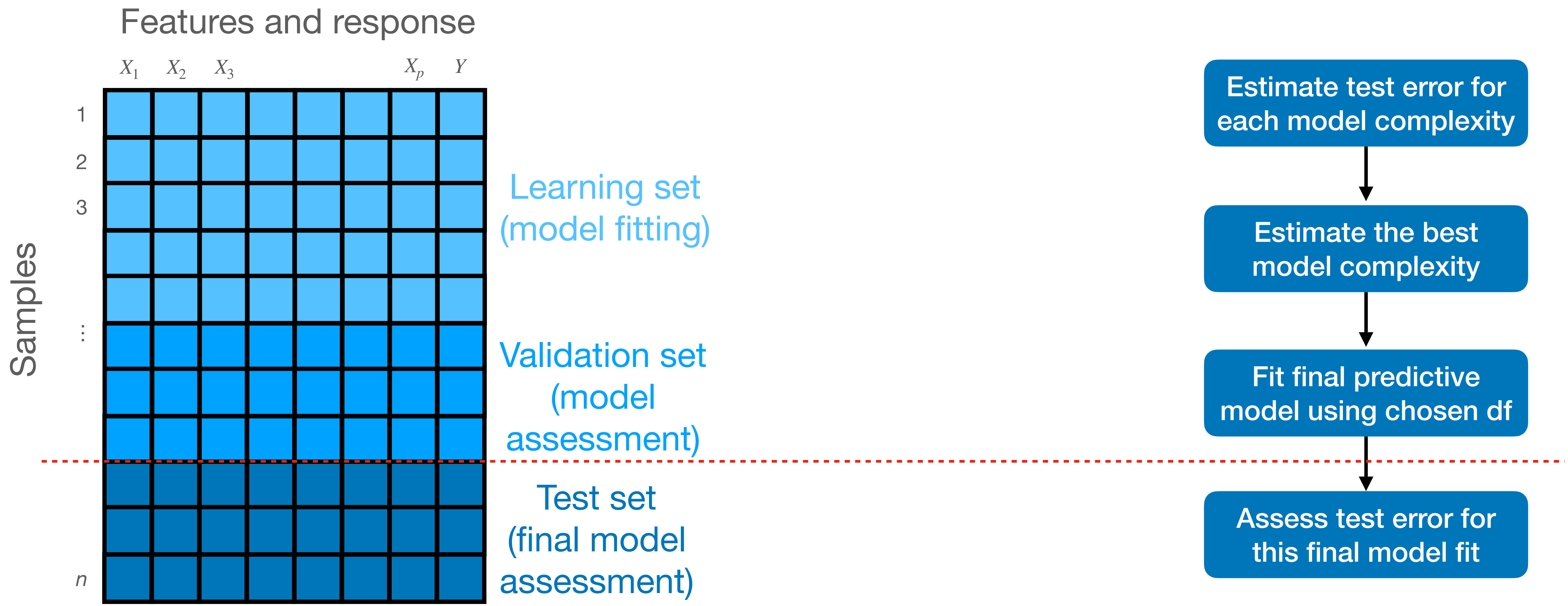
Separating data for model building and assessment



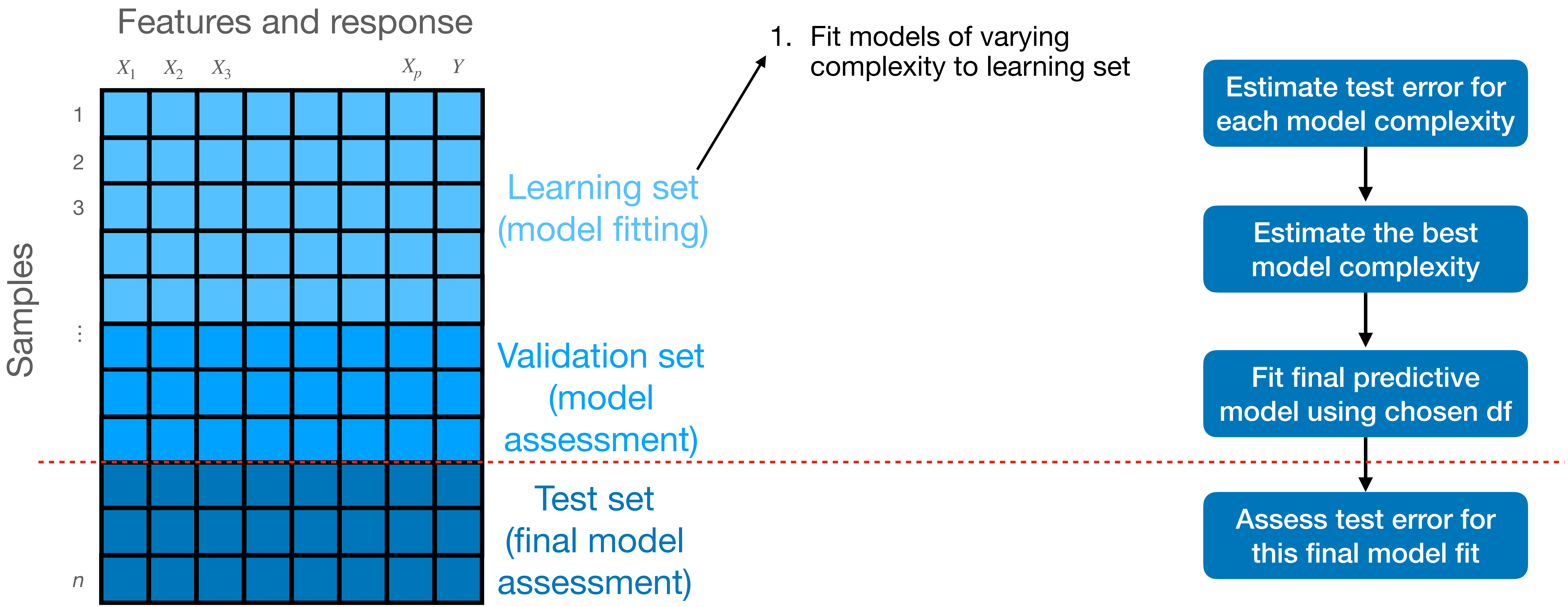
More samples for training → better fitted model;

More samples for testing → better estimate of test error.

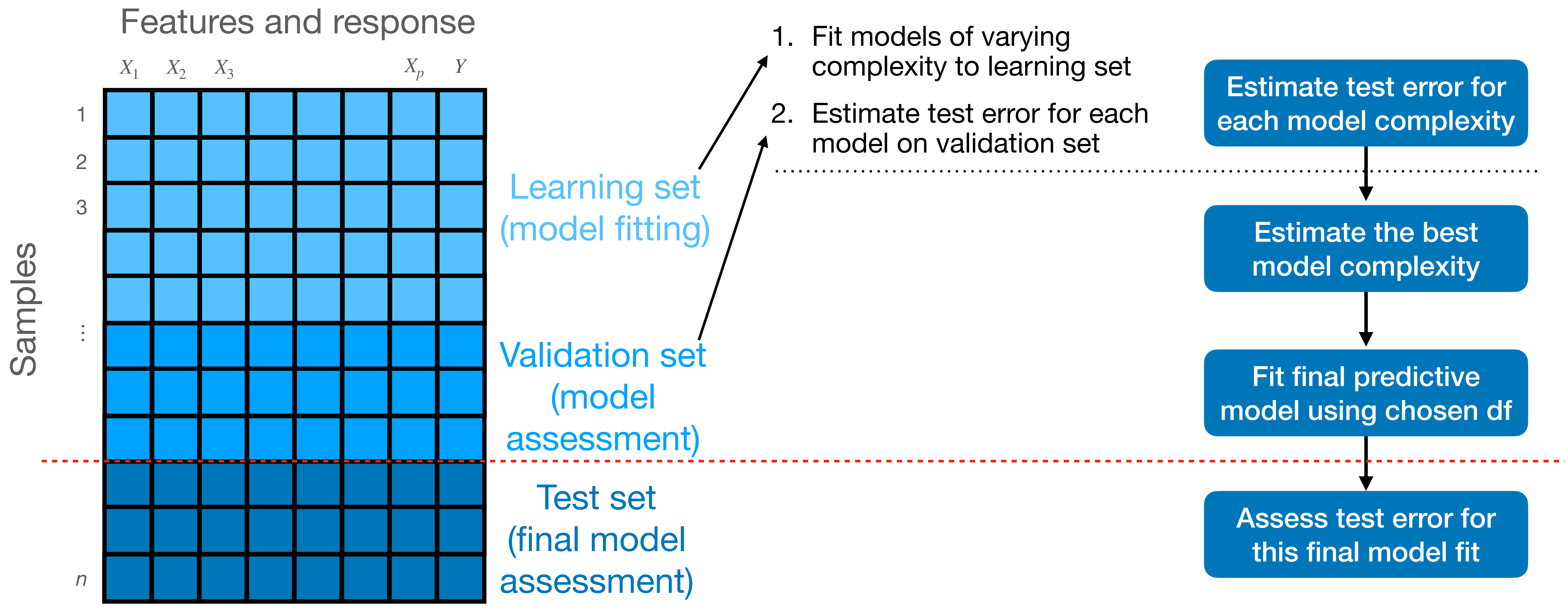
Validation set approach for model selection



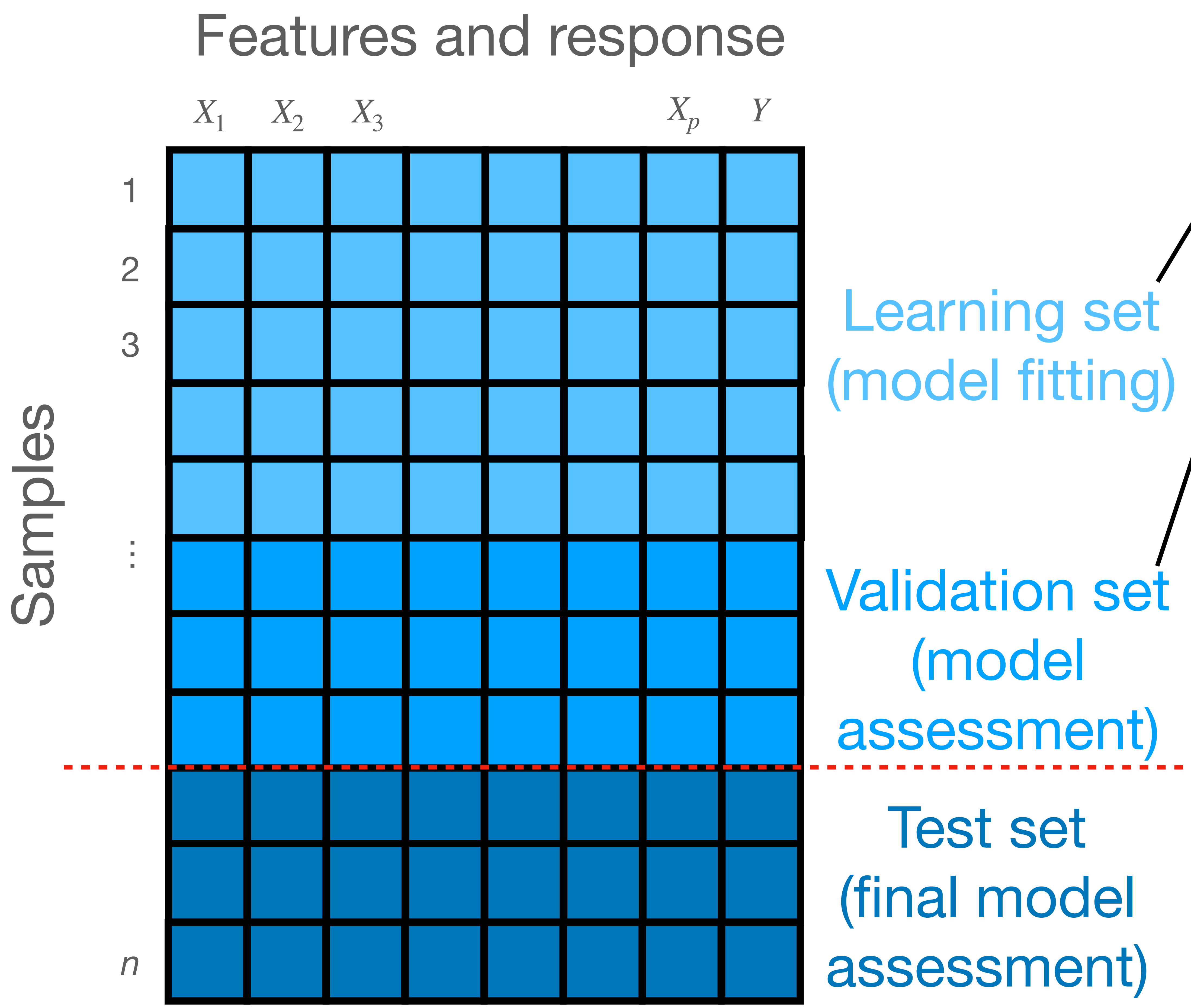
Validation set approach for model selection



Validation set approach for model selection



Validation set approach for model selection



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

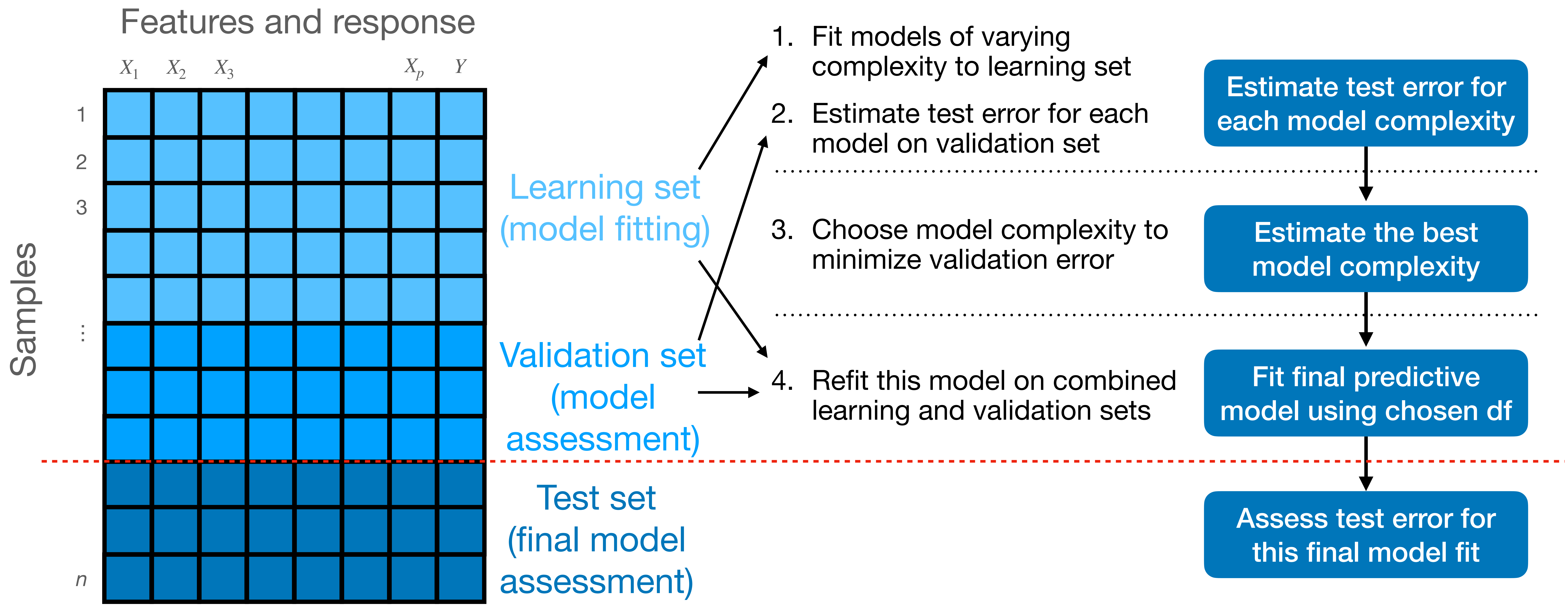
Estimate test error for each model complexity

Estimate the best model complexity

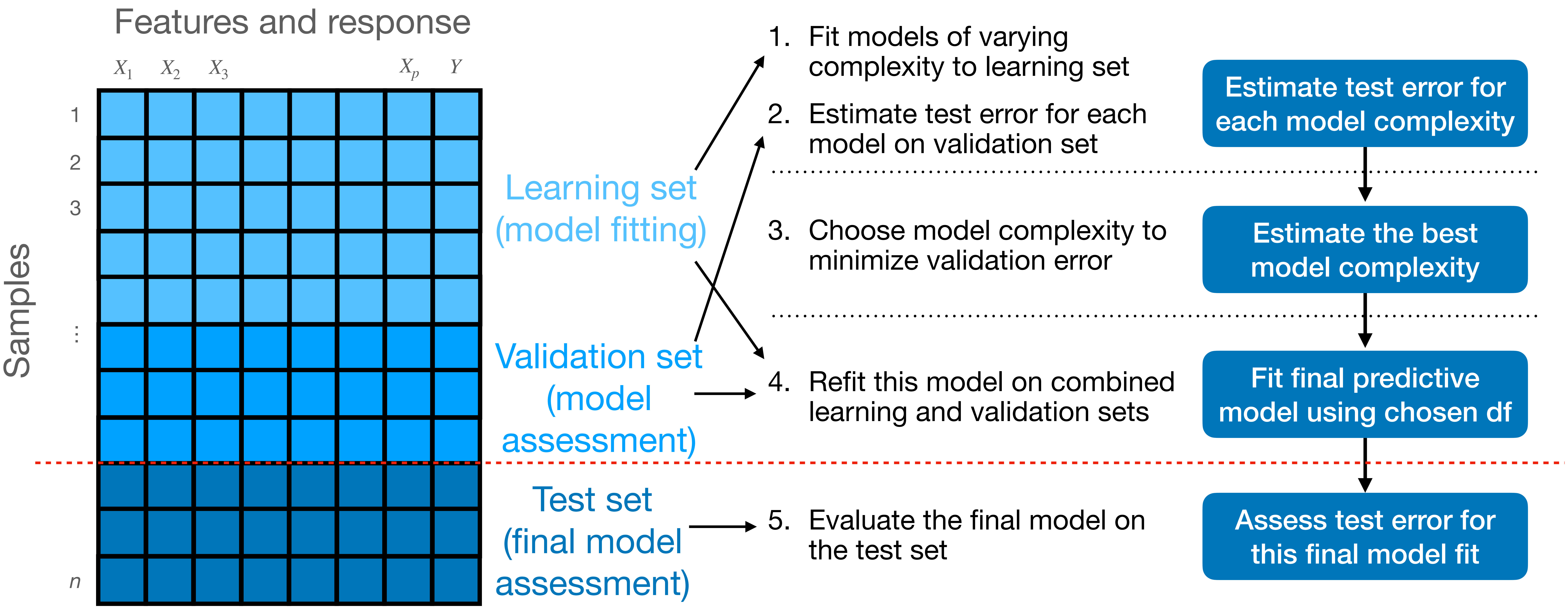
Fit final predictive model using chosen df

Assess test error for this final model fit

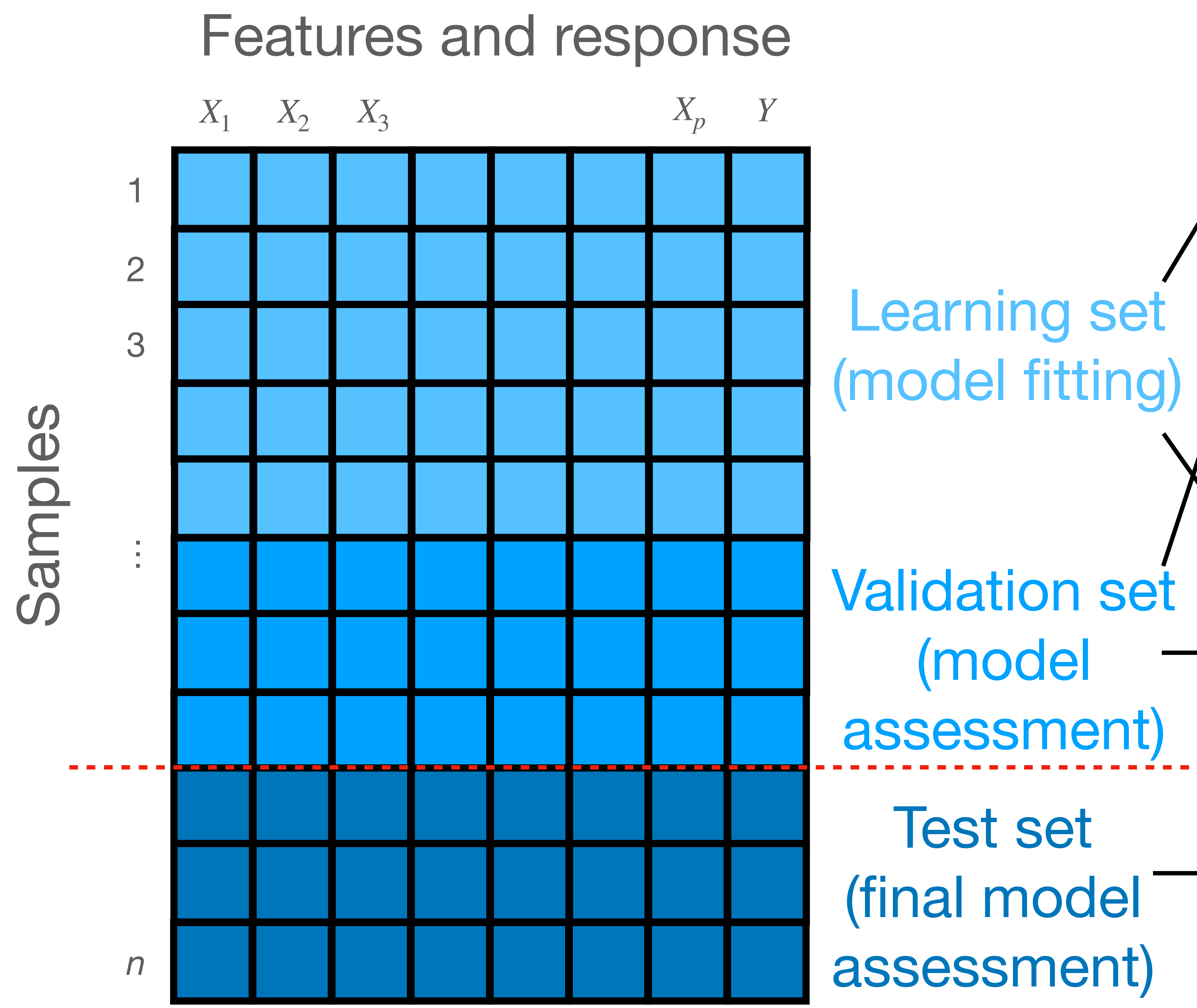
Validation set approach for model selection



Validation set approach for model selection



Validation set approach for model selection



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

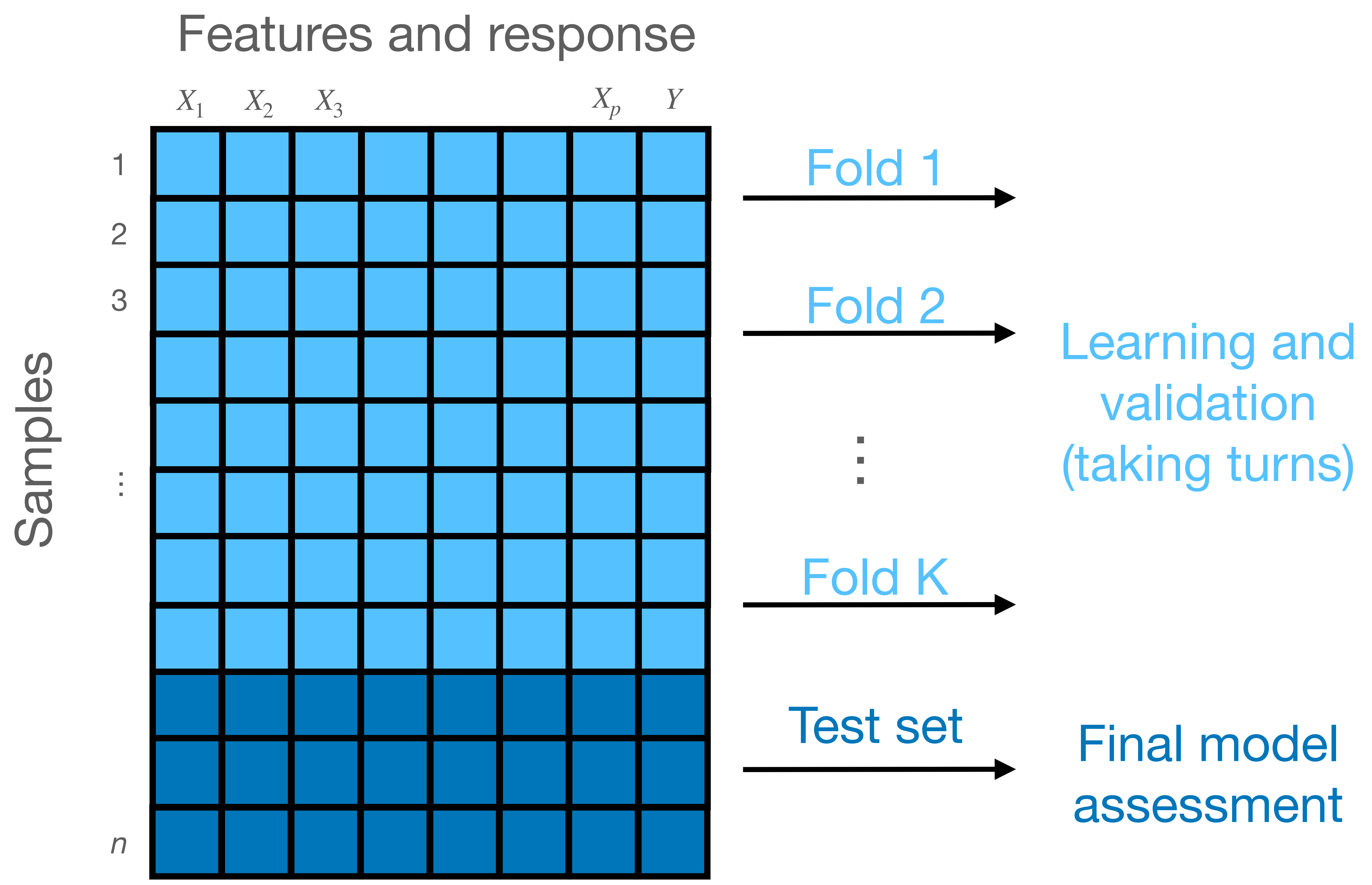
Estimate the best model complexity

Fit final predictive model using chosen df

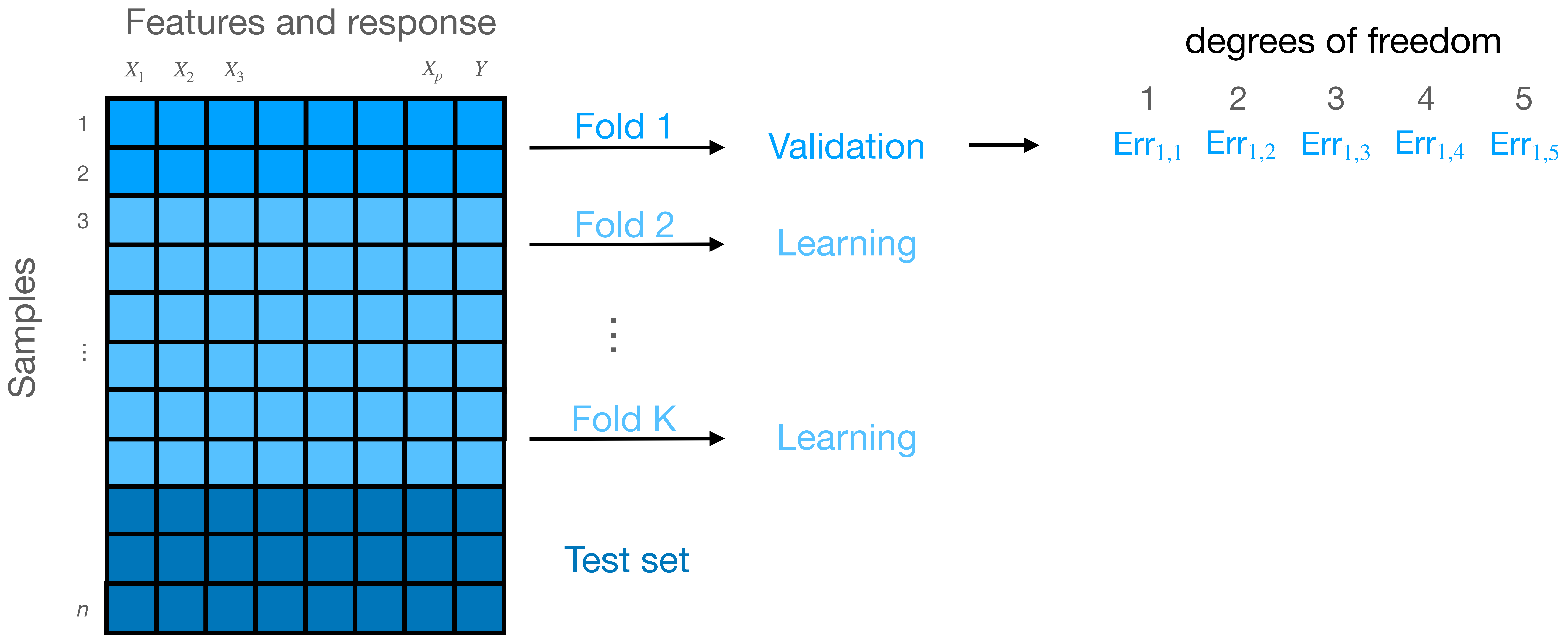
Assess test error for this final model fit

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

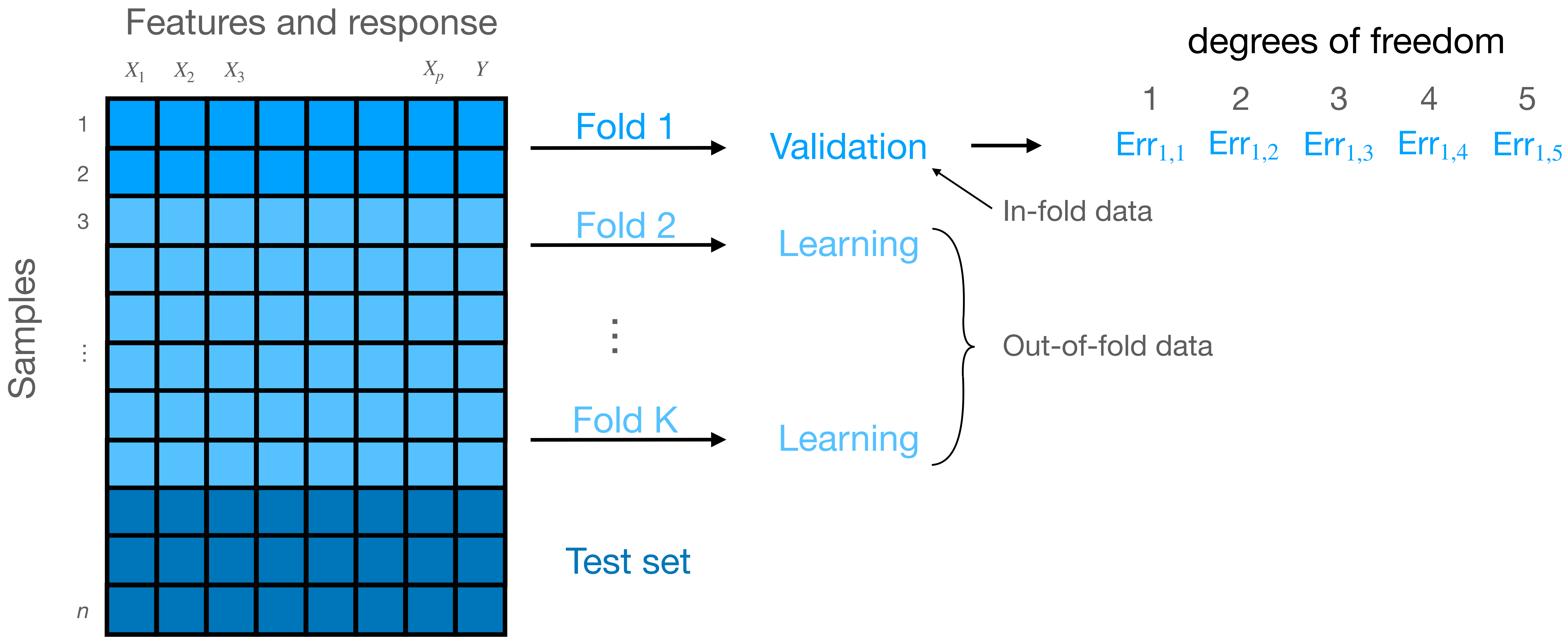
Cross-validation for model selection



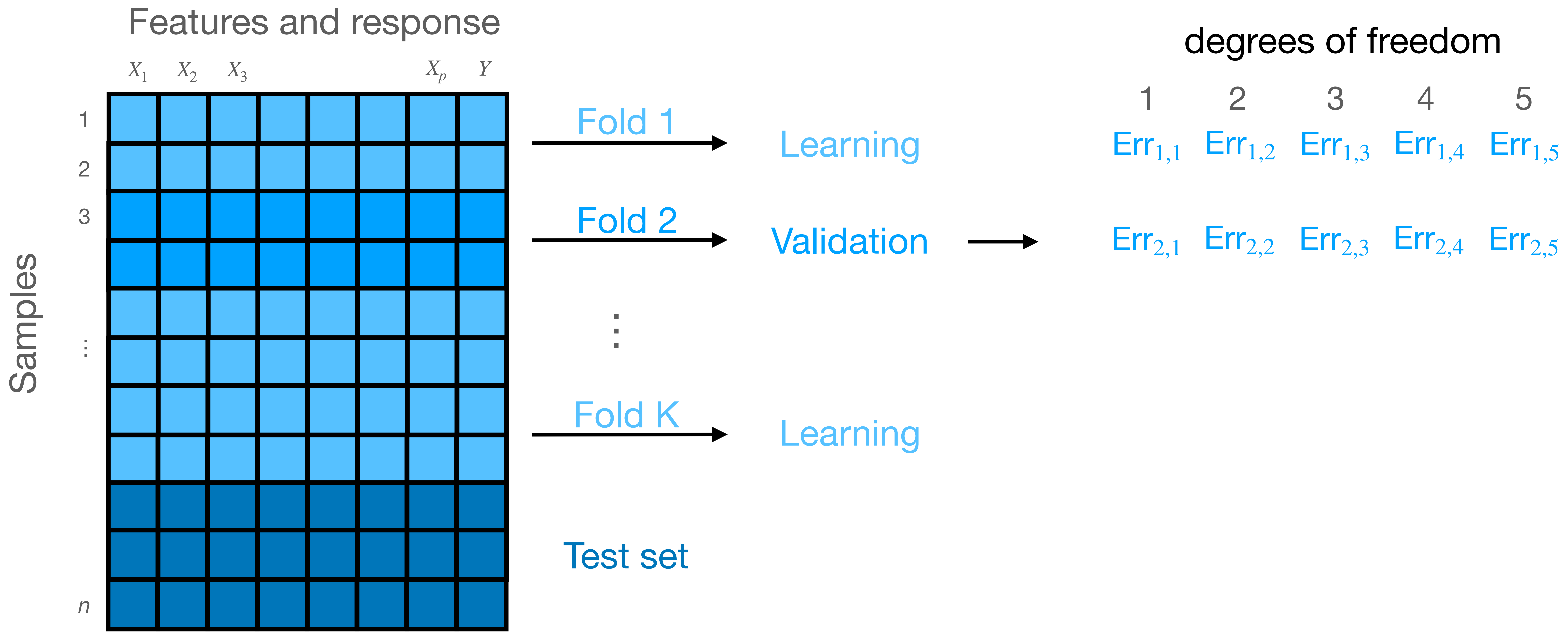
Cross-validation for model selection



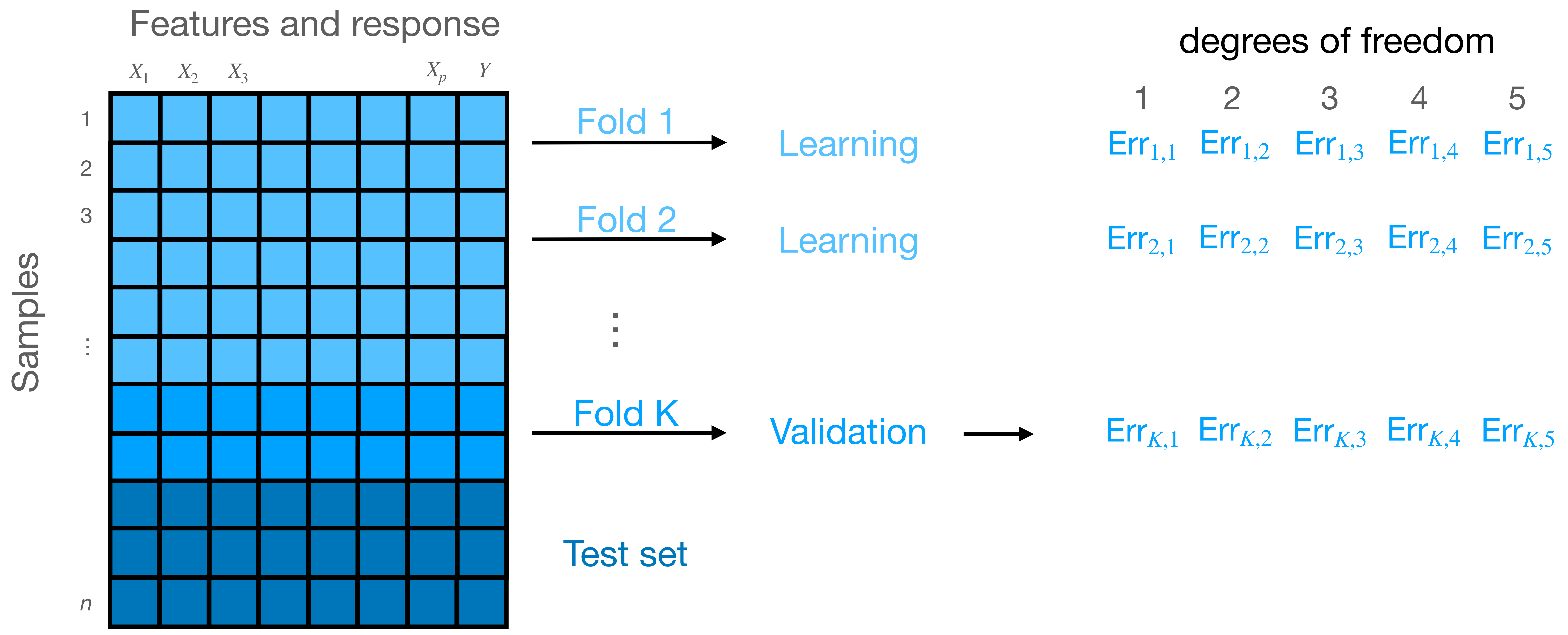
Cross-validation for model selection



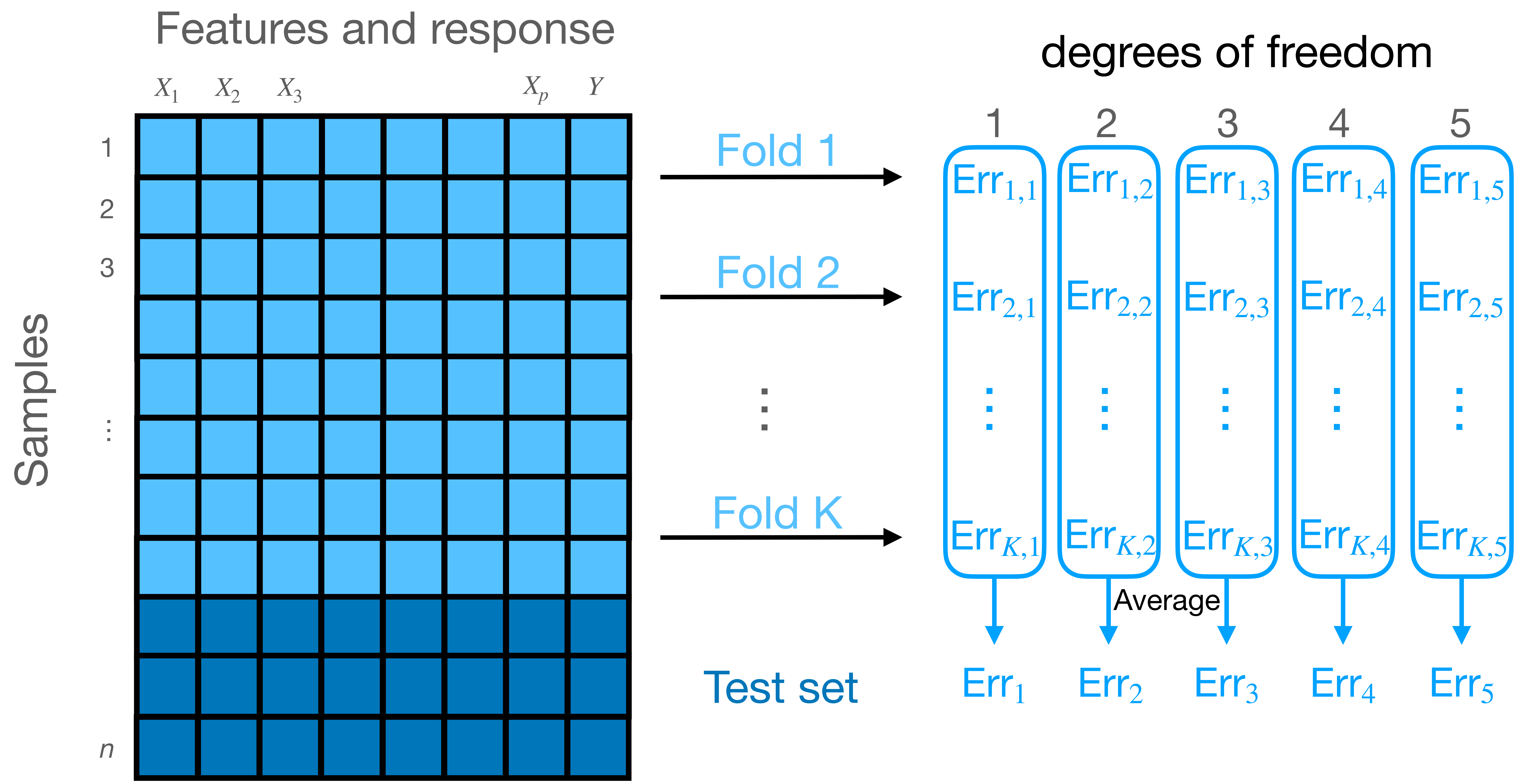
Cross-validation for model selection



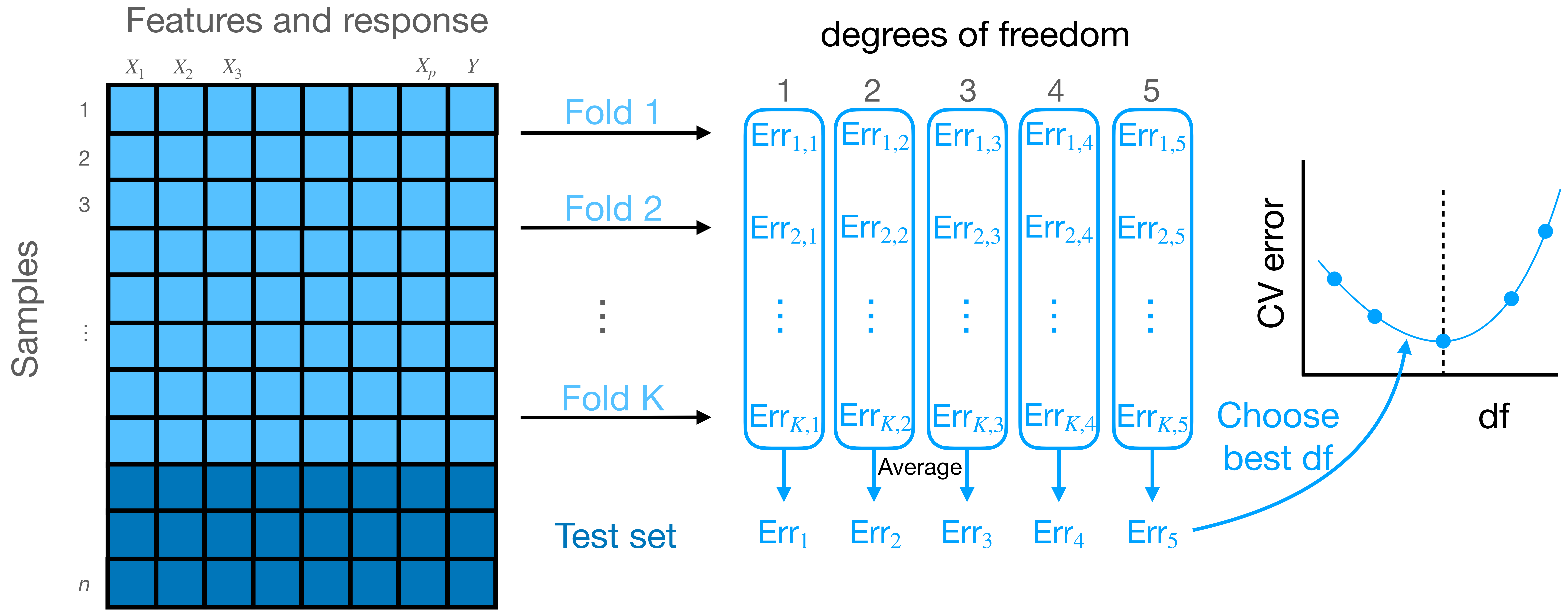
Cross-validation for model selection



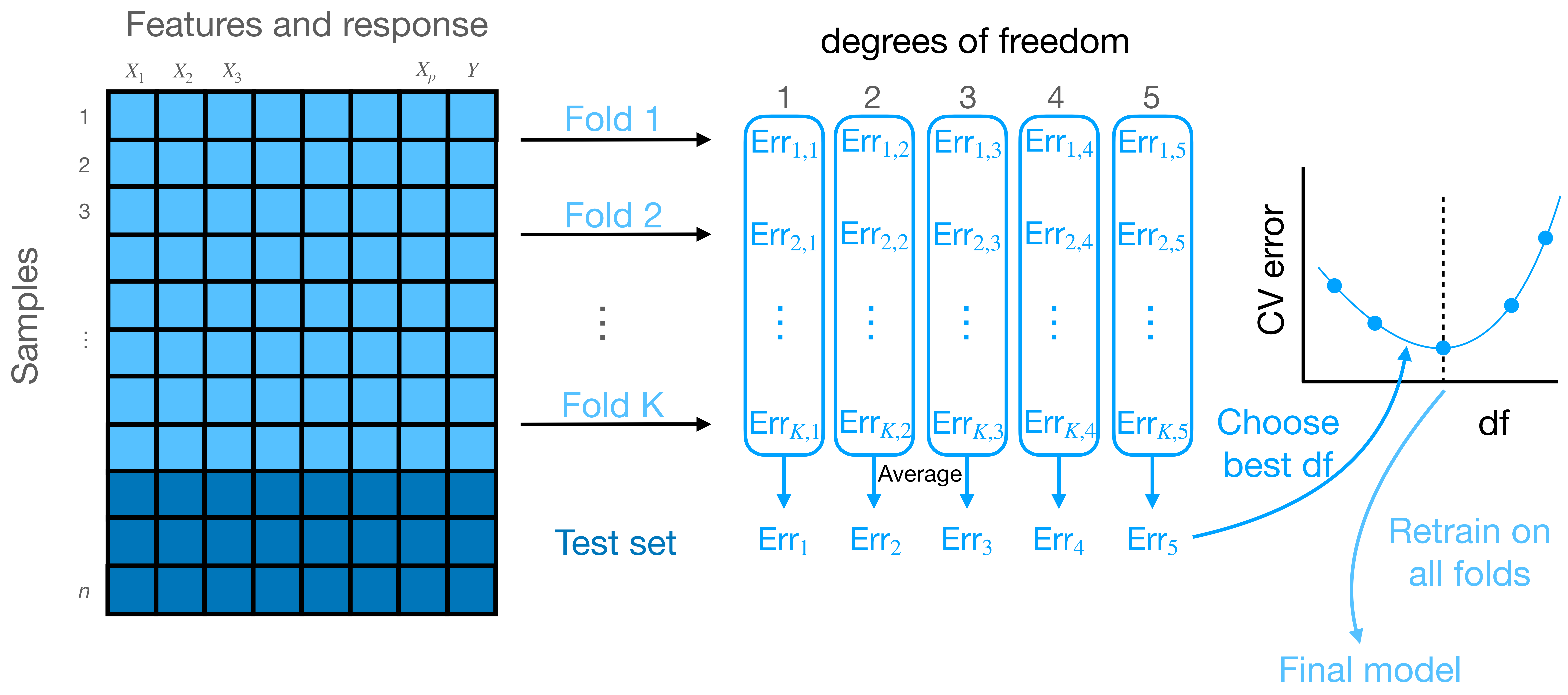
Cross-validation for model selection



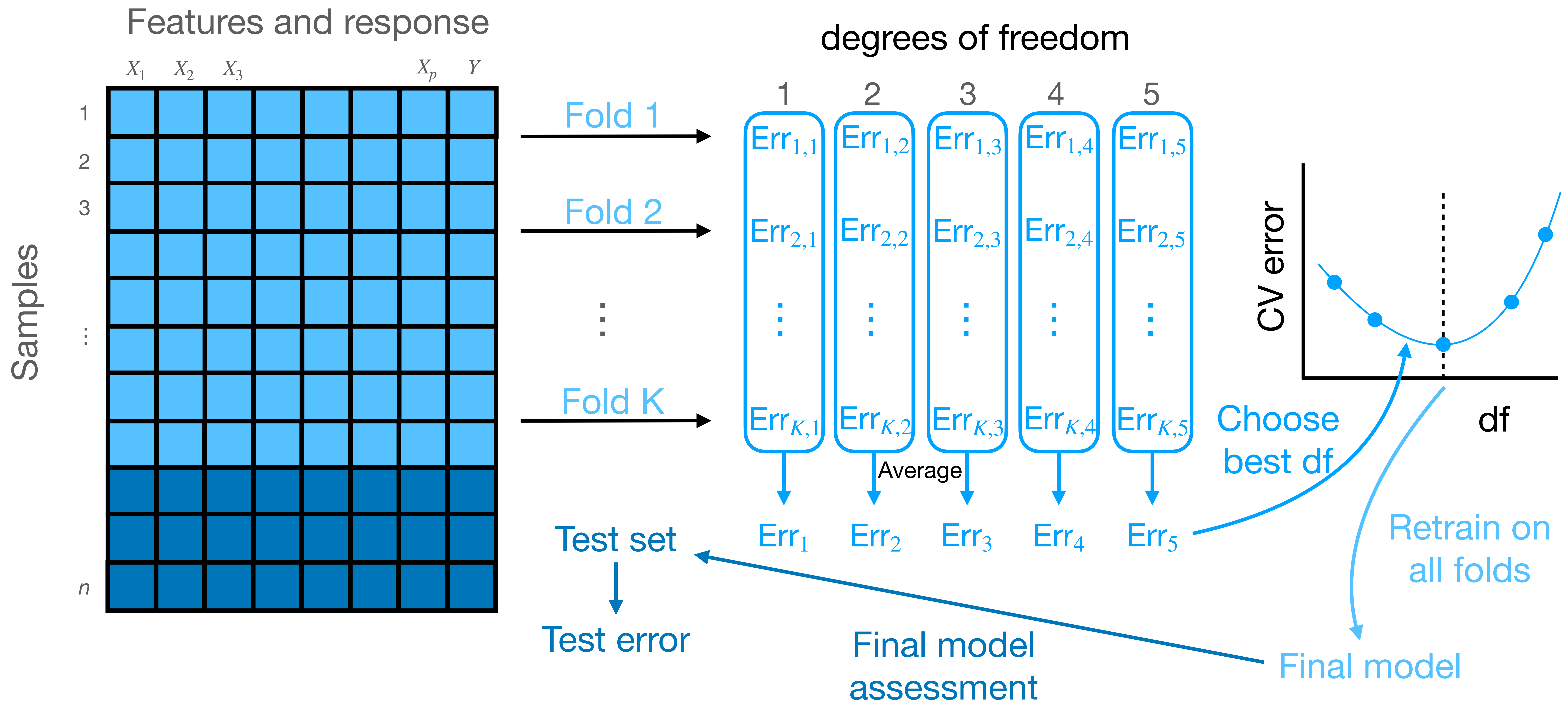
Cross-validation for model selection



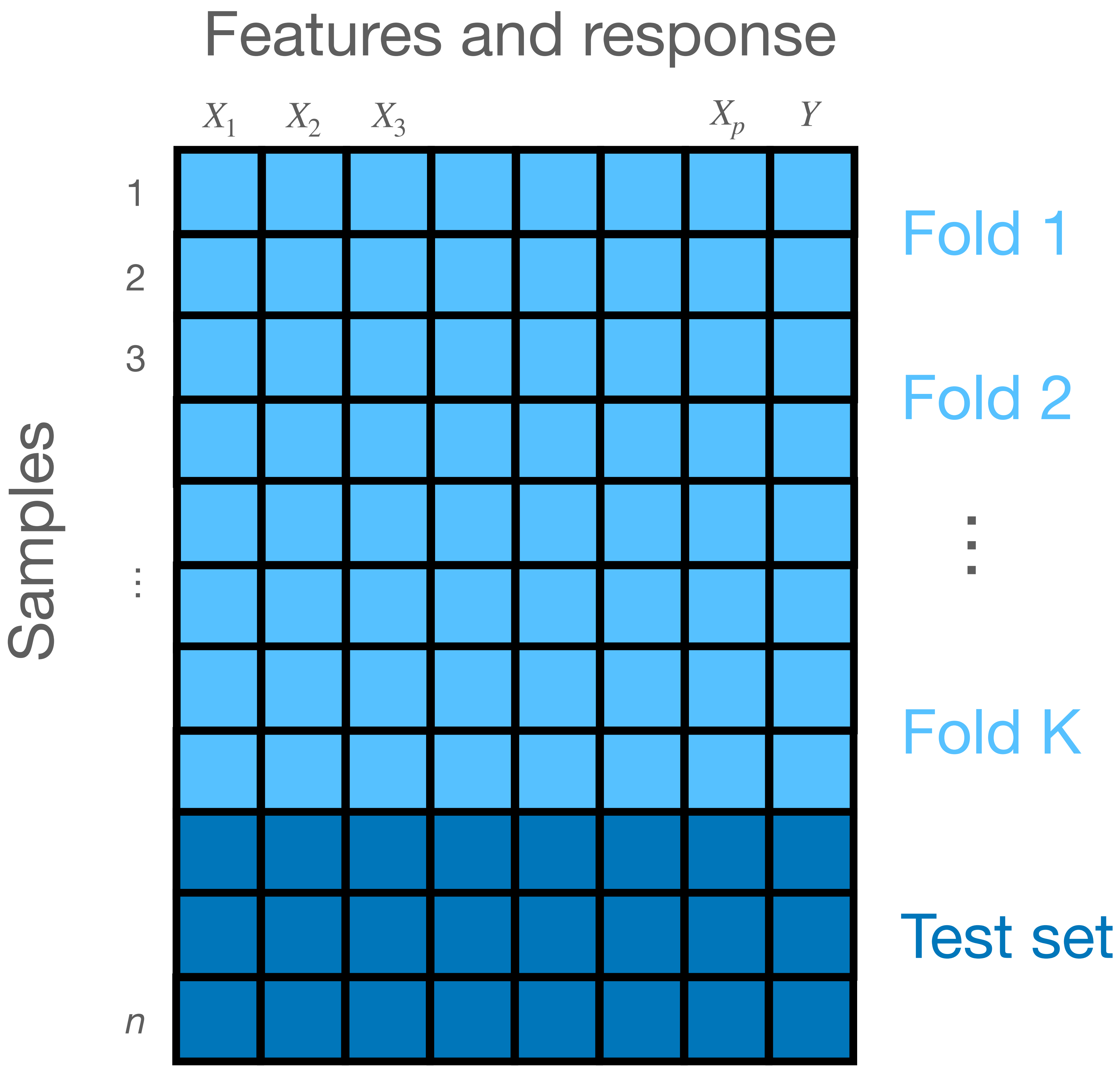
Cross-validation for model selection



Cross-validation for model selection



Cross-validation (summary)



1. Split training 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

Estimate test error for each model complexity

4. Choose model complexity to minimize CV error

Estimate the best model complexity

5. Refit this model on all folds

Fit final predictive model using chosen df

6. Evaluate final model on the test set

Assess test error for this final model fit

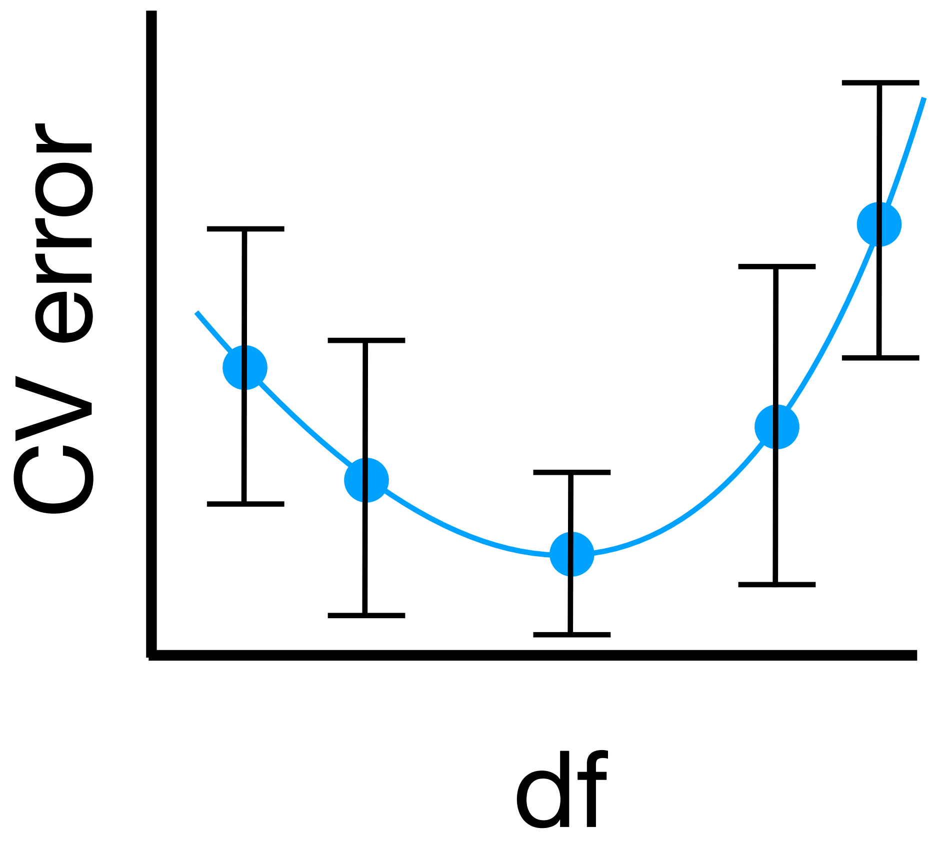
Choosing the number of folds

- More folds means more computation
- Fewer folds means the training sets used for model selection are much smaller than the actual training set
- (Advanced: More folds means that CV error estimates are more similar across folds, so the overall CV estimate defined as their average has more variance.)
- In practice, $K = 5$ or $K = 10$ are common choices

Cross-validation standard error

degrees of freedom

1	2	3	4	5
$\mathrm{Err}_{1,1}$	$\mathrm{Err}_{1,2}$	$\mathrm{Err}_{1,3}$	$\mathrm{Err}_{1,4}$	$\mathrm{Err}_{1,5}$
$\mathrm{Err}_{2,1}$	$\mathrm{Err}_{2,2}$	$\mathrm{Err}_{2,3}$	$\mathrm{Err}_{2,4}$	$\mathrm{Err}_{2,5}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathrm{Err}_{K,1}$	$\mathrm{Err}_{K,2}$	$\mathrm{Err}_{K,3}$	$\mathrm{Err}_{K,4}$	$\mathrm{Err}_{K,5}$
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Err_1	Err_2	Err_3	Err_4	Err_5
se_1	se_2	se_3	se_4	se_5

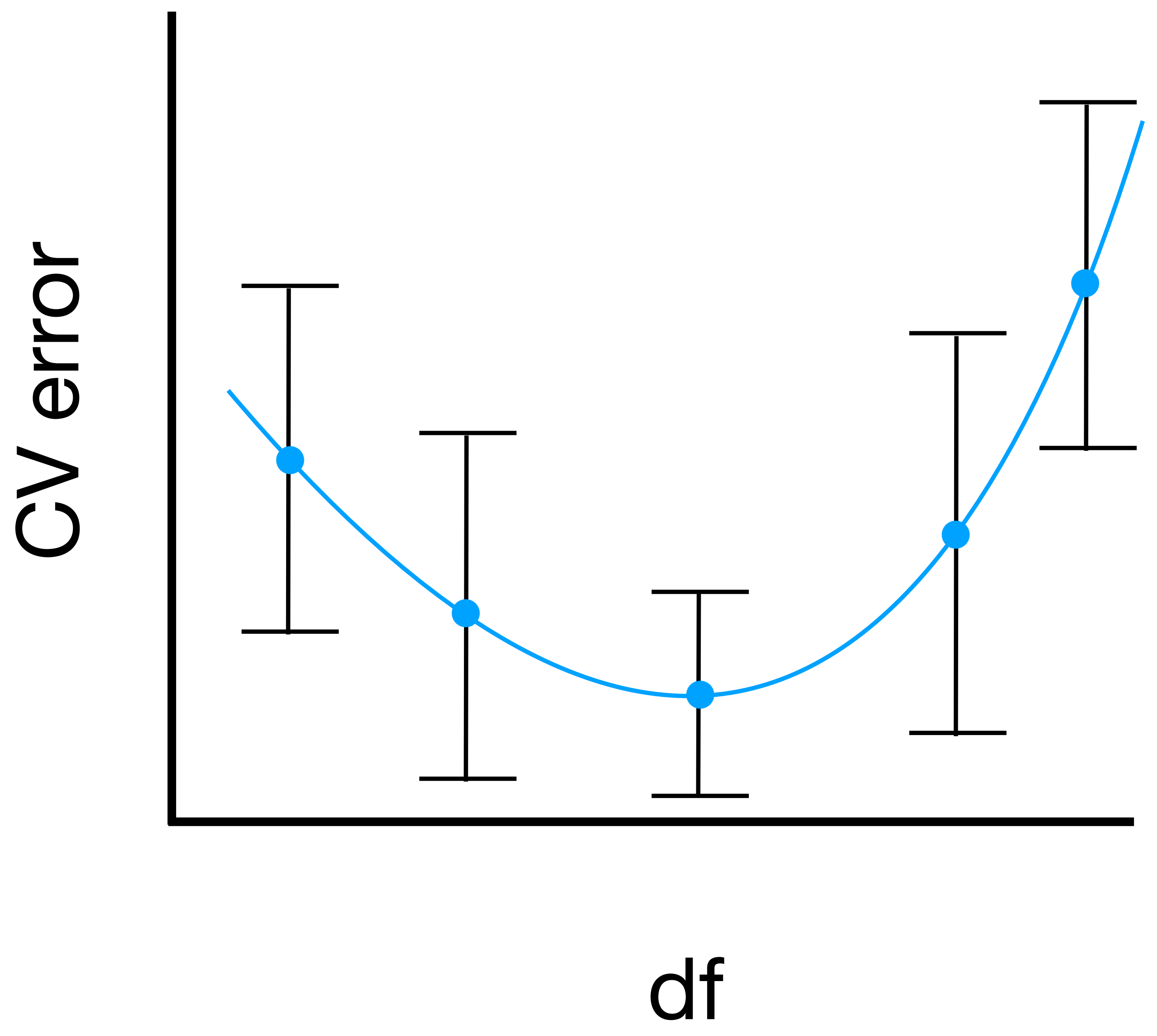


$$\mathrm{se}_{\mathrm{df}} = \frac{1}{\sqrt{K}} \times \mathrm{s.d.}(\mathrm{Err}_{1,\mathrm{df}}, \ldots, \mathrm{Err}_{K,\mathrm{df}})$$

One standard error rule

Occam's razor:

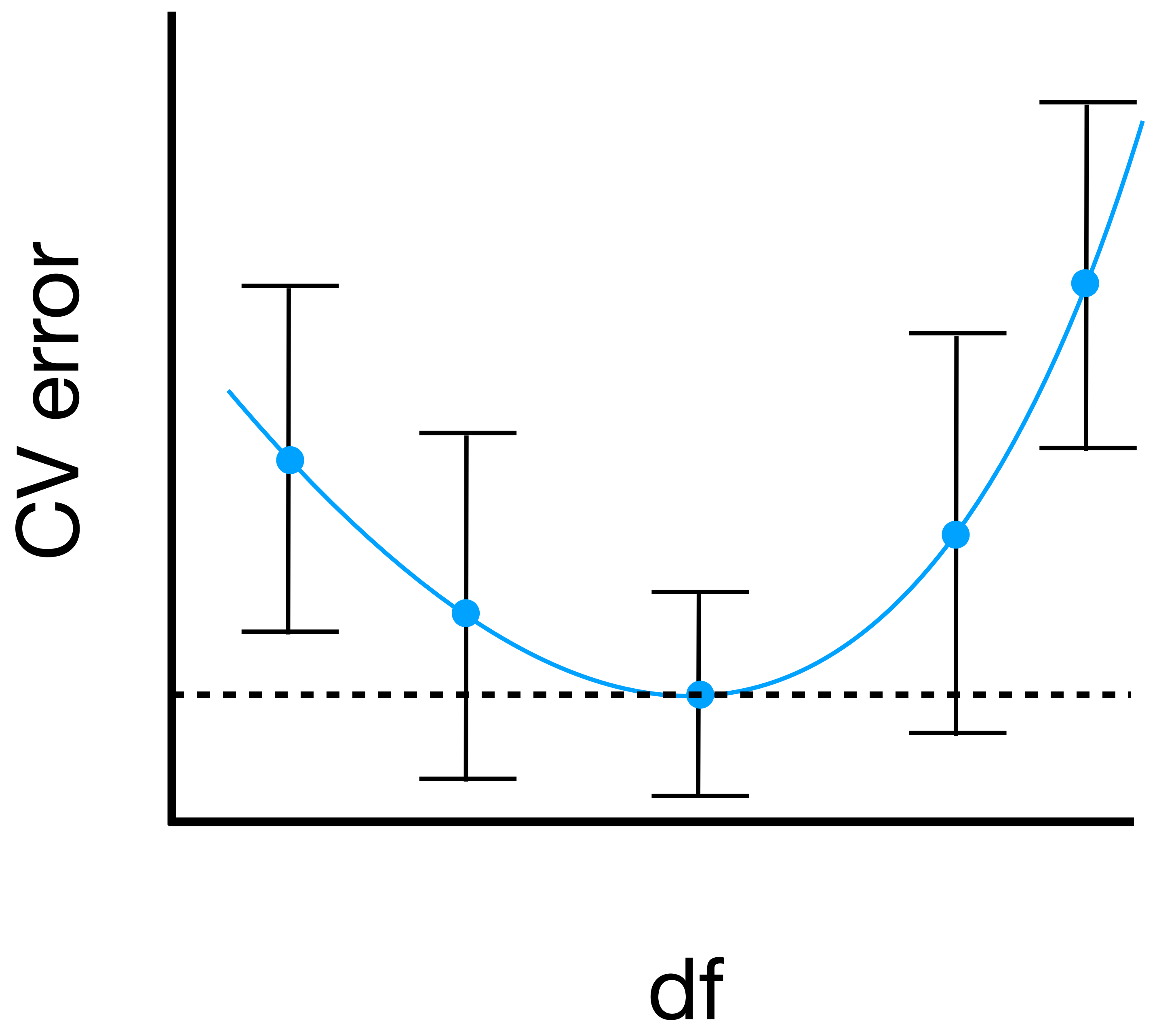
Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.



One standard error rule

Occam's razor:

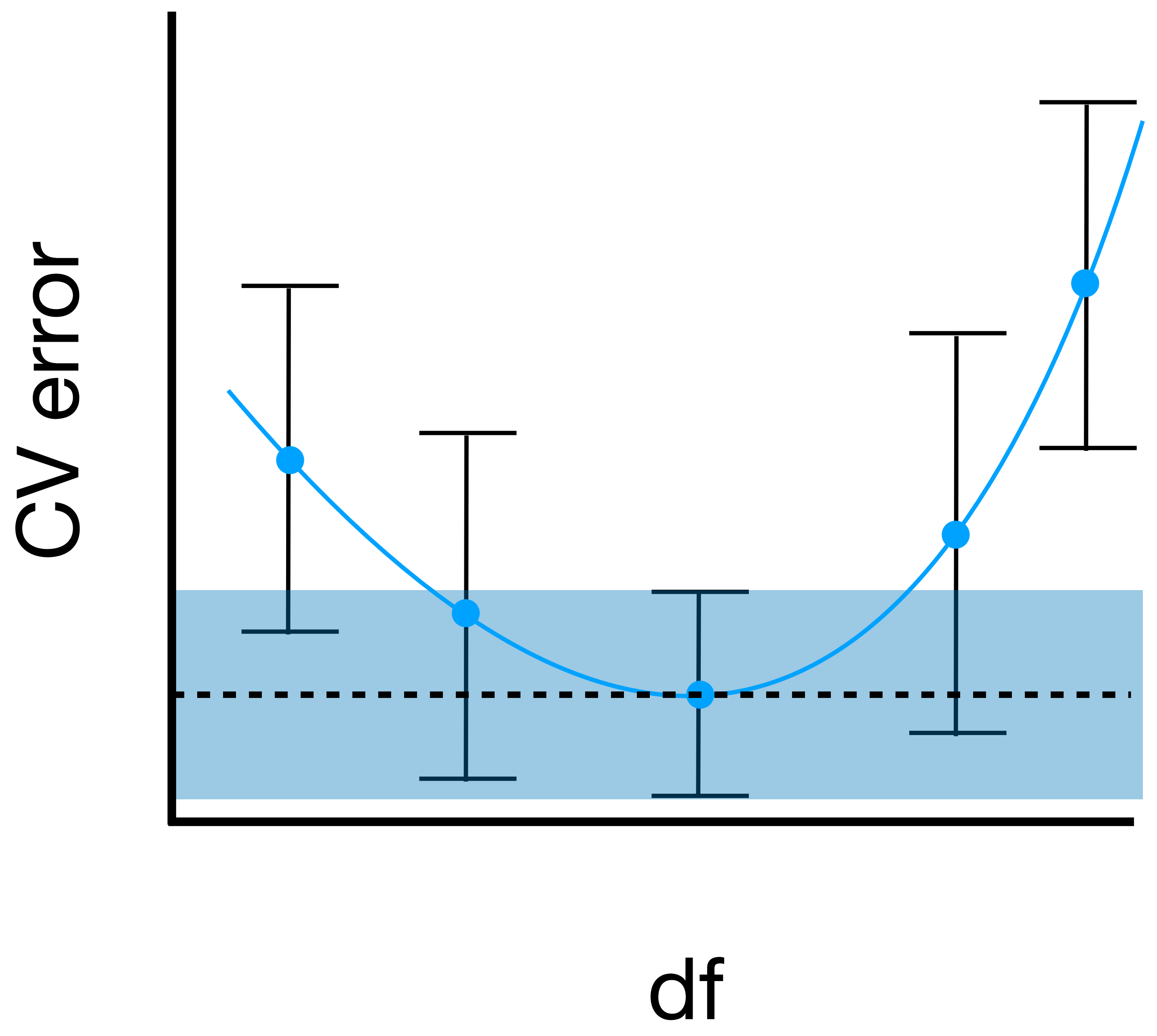
Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.



One standard error rule

Occam's razor:

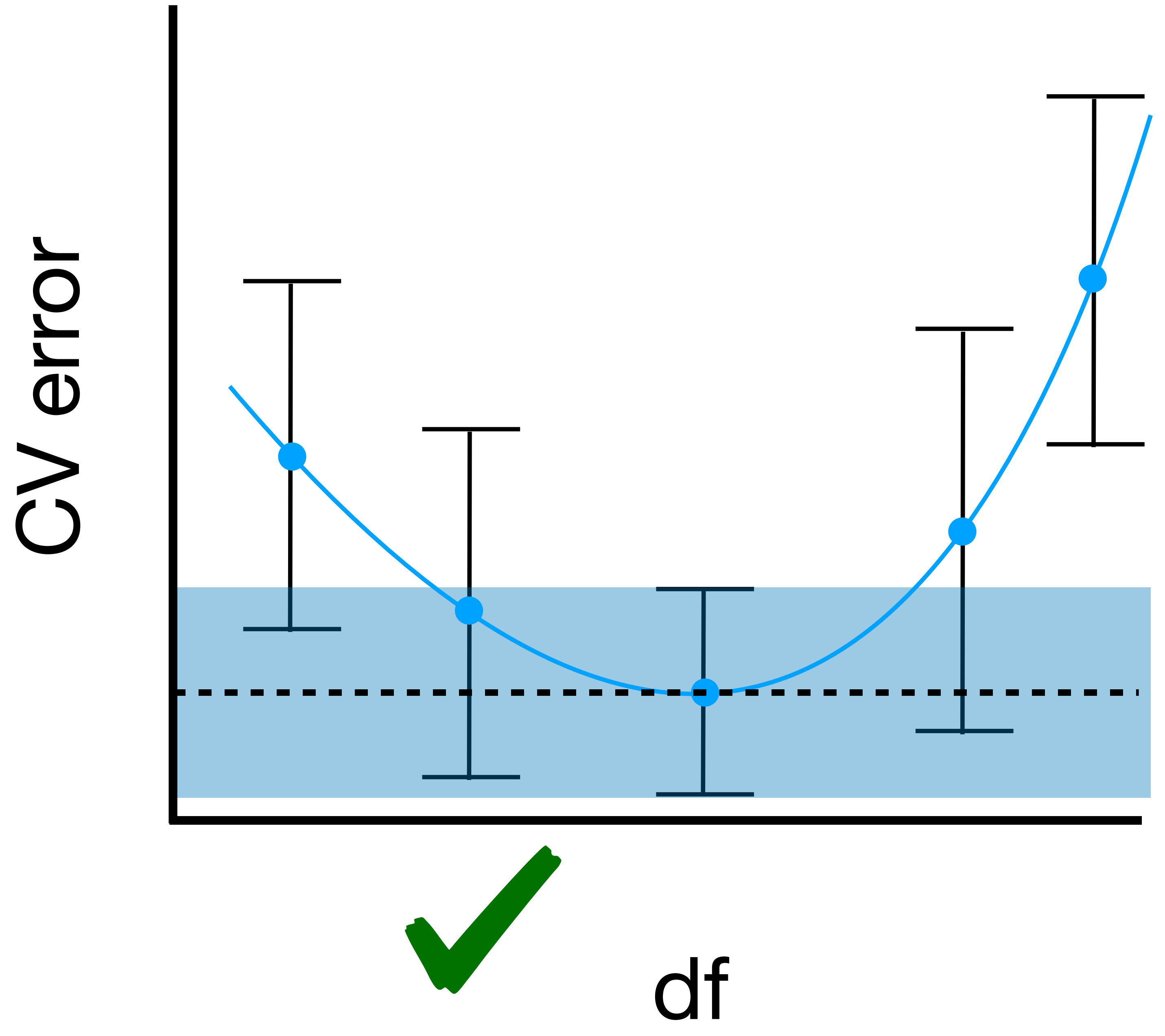
Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.



One standard error rule

Occam's razor:

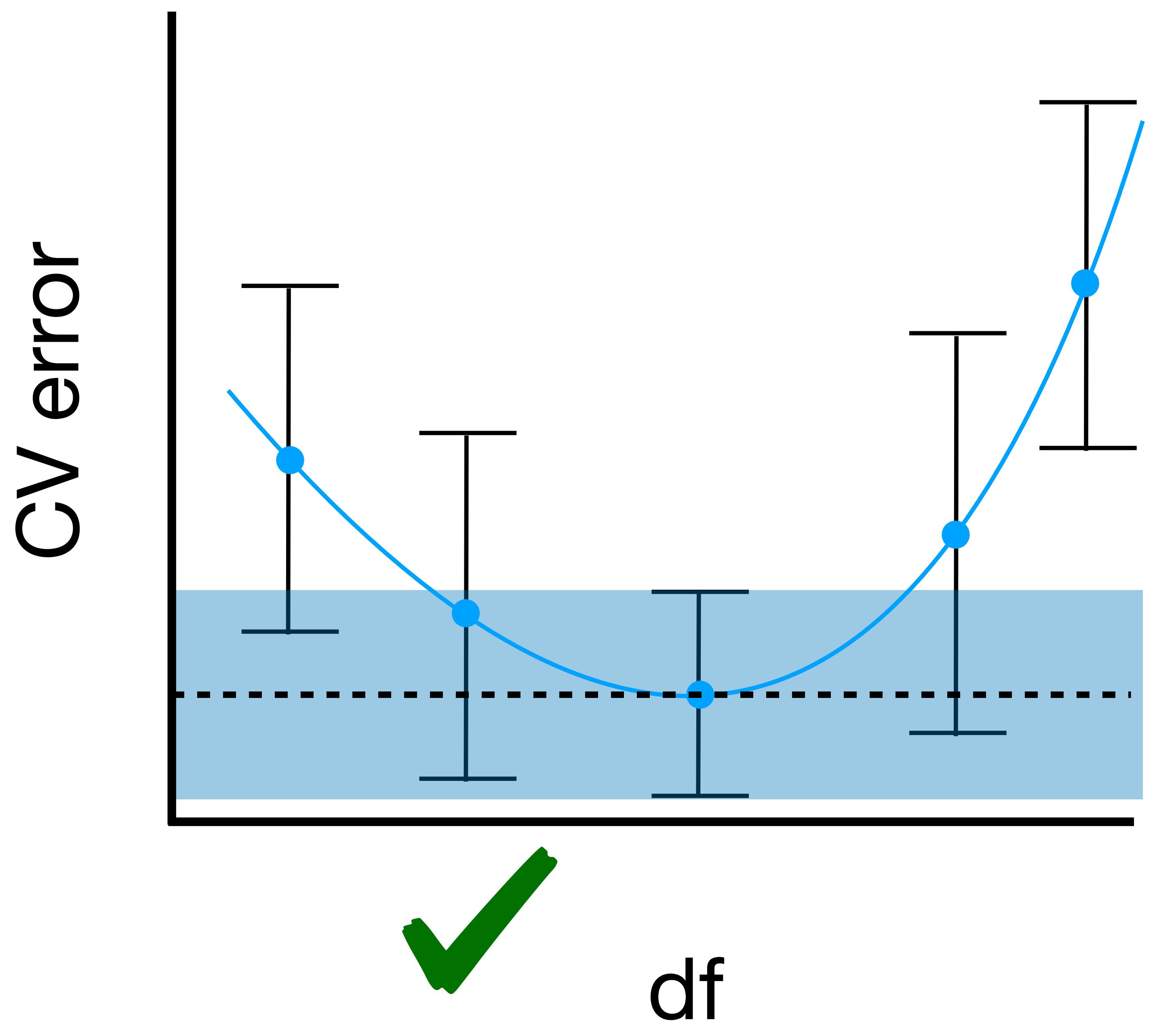
Select the smallest model for which the CV error is within one standard error of the lowest point on the curve.



One standard error rule

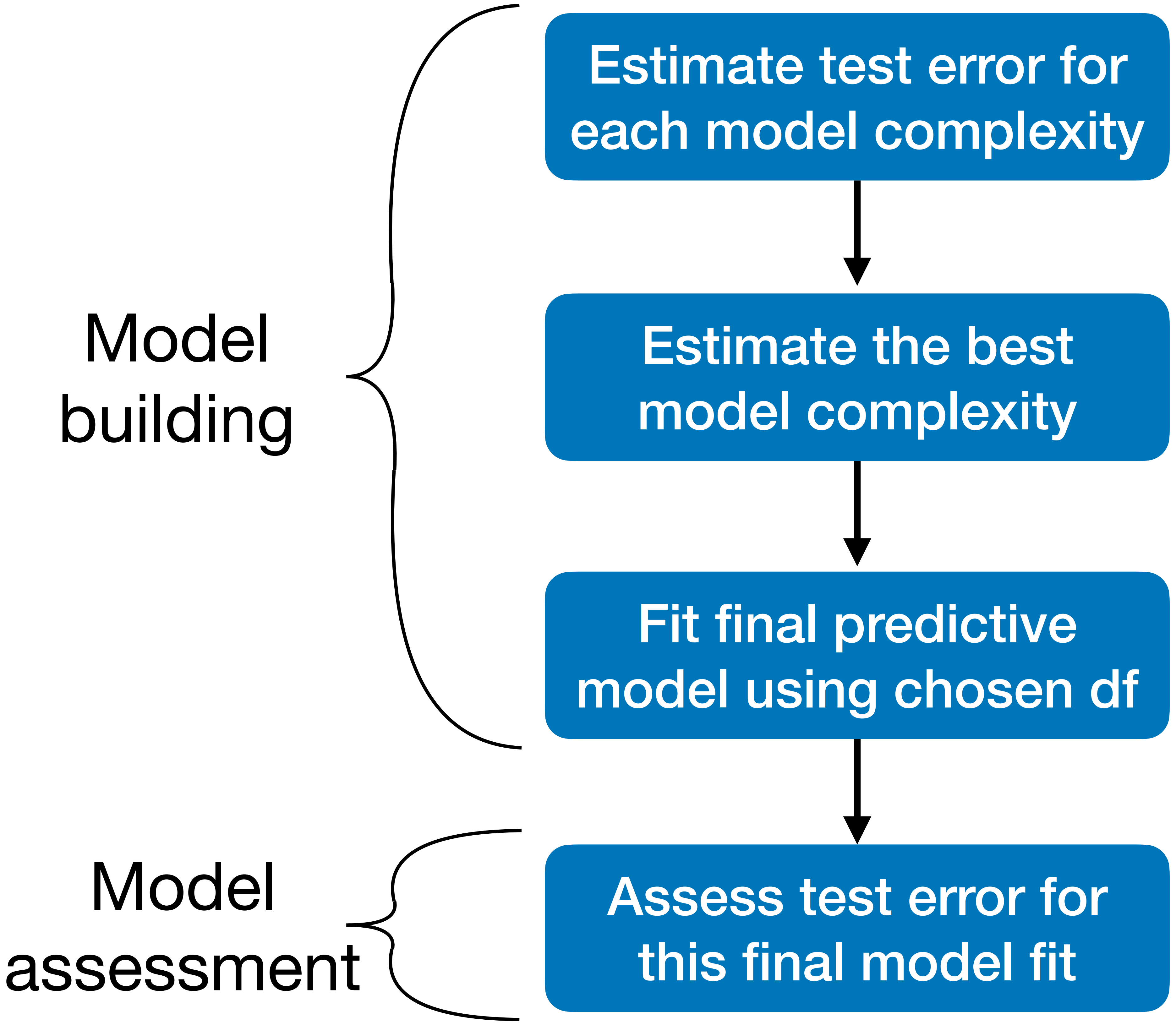
Occam's razor:

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

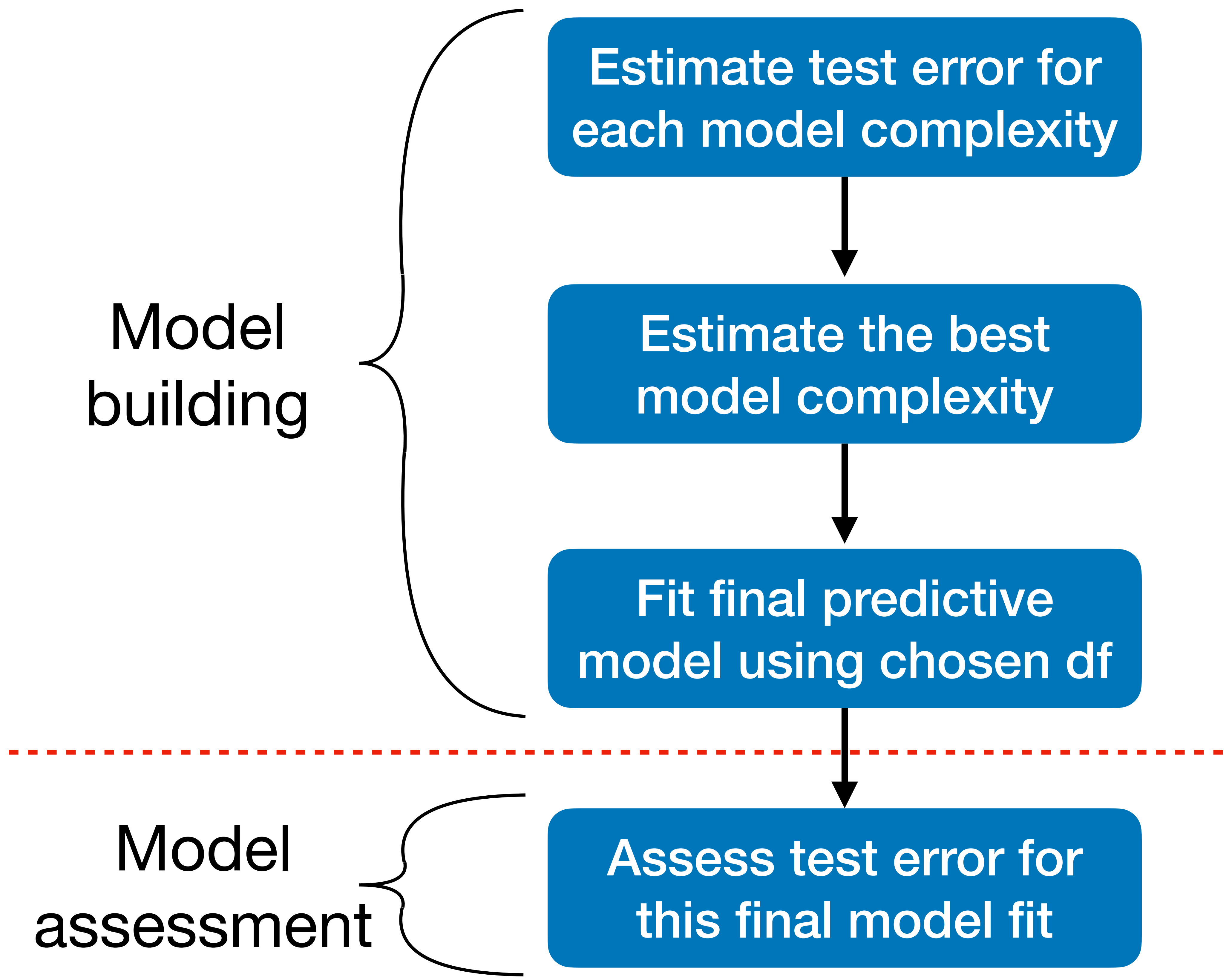


Often used instead of choosing the minimum of CV curve.

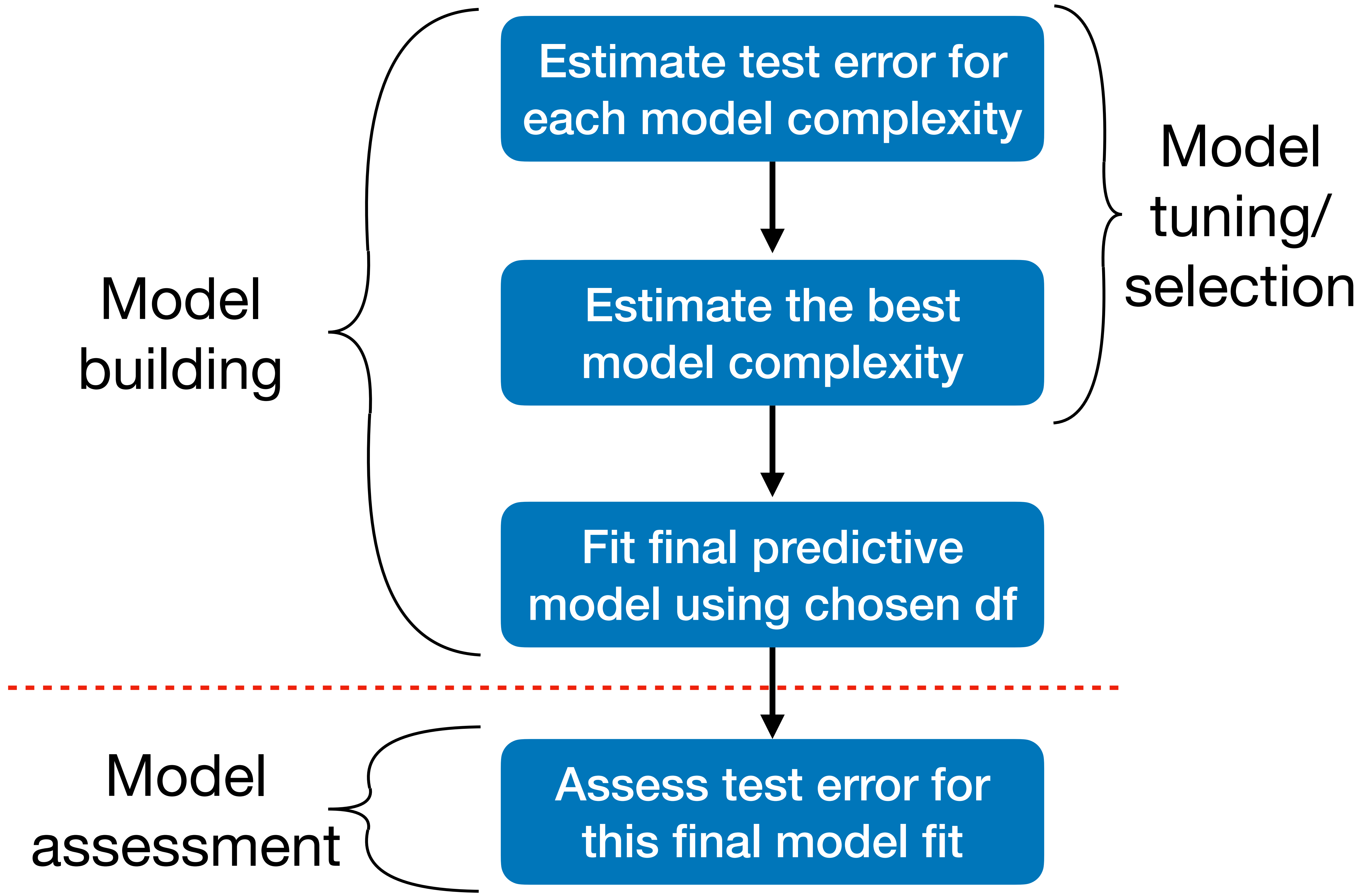
Summary


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

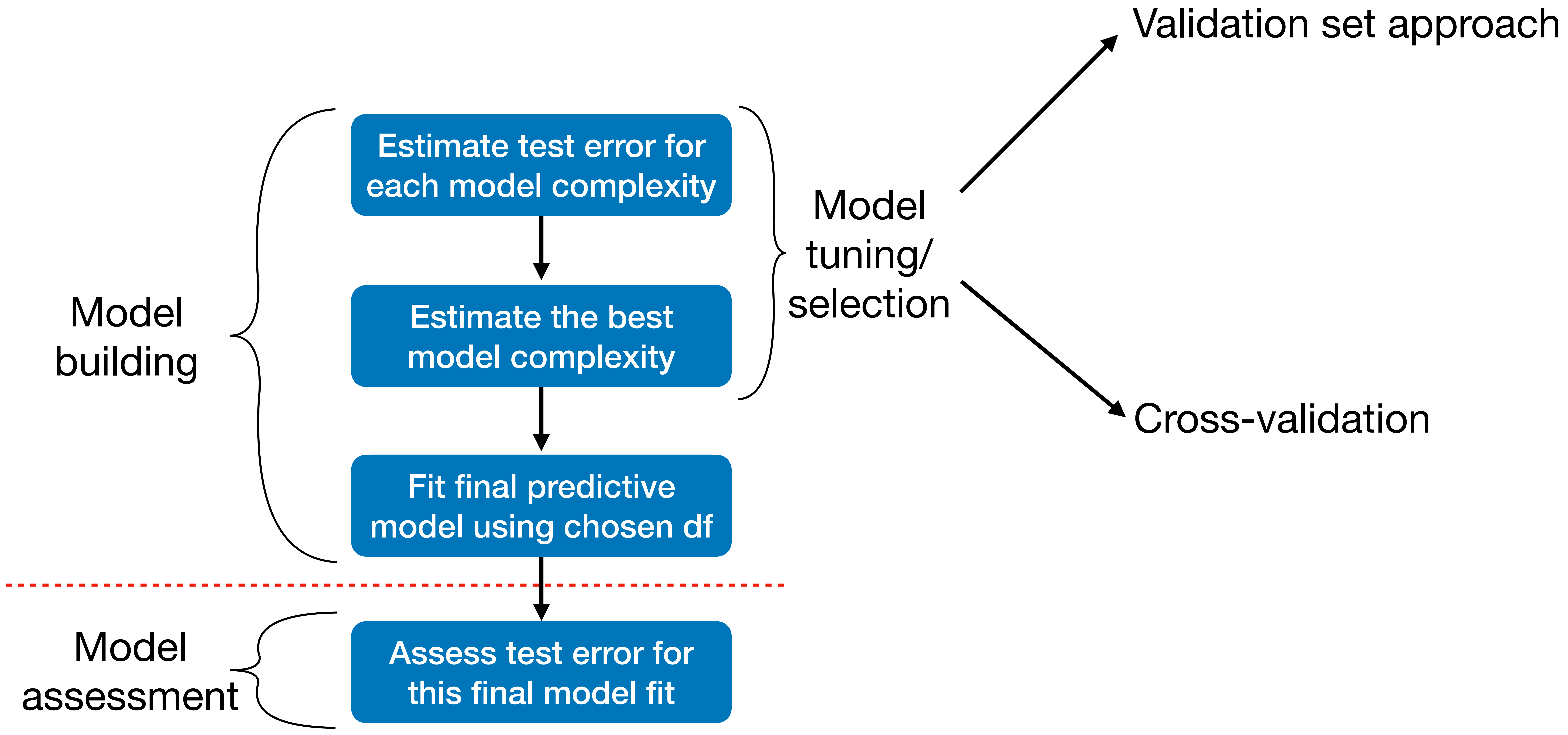
Summary


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

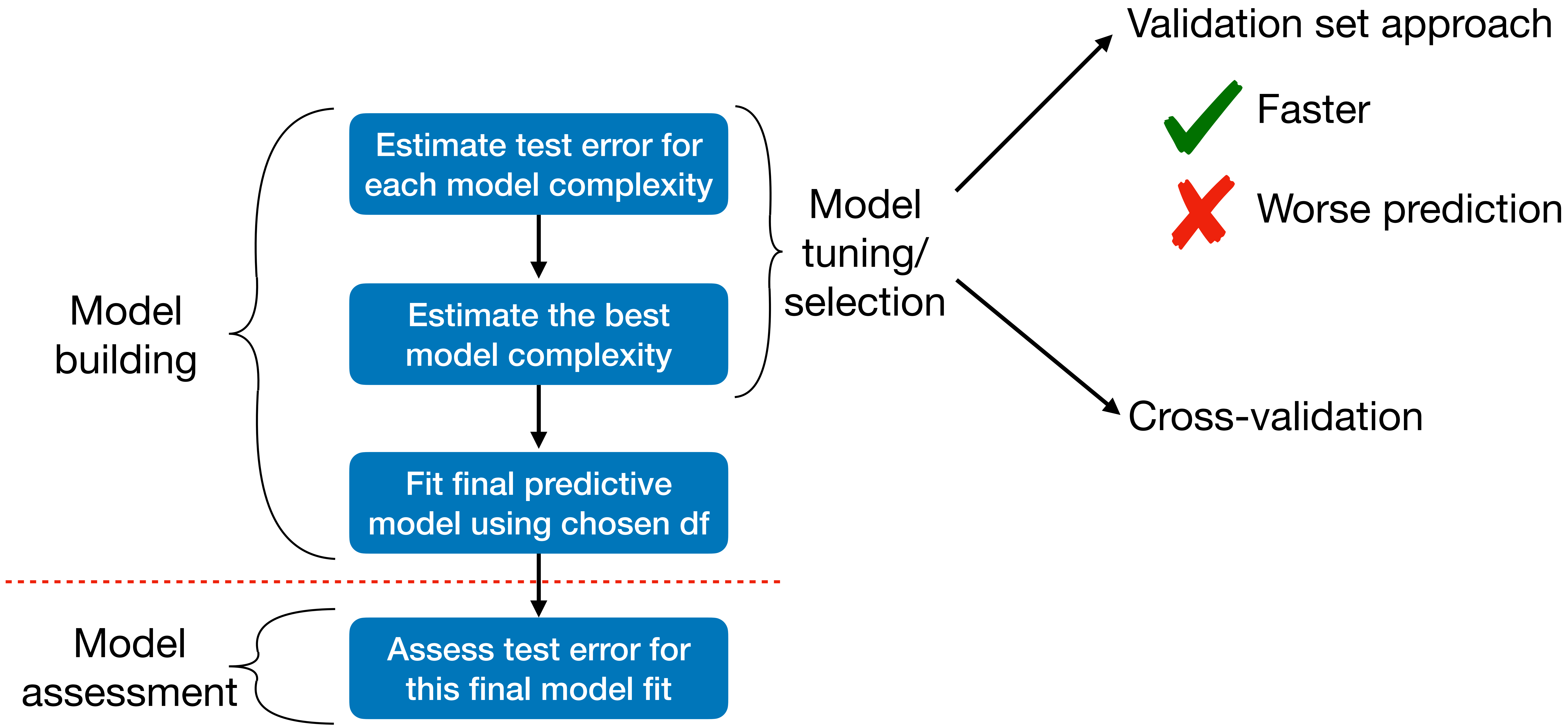

Estimate test error for each model complexity
Estimate the best model complexity
Fit final predictive model using chosen df
Assess test error for this final model fit
Model building
Model tuning/ selection
Model assessment

Summary

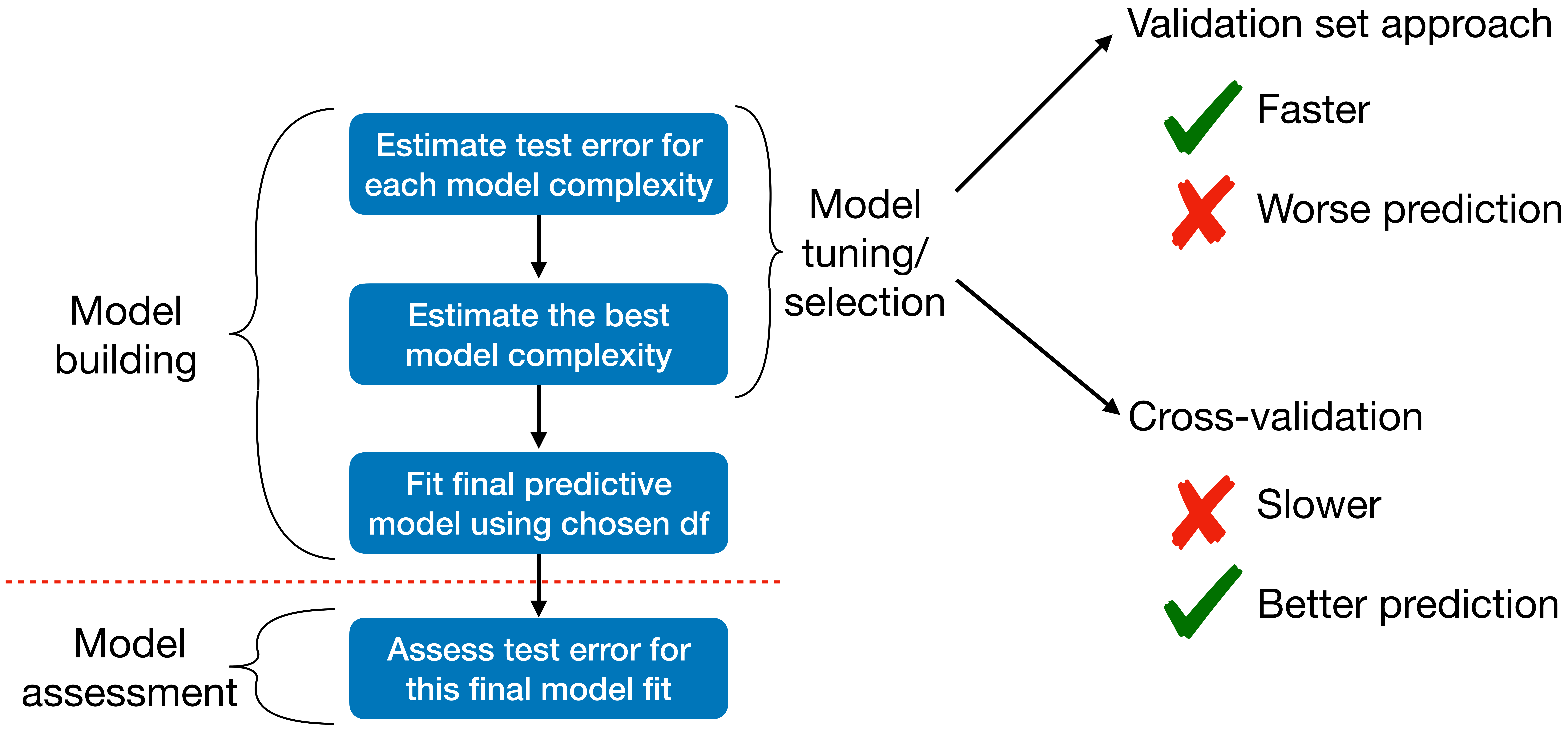


Validation set approach
Estimate test error for each model complexity
Model tuning/ selection
Model building
Estimate the best model complexity
Cross-validation
Fit final predictive model using chosen df
Model assessment
Assess test error for this final model fit

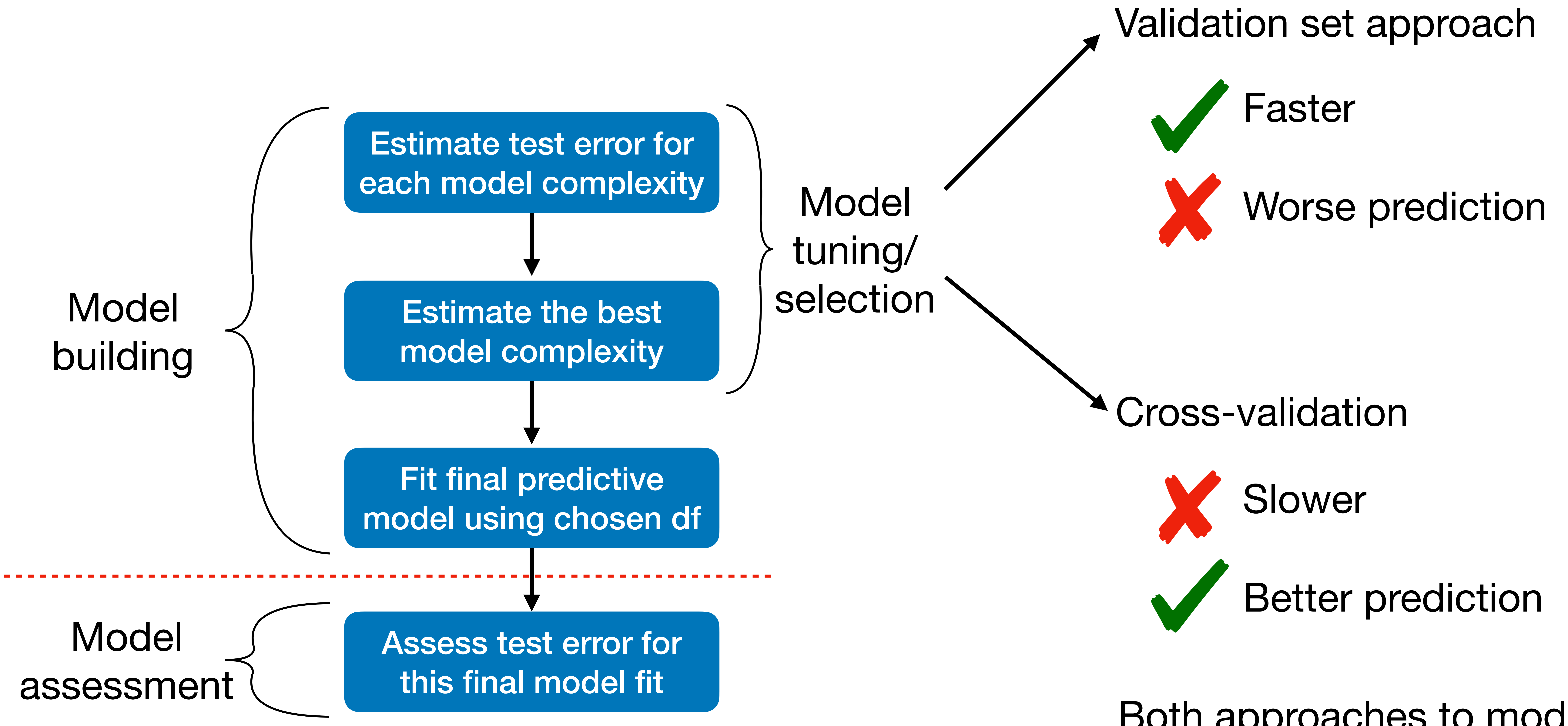
Summary


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 assessment
Model tuning/ selection
Validation set approach
Faster
Worse prediction
Cross-validation

Summary


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

Summary



Both approaches to model tuning used in practice.