Validation and Optimization of Machine Learning Models Regression

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

2021-06-17
ESCAPE Summer School

Claudia Beleites
Chemometrix GmbH, Södeler Weg 19, 61200 Wölfersheim, Germany

Topics

1. Introduction
2. Figures of Merit
3. Verification Schemes
4. Resampling Techniques
5. Model Stability
6. Sample Size Planning
7. Validation
8. Data-driven Model Optimization and Hyperparameter Tuning
9. Regression

CHEMO METRIX

Validation and Verification

C. Beleites

Verification: Making sure/measuring/showing that the model meets the specifications.

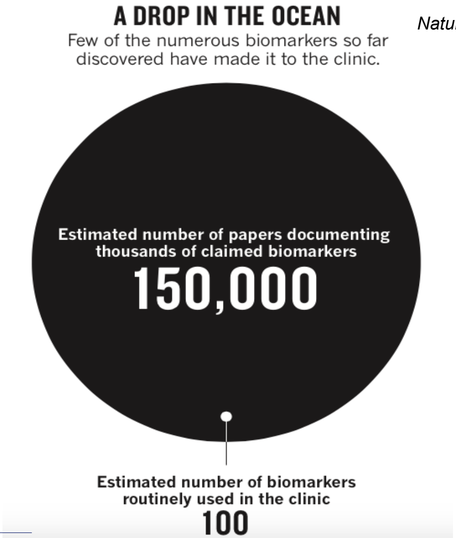
Validation: Making sure that the model meets the application needs.

- Chemometric model validation
  ⇝ typically verification rather than validation is done.
- Characterize model by measuring its predictive performance

CHEMO METRIX

Reproducibility!?

Nature **469**, 156-157



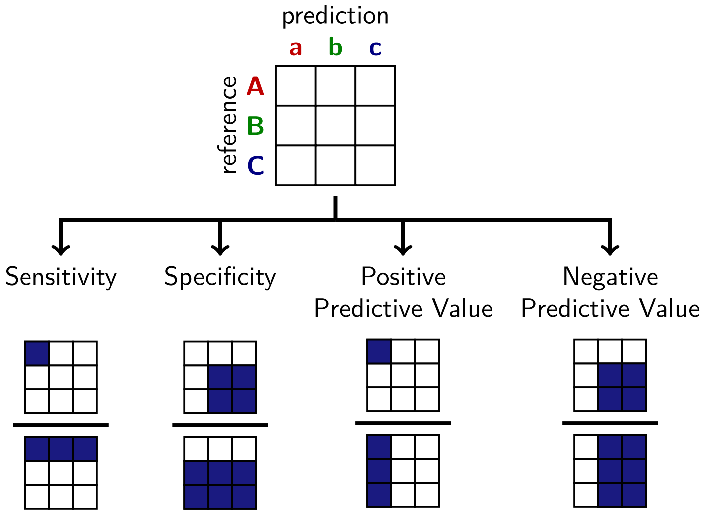
Poste: Bring on the biomarkers. Nature, 2011, 469, 156-157.

Recipe: Verification/Validation/Testing

Ingredients

- Ready-to-use model
  treated as black box: *case* $\mapsto$ *prediction*
- Figures of merit (performance measure)
  Overall Accuracy, Sensitivity, Specificity, Predictive Values, MSE, RMSE, R^2, ...
- Validation *scheme*: *How to get test cases?*
  ~~Autoprediction~~, Resampling, Test Set, Validation study

Figures of Merit: Proportions



Proportion Questions

Sensitivity = Recall: of all truly class *A* cases, which fraction is correctly recognized as class *A*?

Specificity: of all cases truly not belonging to class *A*, which fraction is correctly recognized as not belonging to class *A*?

Positive Predictive Value = Precision: of all cases predicted to belong to class *A*, which fraction does truly belong to class *A*?

Negative Predictive Value: of all cases predicted not to belong to class *A*, which fraction does truly not belong to class *A*?

accuracy: correct proportion among all predicted cases

error rate: misclassified proportion among all predicted cases

κ: chance-corrected accuracy, inter-observer agreement

Proportions: Characteristics

✔ well-known, widely used

✘ often misunderstood:

- sensitivity & specificity
  - ✔ easy to measure: test *n* cases of each class, record results
  - ✘ low relevance for application
- predictive values (positive/negative)
  - ✔ high relevance for application
  - ✘ difficult to measure: need to know relative class fequencies under application conditions
    weight rows of confusion matrix accordingly
- figures of merit spanning rows of confusion matrix
  - ✘ correct for relative class fequencies under application conditions

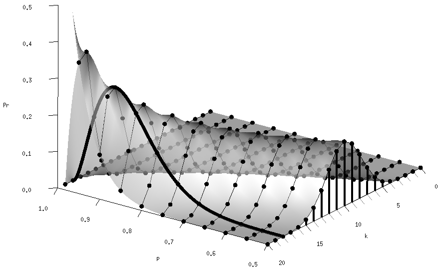
More figures of merit

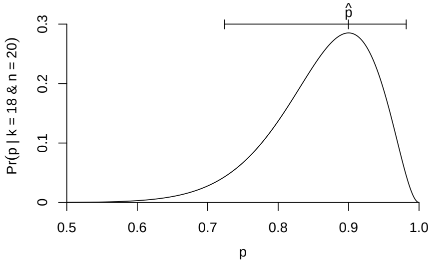
- chance-corrected: κ
  - ✔ rescaling possible for other figures of merit
  - ✔ alternative: report chance agreement (or naive model performance) together with figure of merit

- Information gain
  - **positive likelihood ratio:** $LR_A^+ = \frac{Sens_A}{1 - Spez_A}$
    How much do the odds to belong to class A increase when a case is predicted to belong to class A?
  - **negative likelihood ratio:** $LR_A^- = \frac{Spez_A}{1 - Sens_A}$
    How much do the odds to belong to class A decrease when a case is predicted not to belong to class A?
  - ✔ independent of relative class fequencies under application conditions

CHEMO METRIX

Confidence Intervals for Sensitivity





- Statistical description: Bernoulli trial

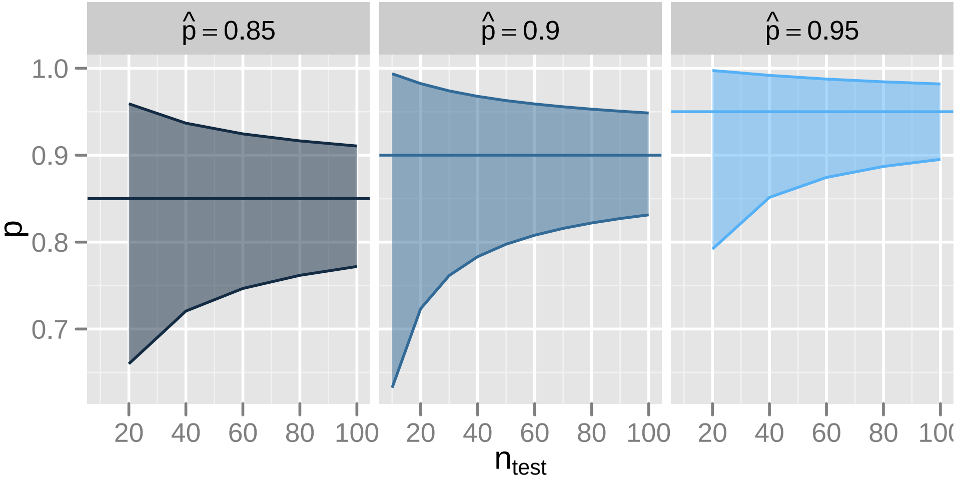
✔ ⇝ use binomial distribution

- Uncertainty on proportion: $var(\hat{p}) = \frac{p(1-p)}{n_{test}}$

✘ normal approximation appropriate only with $np \geq 5$ and $n(1-p) \geq 5$

⇝ Estimate necessary n_{test}

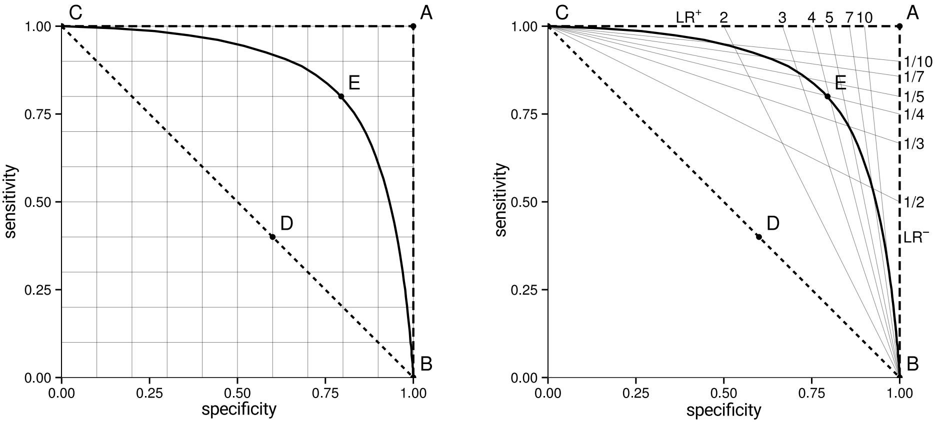
Sample size from Confidence Interval



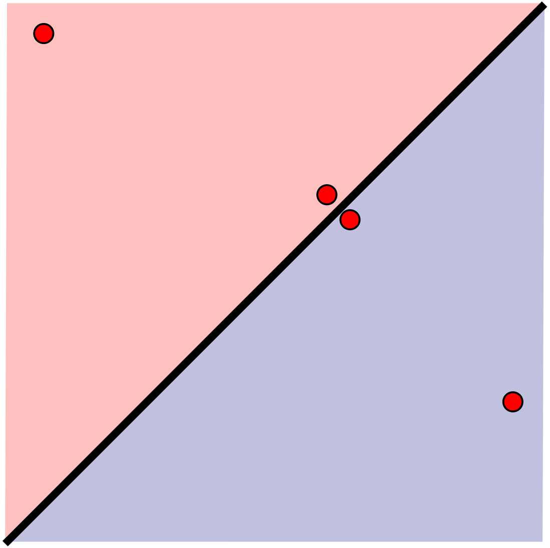
CHEMO METRIX

Beleites, C. *et al.*:Sample size planning for classification models. Anal Chim Acta, 2013, 760, 25-33

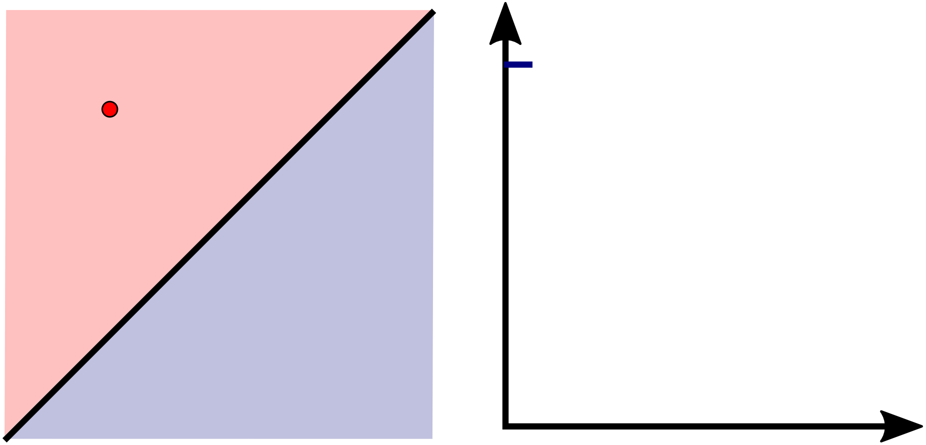
Receiver Operating Charcteristic/Specificity-Sensitivity-Diagram



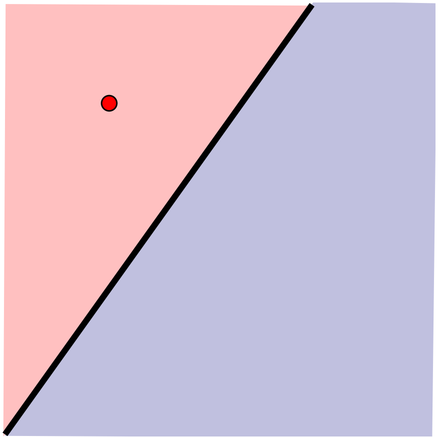
Proportions: Behaviour in Hyperparameter Optimization

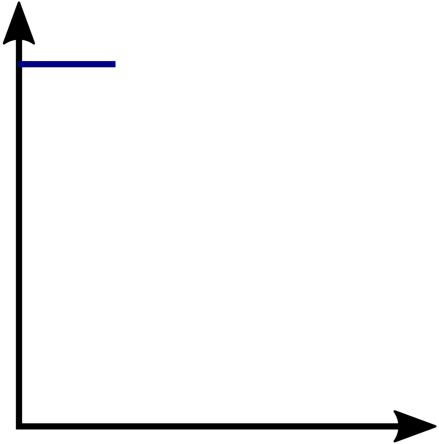


Proportions: Behaviour

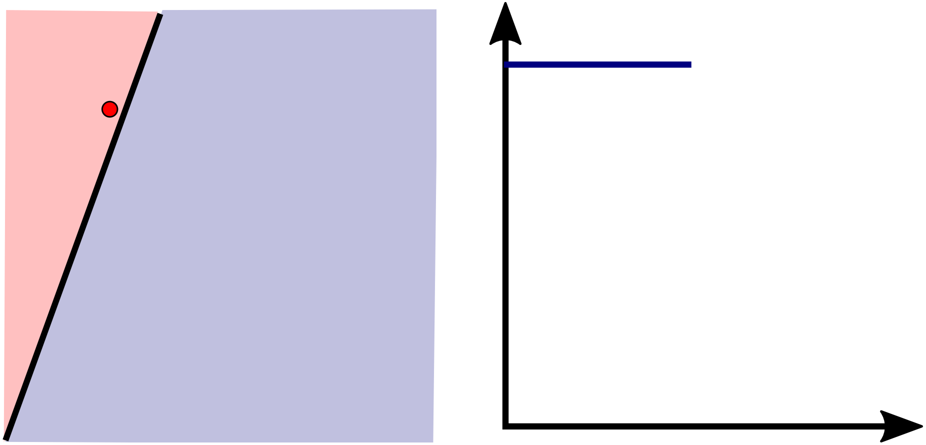


Proportions: Behaviour

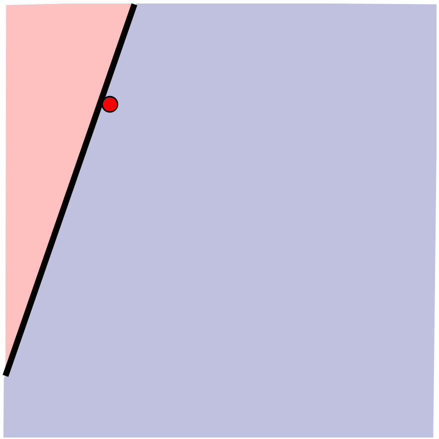


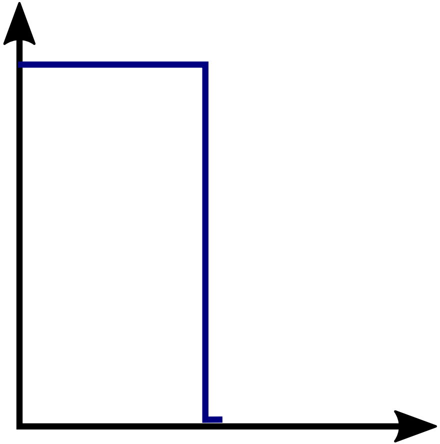


Proportions: Behaviour

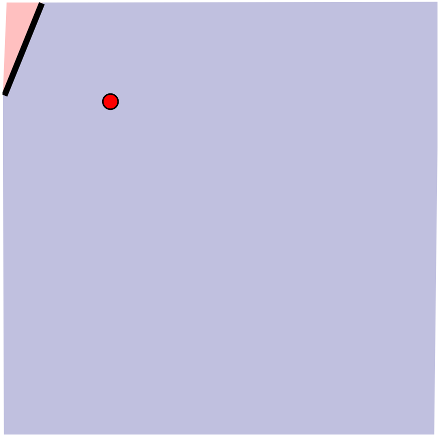


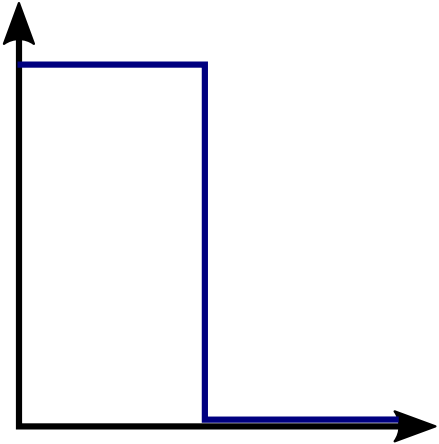
Proportions: Behaviour



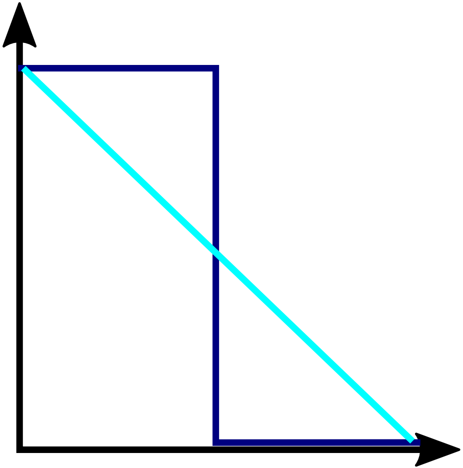


Proportions: Behaviour





(Strictly) Proper Scoring Rules



Wanted: Figure of merit that . . .

- . . . continuously penalizes closeness to class boundary
- . . . continuously reacts to changes in the model
- . . . slight deterioration ⇝ slight drop in measured performance
- . . . has exactly one optimum
- at the best classifier.

Regression: (Root) Mean Squared Error

- Loss behind e.g. Gaussian Least Squares
- penalizes large deviations
- $MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$
  $N \ldots$ number of cases
  $i \ldots$ case in question
- $MSE = bias^2 + variance$
- $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}$
  same scale as y

Brier's Score: Mean Squared Error for Classification

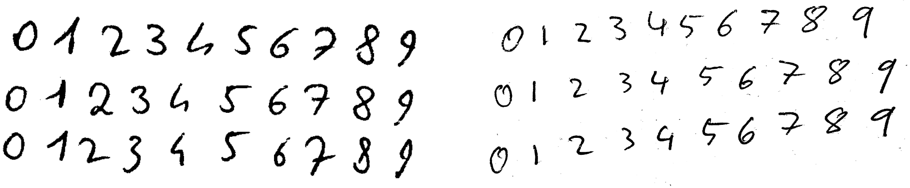
- Classifier that predicts class membership probability rather than labels
- Idea: of all cases where classifier predicts x % class membership, x % should belong to class in question
  a.k.a. *well calibrated prediction*
- Brier's score: $BS = \frac{1}{N} \sum_{i=1}^{N} (\hat{p}_i - p_i)^2$ or
  $BS = \frac{1}{N} \sum_{j=1}^{R} \sum_{i=1}^{N} (\hat{p}_i j - p_i j)^2$ (multiclass version) with
  N ... number of cases
  i ... case in question
  R ... number of classes
  j ... class in question
  p ... class membership, usually $\in \{0, 1\}$
  $\hat{p}$ ... predicted class membership $\in [0, 1]$

Summary Figures of Merit

- Many exist ⇝ choose relevant ones
- Usually, several figures of merit are needed for characterization
- Figures of merit are *measured* ⇝ subject to bias and variance like any other measurement
- Regression: figures of merit “well-behaved”,
  but no back-of-the-envelope variance guesstimates
- Classification: proportions easy to understand & widespread
  but have bad variance properties & discontinuous behaviour

⇝ use (strictly) proper scoring rules for optimization

Cases and Statistical Independence



- structure in data: clusters, spatial and/or temporal
- e.g. repeated measurements: repetitions are more similar to each other
- special case: time series

✔ Think hard about factors affecting independence

Model Testing: Measure the Model's Performance

Different kinds of test samples
⇝ different performance measures

Goodness of fit: *training samples*
⇝ residuals

Generalization error: statistically independent samples
resampling,
test set measured at same time as training set

Future performance: samples measured *after* training samples
dedicated test set for detection of drift

Resampling for Model Validation

✘ We don't have enough samples

- Training:
  - Model quality depends on ratio $n_{train} : d.f.$
  - Linear model: $5\,\text{samples}/(\text{variate} \cdot \text{class})$
  - ✔ We want to use all samples for training

Resampling for Model Validation

✘ We don't have enough samples

- Training:
  - Model quality depends on ratio $n_{train} : d.f.$
  - Linear model: 5 samples/(variate · class)
  - ✔ We want to use all samples for training

- Testing:
  - ✔ We want to know whether the model is stable
  - Quality of the performance measure depends on n_{test}
  - Width of 95 % confidence interval $\overset{!}{\leq}$ 10 % for $p = 90\,\%$: $n_{test} \geq 140$
  - ✔ We want to use all samples for testing

CHEMO METRIX

Validation Schemes: Overview

autoprediction
reclassification
resubstitution

Resampling Validation

set validation

leave one out

cross validation

iterated *k*-fold
repeated *k*-fold

out-of-bootstrap

.632+ bootstrap

.632 bootstrap

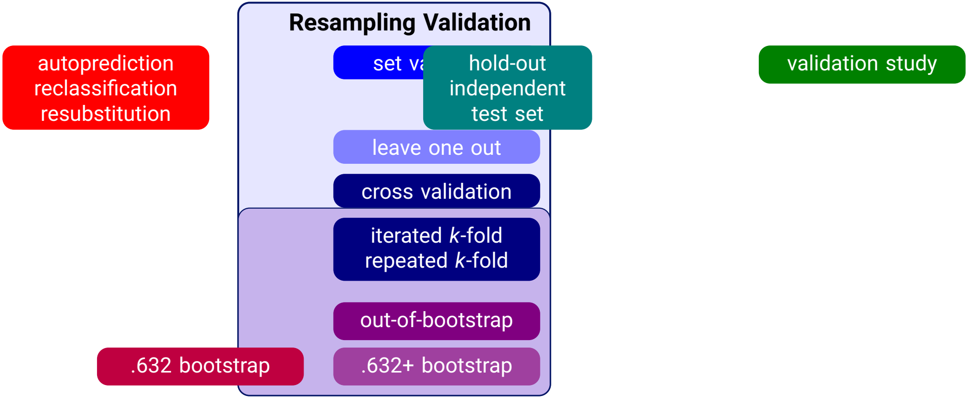
hold-out
independent
test set

validation study

R Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, 1995, Morgan Kaufmann, USA, 1995, 1137–1145.

Beleites, C. et al.: Variance reduction in estimating classification error using sparse datasets, Chemom Intell Lab Syst, 2005, 79, 91 - 100

Validation Schemes: Overview



R Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, 1995, Morgan Kaufmann, USA, 1995, 1137–1145.

Beleites, C. et al.: Variance reduction in estimating classification error using sparse datasets, Chemom Intell Lab Syst, 2005, 79, 91 - 100

Validation Schemes: Overview

autoprediction
reclassification
resubstitution

Resampling Validation

set validation

leave one out

cross validation

iterated *k*-fold
repeated *k*-fold

out-of-bootstrap

.632 bootstrap

.632+ bootstrap

hold-out
independent
test set

R Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, 1995, Morgan Kaufmann, USA, 1995, 1137–1145.

Beleites, C. et al.: Variance reduction in estimating classification error using sparse datasets, Chemom Intell Lab Syst, 2005, 79, 91 - 100

Validation Schemes: Overview

autoprediction
reclassification
resubstitution

Resampling Validation

set validation

leave one out

cross validation

iterated *k*-fold
repeated *k*-fold

out-of-bootstrap

.632 bootstrap

.632+ bootstrap

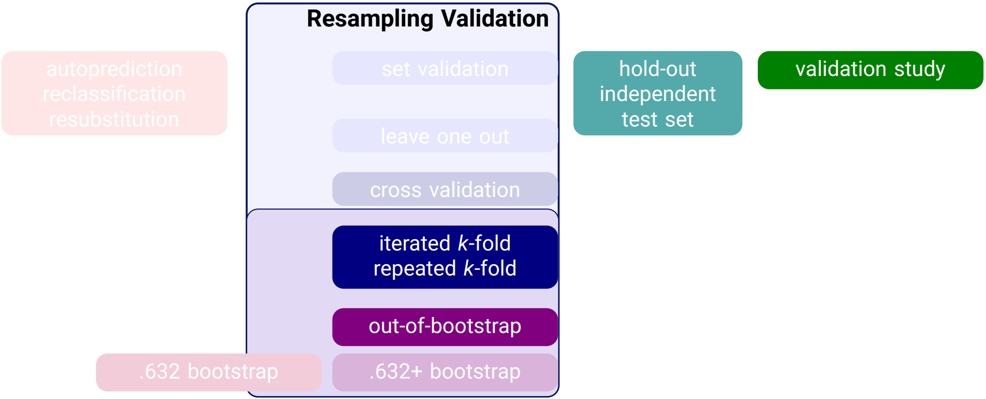
hold-out
independent
test set

validation study

R Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, 1995, Morgan Kaufmann, USA, 1995, 1137–1145.

Beleites, C. et al.: Variance reduction in estimating classification error using sparse datasets, Chemom Intell Lab Syst, 2005, 79, 91 - 100

Validation Schemes: Recommendations



R Kohavi: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, 1995, Morgan Kaufmann, USA, 1995, 1137–1145.

Beleites, C. et al.: Variance reduction in estimating classification error using sparse datasets, Chemom Intell Lab Syst, 2005, 79, 91 - 100

CHEMOMETRIX

Resampling vs. Validation Study

	Resampling	Validation Study
statistical properties		
bias	✔ pessimistic (low)	✔ unbiased
variance	$f(n)$	$f(n_{test})$
efficient use of cases	✔	✔
measure model stability	✔ iterated	✔/✘
measure drift	✘	✔ DoE
future case performance	✘	✔ DoE
out-of-spec cases	✘	✔ DoE
practical properties		
independence	⚠ splitting error prone	✔
effort	✔ computational	✘ experimental

Beleites, C. & Salzer, R.: Assessing and improving the stability of chemometric models in small sample size situations Anal Bioanal Chem, 2008, 390, 1261-1271
Esbensen, K. H. & Geladi, P.: Principles of Proper Validation: use and abuse of re-sampling for validation J Chemom, 2010, 24, 168-187

CHEMO METRIX

Resampling vs. Hold Out

	Resampling	Split off Hold Out Set
statistical properties		
bias	✔ pessimistic (low)	✔ unbiased
variance	✔ $f(n)$ lower	✘ $f(n_{test})$ large
efficient use of cases	✔	✘
measure model stability	✔ iterated	✘
measure drift	✘	✘
future case performance	✘	✘
out-of-spec cases	✘	✘(✔)
practical properties		
independence	⚠ splitting error prone	⚠ same as resampling
effort	✔ computational	✔ low

Beleites, C. & Salzer, R.: Assessing and improving the stability of chemometric models in small sample size situations Anal Bioanal Chem, 2008, 390, 1261-1271
Esbensen, K. H. & Geladi, P.: Principles of Proper Validation: use and abuse of re-sampling for validation J Chemom, 2010, 24, 168-187

CHEMO METRIX

Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
- Split before 1st step that involves multiple cases
- Additional independent validation for data-driven optimization/tuning/model selection
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
- Split before 1st step that involves multiple cases
- Additional independent validation for data-driven optimization/tuning/model selection
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
  day of measurement, before/after new calibration, . . .
- Split before 1st step that involves multiple cases
- Additional independent validation for data-driven optimization/tuning/model selection
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
  day of measurement, before/after new calibration, . . .
- Split before 1st step that involves multiple cases
  centering, PCA preprocessing, . . .
- Additional independent validation for data-driven optimization/tuning/model selection
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
  day of measurement, before/after new calibration, . . .
- Split before 1st step that involves multiple cases
  centering, PCA preprocessing, . . .
- Additional independent validation for data-driven optimization/tuning/model selection
  nested/double cross validation or train-validate-test ⇝ necessary case numbers HUGE
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

CHEMO METRIX

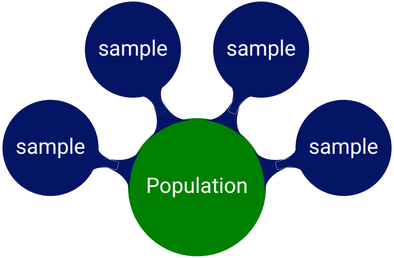
Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
  day of measurement, before/after new calibration, . . .
- Split before 1st step that involves multiple cases
  centering, PCA preprocessing, . . .
- Additional independent validation for data-driven optimization/tuning/model selection
  nested/double cross validation or train-validate-test ⇝ necessary case numbers HUGE
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
  cluster analysis to assign labels → OK for training cases
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

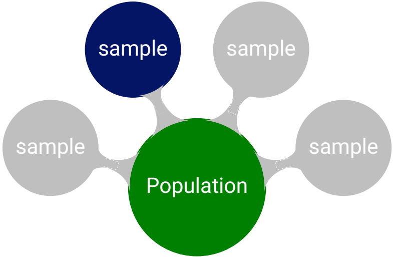
Validation Check List

- Randomize order of measurements
- Split at highest level in sample hierarchy/data structure
  patients, strains, cell lines,
  day of measurement, before/after new calibration, . . .
- Split before 1st step that involves multiple cases
  centering, PCA preprocessing, . . .
- Additional independent validation for data-driven optimization/tuning/model selection
  nested/double cross validation or train-validate-test ⇝ necessary case numbers HUGE
- Test cases: reference labels must be independent of cases (measurements, spectra, . . .)
  cluster analysis to assign labels → OK for training cases semi-supervised learning → OK for training cases
- Make sure labelling procedure does not distort difficulty for test cases
- Ensure correctness of code

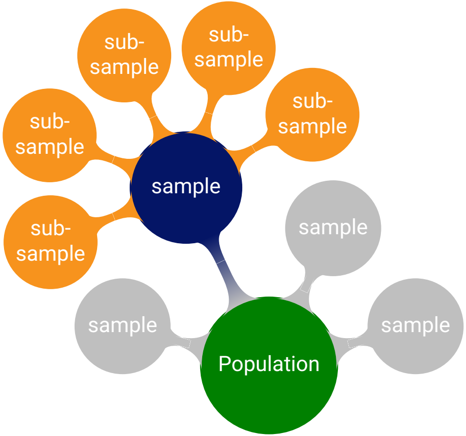
The Concept behind Resampling



The Concept behind Resampling



The Concept behind Resampling



- Subsamples are approximations of (more) real samples
- Subsample is perturbed version of the real sample

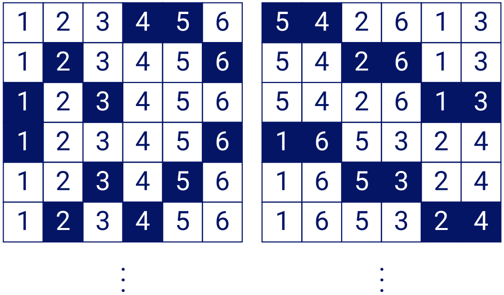
Cross Validation: Drawing without Replacement

1	2	3	**4**	**5**	6
1	**2**	3	4	5	**6**
1	2	**3**	4	5	6

5	**4**	2	6	1	3
5	4	**2**	**6**	1	3
5	4	2	6	**1**	**3**

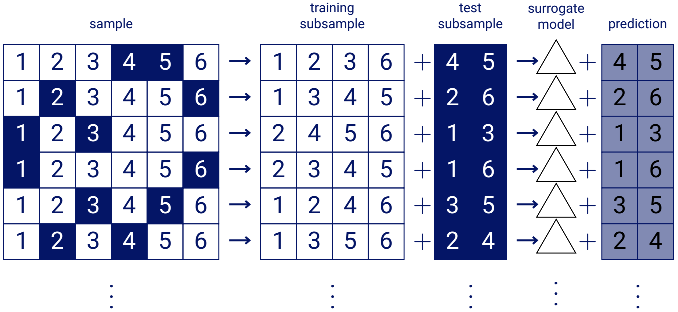
✔ Each case is left out exactly once

Cross Validation: Drawing without Replacement



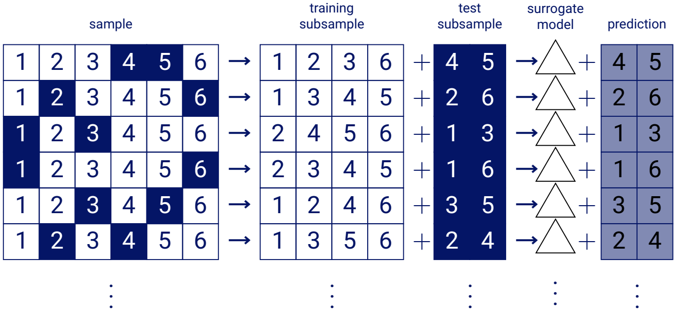
- ✔ Each case is left out exactly once per iteration
- Repetitions aka iterations possible with *k*-fold or leave-*n*-out cross validation
- ✘ Leave-one-out cannot be iterated

Resampling for Model Validation: Assumptions



- Surrogate model equals model of whole sample
- Surrogate models equal to each other
- All cases come from the same distribution

Resampling for Model Validation: Assumptions



- Surrogate model equals model of whole sample

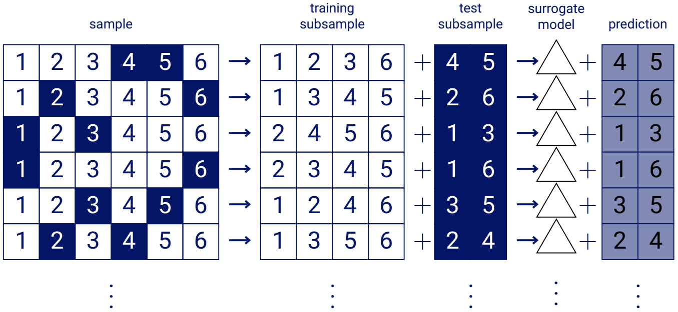
✘ Violation ⇝ pessimistic bias

- Surrogate models equal to each other

- All cases come from the same distribution

CHEMO METRIX

Resampling for Model Validation: Assumptions



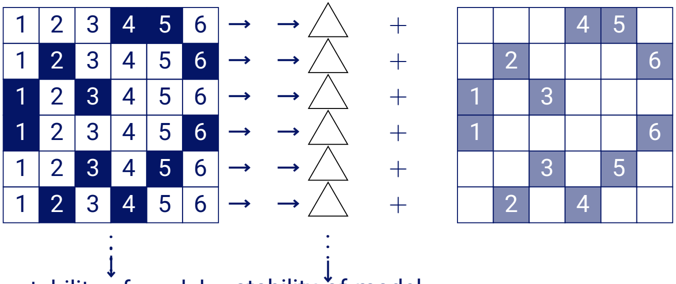
- Surrogate model equals model of whole sample
- ✘ Violation ⇝ pessimistic bias

- Surrogate models equal to each other
- ✘ Violation (instability) ⇝ higher variance

- All cases come from the same distribution

CHEMO METRIX

Model Stability



stability of model predictions

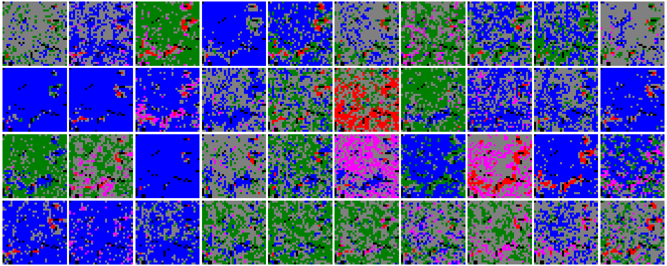
stability of model parameters

- Subsamples are perturbed versions of real sample

✔ Measure stability of model
  - Stability of model parameters
  - Stability of predictions

- Repetitions reduce variance due to instability of surrogate models.

Model Stability: 40× 8-fold cross validation



- FTIR images of tumour sections (normal, °II, °III, °IV)
- total: 150 images of 58 patients: 133 000 spectra
  smallest class: °II, 4 800 spectra (3 patients, 5 images)
- LDA after automatic selection of 8 spectral regions
- reject spectra with posterior probability <0.85

How many cases do we need?

. . . **to train a good classifier?**

– rules of thumb
linear model: $\frac{n}{p} \geq 3 - 5$ in each class

⇒ learning curve

. . . **to measure the model's performance?**

⇝ confidence intervals for test results

– Rules of thumb
100 test cases to estimate a proportion

– Regression ⇝ needs preliminary experiment

CHEMO METRIX

Validation: Questions

- Did you ask the right question?
- Or did you use a surrogate?
  - Is that surrogate appropriate?
  - What are the limits?
- Is your model set up correctly?
  - Is it really a classification problem?
  - one-class vs. discriminative?
  - open-world vs. closed-world?
  - correct scale of *y*? Other transformation better?
- Do you use the correct controls/base class or correct 0-point (center, origin) of regression?
- What happens with out-of-spec cases (unknown class? bad measurements?)

CHEMO METRIX

Validation: Questions

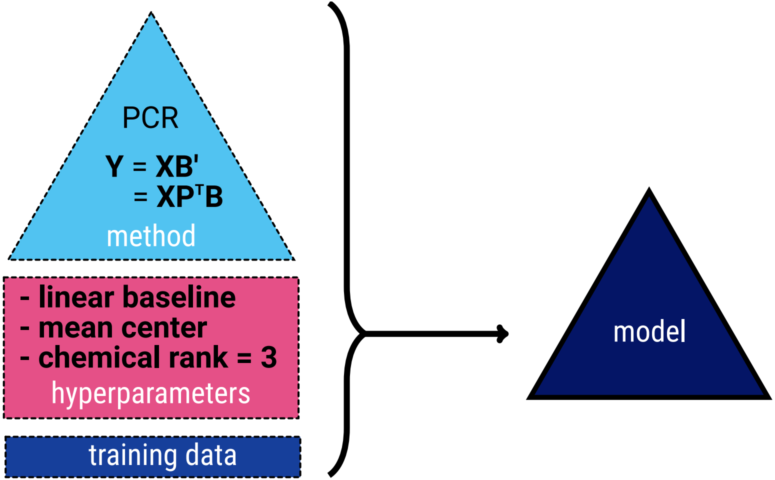
- Bias introduced by data acquisition procedure?
  - Labeling procedure with self-fulfilling prophecies (e.g. cluster analysis as basis for labeling, semi-supervised label generation)?
- What about borderline cases?
  - Do your *labeled* cases correctly represent them?
  - No exclusion of “difficult” cases in the reference labeling step?
- What other confounders could exist?
- What are the limits of your method?
- reading suggestions on reproducibility issues in medical research:
  - Buchen, L.: Cancer: Missing the mark. Nature, 2011, 471, 428 – 432
  - Begley, C. G. & Ellis, L. M.: Drug development: Raise standards for preclinical cancer research. Nature, 2012, 483, 531 – 533
  - Ioannidis, J. P. A.: Why Most Published Research Findings Are False, PLoS Med 2(8): e124
  - . . .

Ruggedness: Perturb Data

- How robust are the predictions?
- Which factors (confounders) have most influence?
- Perturb Data
  - Repeated cross validation:
    How do predictions vary if *a few* training cases are exchanged?
    ⇝ stability of predictions
  - Simulate instrument related distortions:
    Measure respective drop in performance

Beleites, C. & Salzer, R.: Assessing and improving the stability of chemometric models in small sample size situations Anal Bioanal Chem, 2008, 390, 1261-1271
Sattlecker, M. et al.: Assessment of robustness and transferability of classification models built for cancer diagnostics using Raman spectroscopy J Raman Spectrosc, 2010, 897-903

Hyperparameters



- available: PCR ($\mathbf{X}_{train}$, no_PCs, center)
- wanted: PCR_tuned ($\mathbf{X}_{train}$)

Data-driven Model Optimization

training data

Idea:

1. sensible range of hyperparameters
2. build covering this search space
3. measure performance
4. take the best

⇒ **Optimize predictive performance**

✔ Large variety of numerical optimizers available

exhaustive grid search, genetic optimizers, simulated annealing, . . .

Data-driven Model Optimization

training data | validation data

Idea:

1. sensible range of hyperparameters
2. build covering this search space
3. measure performance
4. take the best

⇒ **Optimize predictive performance**

✔ Large variety of numerical optimizers available

exhaustive grid search, genetic optimizers, simulated annealing, . . .

Data-driven Model Optimization

training data | validation data

Idea:

1. sensible range of hyperparameters
2. build covering this search space
3. measure performance
4. take the best

⇒ **Optimize predictive performance**

✔ Large variety of numerical optimizers available

exhaustive grid search, genetic optimizers, simulated annealing, . . .

✘ Careful: valdiation data enters model building process

⇒ need another independent set to validate the *final* model

Data-driven Model Optimization

training data | validation data | test data

Idea:

1. sensible range of hyperparameters
2. build covering this search space
3. measure performance
4. take the best

⇒ **Optimize predictive performance**

✔ Large variety of numerical optimizers available

exhaustive grid search, genetic optimizers, simulated annealing, . . .

✘ Careful: valdiation data enters model building process

⇒ need another independent set to validate the *final* model

Data-driven Model Optimization

training data

validation data

test data

- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with validation set
- validate chosen model with test set

Data-driven Model Optimization

- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with validation set
- validate chosen model with test set

Data-driven Model Optimization

training data | optimization data | verification data

- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with ~~validation set~~ optimization aka development set
- validate chosen model with ~~test set~~ final verification set

Data-driven Model Optimization

- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with ~~validation set~~ optimization aka development set
- validate chosen model with ~~test set~~ final verification set

✔ resampling version: nested/double cross validation

Data-driven Model Optimization

training data | optimization data | verification data

- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with ~~validation set~~ optimization aka development set
- validate chosen model with ~~test set~~ final verification set

✔ resampling version: nested/double cross validation

✔ `train (X, hyperparameters)` vs. `tuned_train (X)`
  - tuned training function: additional internal split for tuning

Data-driven Model Optimization

training data

test data

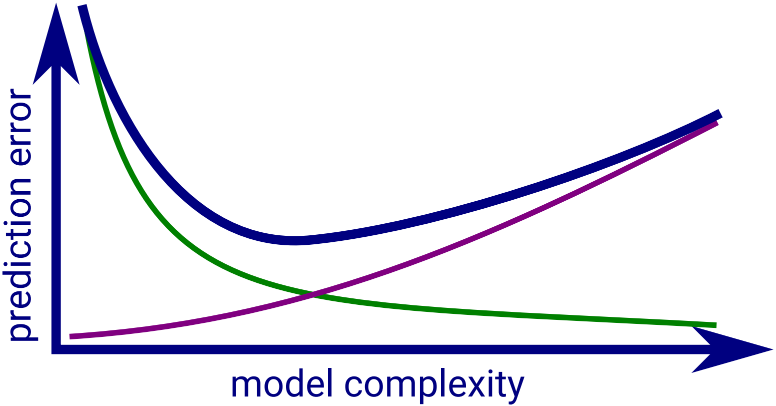
- fit normal parameters (coefficients) with *training* set
- fit hyperparameters with ~~validation set~~ optimization aka development set
- validate chosen model with ~~test set~~ final verification set

✔ resampling version: nested/double cross validation

✔ `train (X, hyperparameters)` vs. `tuned_train (X)`
  - tuned training function: additional internal split for tuning
  - ✔ treat `tuned_train (X)` like any other training function

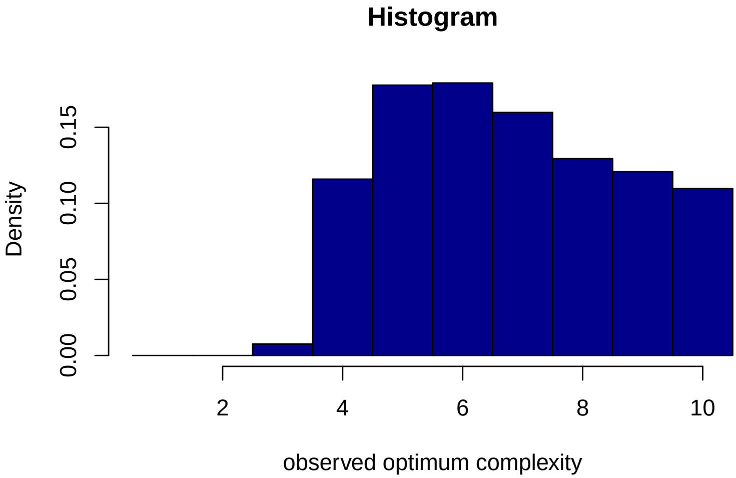
CHEMO METRIX

Grid search: Stability of Solution



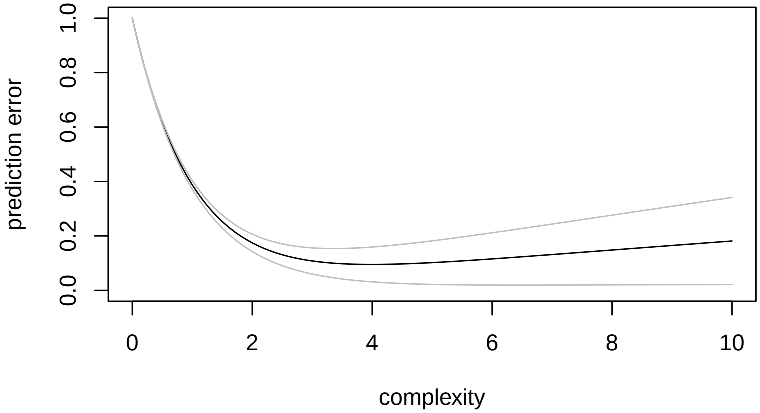
Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Grid search: Stability of Solution



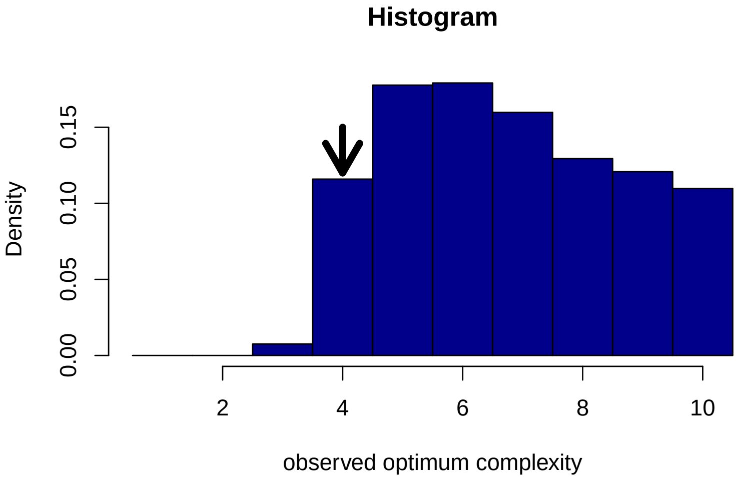
Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Grid search: Stability of Solution



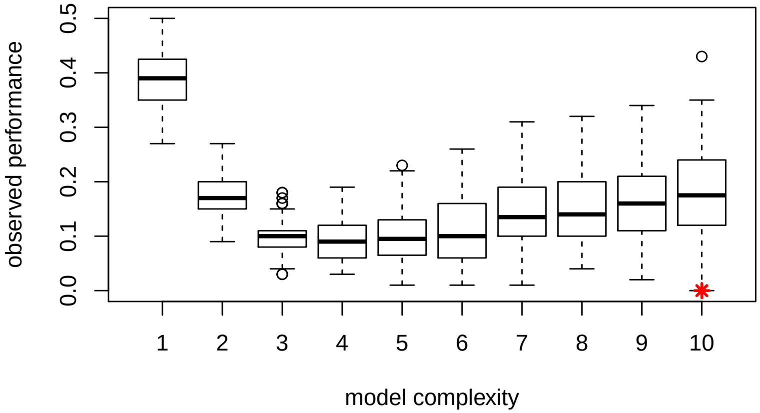
Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Grid search: Stability of Solution



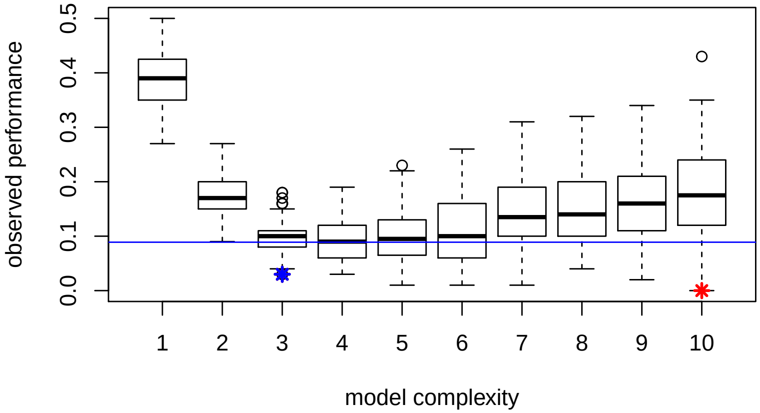
Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Grid search: Stability of Solution



Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Grid search: Stability of Solution



Hastie, Tibshirani, Friedman: The Elements of Statistical Learning, Springer

Summary: Validation

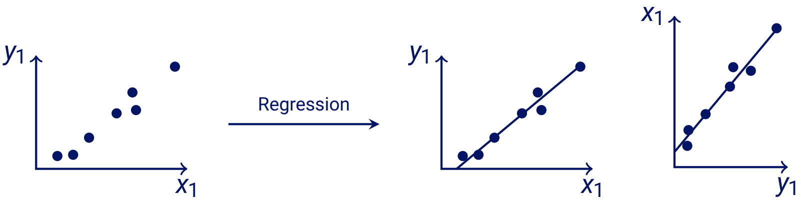
- ✔ Think hard about your data, model, and application!
- ✔ Sample size planning: calculate from required precision of validation results possibly from preliminary experiment
- At some point, validation studies are needed.
  Before that, use repeated cross validation or out-of-bootstrap.
- ✔ Determine independent splitting
- ✔ Check stability of predictions and – if possible – model parameters
- ✘ Resampling cannot detect drift
- ✘ Hold-out is inefficient and prone to the same errors as resampling!

Summary: Data-driven model Optimization

- ✔ Needs internal performance estimate plus outer independent validation
- ✘ ⇝ large sample size required
- ✔ wrap optimization in `tuned_model` function
- ✔ validate output of `tuned_model` like any other model training function
- ✔ Check stability of optimization
- ✔ Use 1-sd-rule to guard against overfitting
- ✔ Class membership probability predicted: MSE (Brier's Score) has low variance and is proper scoring rule
  ⇝ suitable for optimization

CHEMO METRIX

Regression



Ordinary Regression

$\mathbf{Y}^{(n \times m)} = \mathbf{X}^{(n \times p)} \mathbf{B}^{(p \times m)}$

- assume error on y (I)
- ✔ causality: $I = f(c)$
- ✔ efficient estimation of calibration line parameters

Inverse Regression

- assume error on x (c)
- ✔ prediction: $c = f(I)$
- ✔ efficient estimation of y
- ✘ needs $p \leq m$

Univariate Linear Regression

$$\mathbf{Y}^{(n \times m)} = \mathbf{X}^{(n \times p)} \mathbf{B}^{(p \times m)}$$

y	**x**
1	2
2	3
3	4
4	5
5	6

Multivariate Linear Regression

$$\mathbf{Y}^{(n \times m)} = \mathbf{X}^{(n \times p)} \mathbf{B}^{(p \times m)}$$

$\mathbf{y}$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
1	2	7	3
2	3	5	5
3	4	3	-2
4	5	1	7
5	6	-1	0

Linear Models: Polynomial Features

$$\mathbf{Y}^{(n \times m)} = \mathbf{X}^{(n \times p)} \mathbf{B}^{(p \times m)}$$

y	x
1	2
2	3
3	4
4	5
5	6

Linear Models: Polynomial Features

$$\mathbf{Y}^{(n \times m)} = \mathbf{X}^{(n \times p)} \mathbf{B}^{(p \times m)}$$

y	x^0	x^1	x^2
1	1	2	4
2	1	3	9
3	1	4	16
4	1	5	25
5	1	6	36

Questions?

Please contact me (**Claudia.Beleites@chemometrix.gmbh**) if you

- have questions, or
- want to reuse these slides.