- **Modellauswahl**
  - **Auf Basis Validation/ Cross-Validation der Testdaten**
    - **Validation mit Testdaten**
    - **Leave one out cross-validation (LOOC)**
    - **K-Fold**
    - **Walk-forward-selection**
  - **Informationskriterien auf Basis Trainingsdaten**
    - **C**
    - **AIC**
    - **BIC**
    - R^2

Modellauswahl

Die Themengebiete der Veranstaltung verknüpfen Modelle des „Machine Learning“ mit energiewirtschaftlichen Fragestellungen

Energiedatenanalyse- Datamining

- **Übergeordnete Themen**
  - Überblick Machine Learning
  - Deskriptive Datenanalyse
  - Modellauswahl
- **Klassische Regressions-verfahren**
  - Multivariate Lineare Regression
  - PCA-PLS-Regression
  - Polynome, Splines, lokale Regression
- **Regression mit KNN**
  - Neuronale Netze Feedforward
  - Rekurrente Neuronale Netze (LSTM)
  - Convolutionale Neuronale Netze CNN
  - Temporal fusion Transformer
- **Klassifizierung**
  - K-nearest, Bayses Ansatz
  - Rekurrente/ Convolutionale KNN
- **Clustering/ Datenreduktion**
  - Principle Component Analyse (PCA)
  - Autoencoder
  - K-Means, Hierarchical Clustering

Zielsetzung der heutigen Vorlesung:
Einführung in die Verwendung der linearen Regression

Thema	Überblick über Validierungs- und Auswahlwahlverfahren

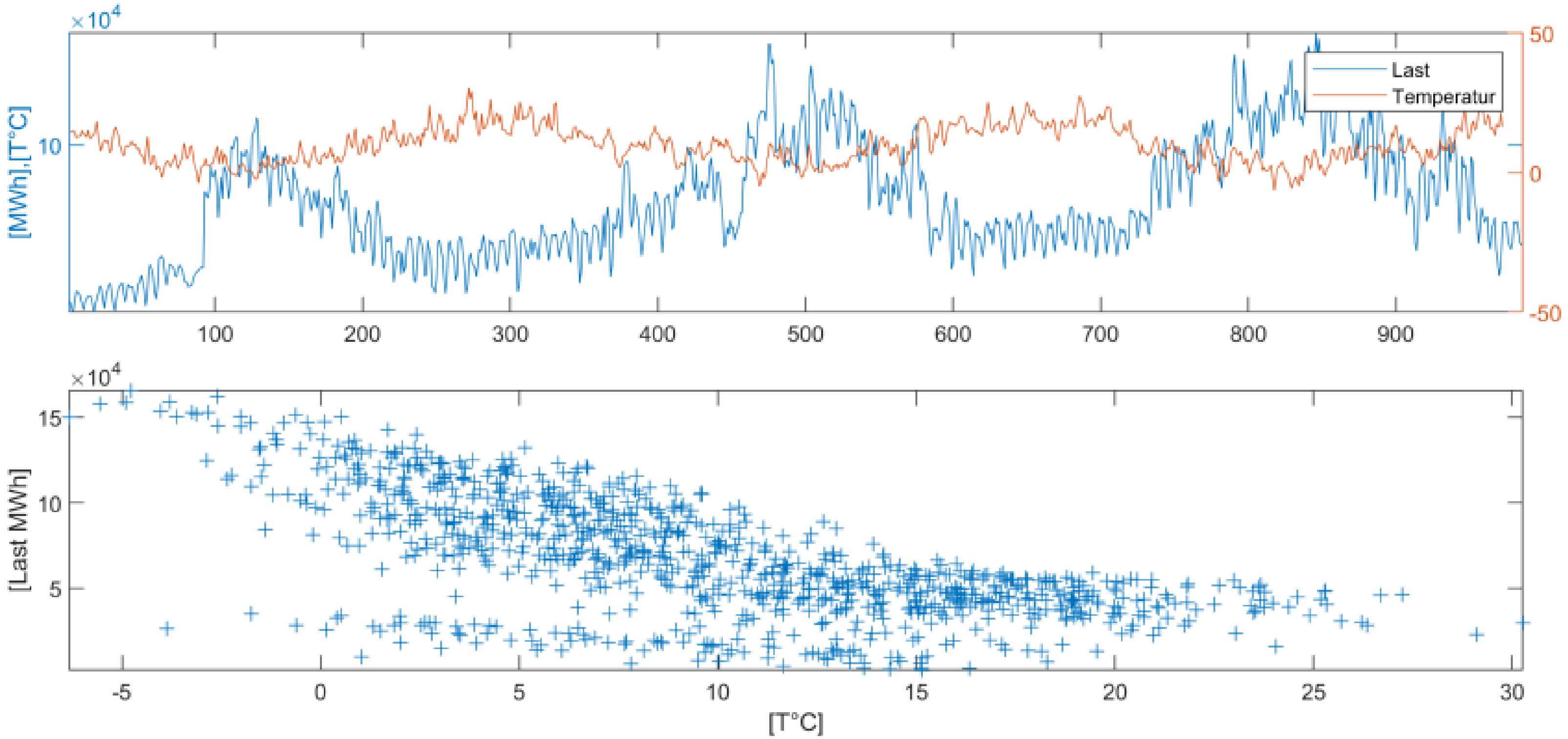
	Aufbau der heutigen Vorlesung:	**Lernziel:**
1	Bedeutung des Bias-Variance Trade off	Notwendigkeit des Einsatzes von Validierungstechniken
2	Validation/ Cross-Validation-Techniken	Anwendung von Techniken zur Modellbeurteilung
3	Best Subset selection Stepwise selection	Prozess zur Modellauswahl

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

1	Bedeutung des Bias-Variance Trade off
2	Generierung von Schätzern für den Test-RMSE
3	Best Subset selection Stepwise selection

Literaturempfehlung und Quellennachweise der Abbildungen: James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics), Springer New York. Kindle-Version.

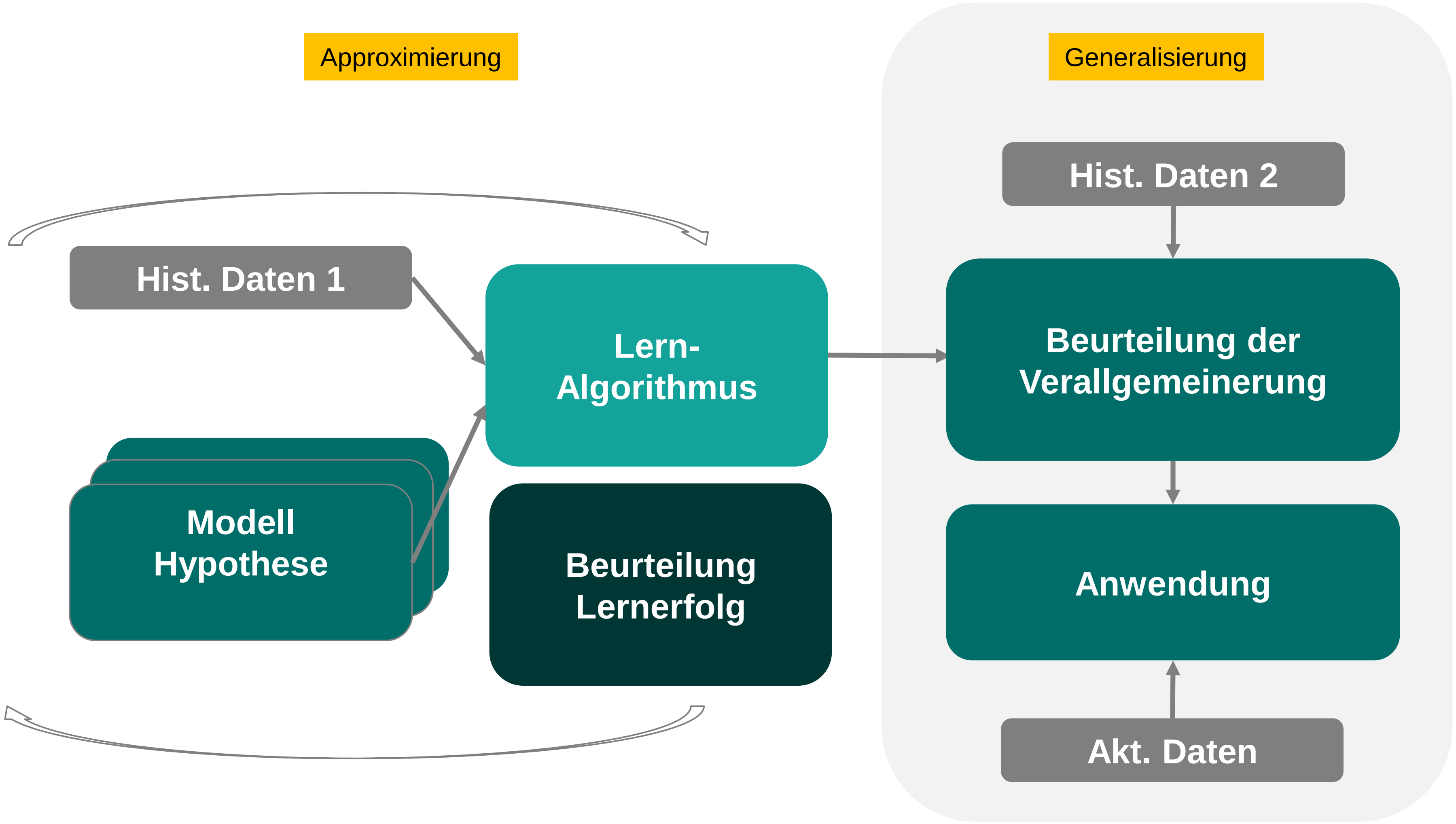
Essentielle Fragen im Rahmen der Modellkalibrierung



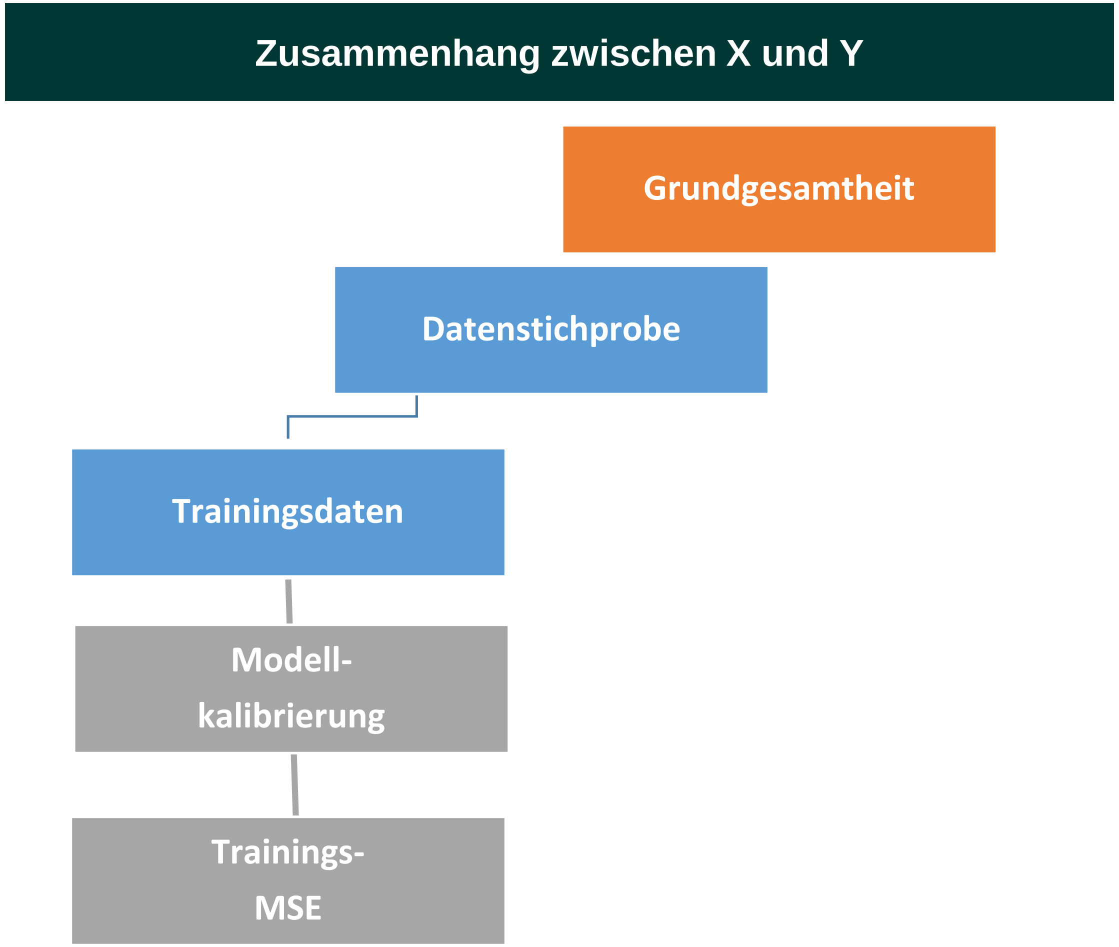
- Ziel 1: Auffinden eines Modells zur Approximation des Zusammenhangs der vorliegenden Daten → **Approximation**

- Ziel 2: Sicherstellung, dass das Modell den Zusammenhang bislang unbekannter Daten abbildet → **Generalisierung**

Eine wesentliche Aufgabe des Anwenders ist die richtige Modellauswahl



Die Approximation erfolgt durch Minimierung eines Zielkriteriums ausgewertet auf einer Datenstichprobe



Ausgangslage:

- der reale funktionale Zusammenhang zwischen Y und X ist unbekannt.

Vorgehensweise:

1. Bildung Modellhypothese,
2. Bildung Datenstichprobe,
3. Modellparametrierung.

- Zielkriterium zur Anpassung der Parameter:
  - Mean Squared Error MSE ($MSE = S_{\hat{e}\hat{e}}$)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

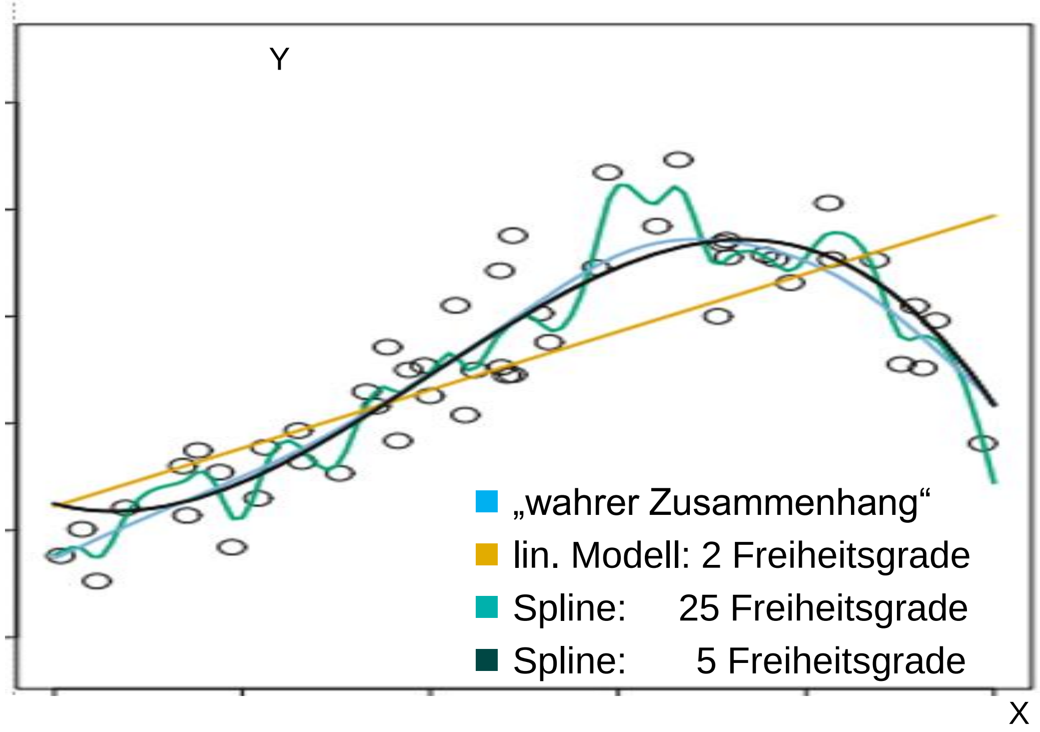
In Sample Fehler = Training MSE

$\hat{y}$ ist der modellierte Wert bzw. Schätzer für y.

Der „In-Sample-Fehler“ misst die Abweichung des Modell auf den Daten anhand derer die Modellparameter optimal eingestellt werden.

Problem: durch hinzufügen weiterer Parameter wird der gesuchte Zusammenhang beliebig approximiert

Zusammenhang zwischen X und Y



- Modelle unterscheiden sich u.a. durch die Anzahl der verwendeten **Parameter.**
- Ein Modellparameter repräsentiert einen Freiheitsgrad.
- Generell gilt:

$$Freiheitsgrade \uparrow \;\rightarrow\; MSE(Training) \downarrow$$

- Konsequenz für die Minimierung der Zielfunktion:
  - Verringerung Abweichung von $\hat{Y}\ zu\ Y$
  - Der MSE sinkt „zwangsläufig“ durch Vergrößerung der Regressorbasis p

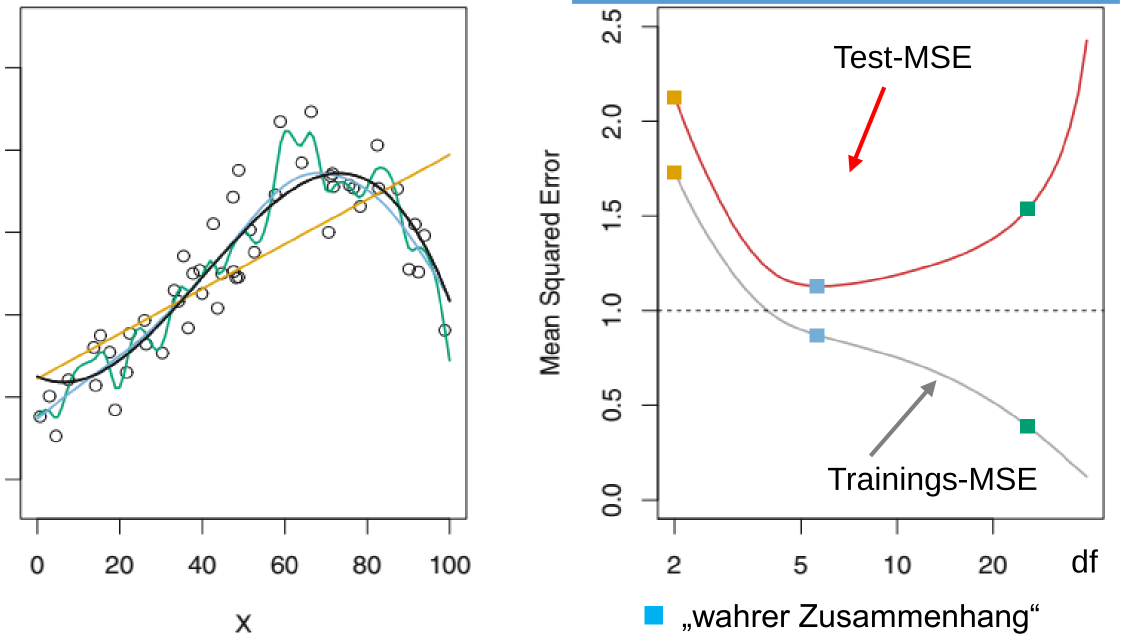
Frage: wie schlägt sich so ein Modell bei „neuen“ bislang unbekannten Daten?

Beobachtung: MSE der Trainingsdaten und MSE der Testdaten verhalten ab einen gewissen Punkt gegensätzlich -> Zustand Overfitting

Verlauf des MSE bei Trainings- und Testdaten

MSE Trainingsdaten := Modellparametrierung

MSE Testdaten := Fit unbekannter Daten



----- MSE des unsystematischen Fehlers (S_{ee})

- „wahrer Zusammenhang"
- lin. Modell: 2 Freiheitsgrade
- Spline: 25 Freiheitsgrade

Wirkungsweise Erhöhung der Freiheitsgrade:

- stetige Verringerung des Trainings-MSE
- Test-MSE verläuft i.d.R. u-förmig mit Minimum oberhalb des unsystematischen Fehlers (---)
- Test-MSE approximiert den „$Out\ of\ Sample\ Fehler$":

$$Out\ of\ Sample\ Fehler = E\left[\left(\hat{f}(x) - f(x)\right)^2\right]$$

Konsequenz:

- Ab einer bestimmten Modellflexibilität setzt ein „Overfitting" ein.
- Overfitting: Verschlechterung der Modellanpassung ggü. einfacheren Modellen.
- Begründung: es wird ein Teil des „unsystematischen" Fehlers in dem Modell „erklärt".

Quelle: James, Gareth, An Introduction to Statistical Learning

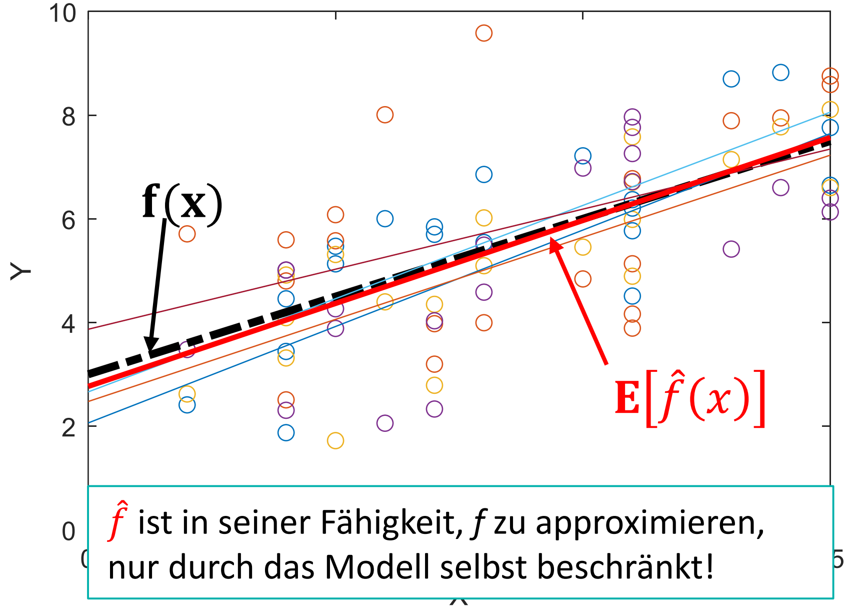
df = Anzahl der Freiheitsgrade (Modellparameter)

Bias und Varianz sind Gütekriterien für die Modellanpassung

BIAS (Verzerrung)

Der Bias misst die Abweichung zwischen **gemittelten** parametrierten **Modell $\mathbf{E}[\hat{f}(x)]$ und dem wahren Zusammenhang** f(x)

$$\mathbf{Bias} = \mathbf{E}\left[\hat{f}(x)\right] - \mathbf{f}(\mathbf{x})$$

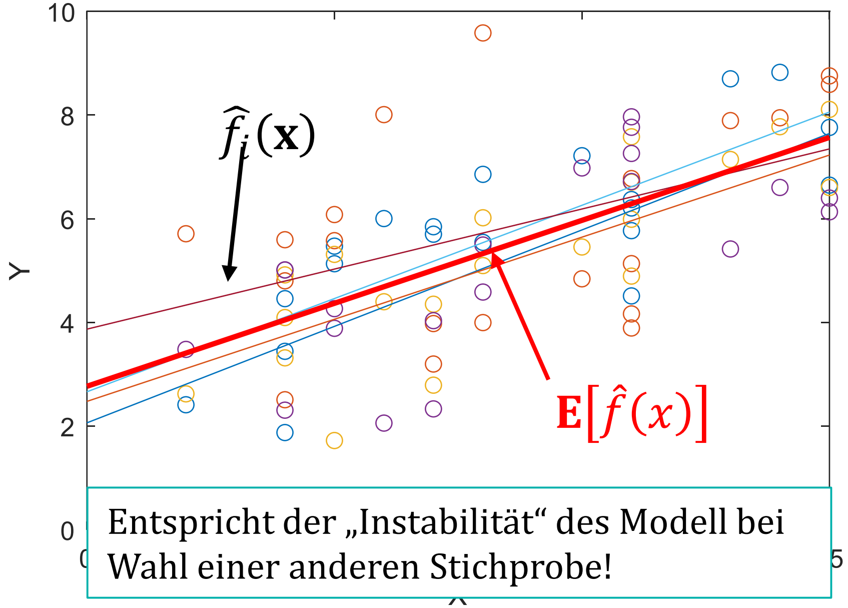


$\hat{f}$ ist in seiner Fähigkeit, f zu approximieren, nur durch das Modell selbst beschränkt!

Varianz

Die Varianz misst die **Variation** des parametrierten Modells $\hat{f}(x)$ um das **gemittelte** Modell $\mathbf{E}[\hat{f}(x)]$

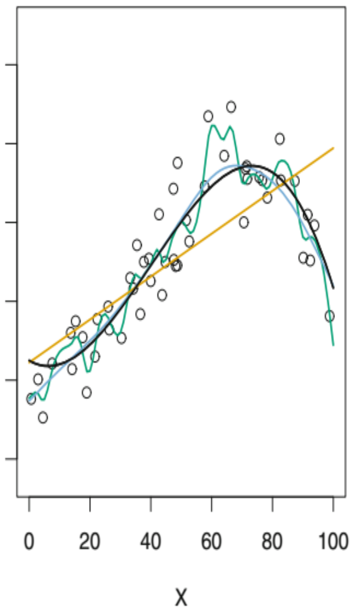
$$\mathbf{Varianz} = \boldsymbol{E}\left[\left(\hat{f}(\mathbf{x}) - \mathbf{E}\left[\hat{f}(x)\right]\right)^{2}\right]$$

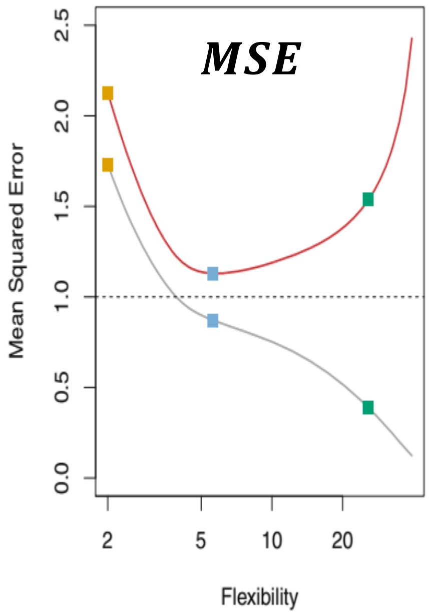


Entspricht der „Instabilität" des Modell bei Wahl einer anderen Stichprobe!

Der Bias-Varianz Trade off hängt stark vom Problem und der gewählter Methodik ab

$$\underbrace{E[MSE]}_{\text{Erwarteter Test-MSE}} = \underbrace{\mathbf{E}[\hat{f}(\mathbf{x})^2] - \left[\mathbf{E}[\hat{f}(x)]\right]^2}_{\text{Varianz der Methodik}} + \underbrace{\left(\mathbf{E}[\hat{f}(x)] - \mathbf{f}(\mathbf{x})\right)^2}_{BIAS^2} + \underbrace{Var(\epsilon)}_{\text{Unsystematischer Fehler}}$$





<u>Bias-Varianz Trade off:</u>

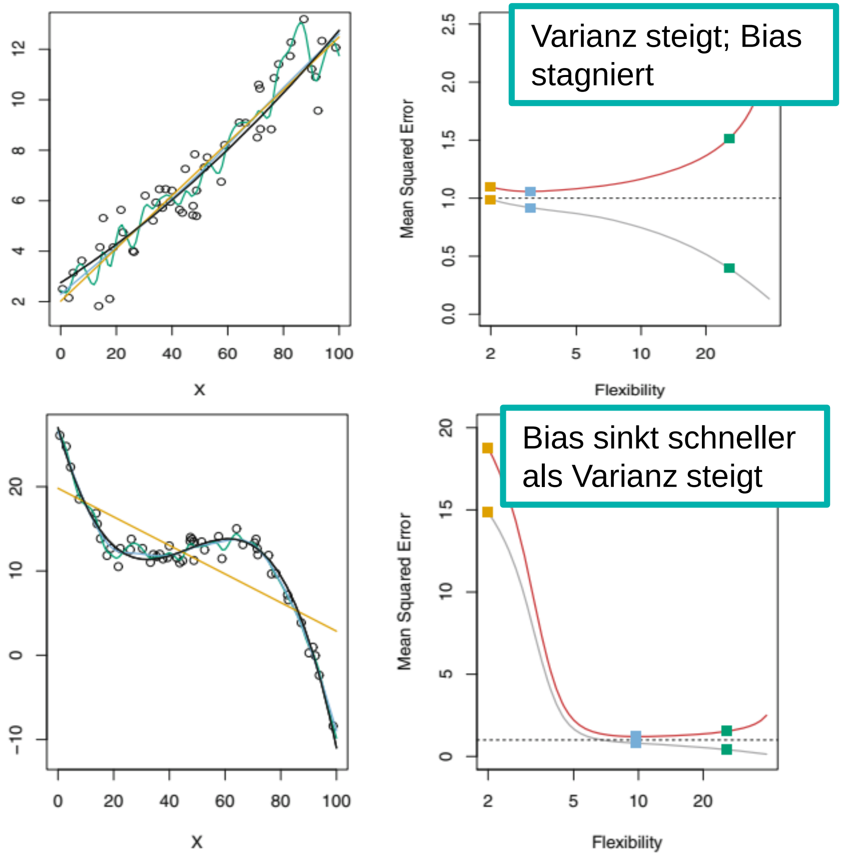
- Flexible Funktionen sind generell besser in der Lage den Bias zu verringern als unflexible Funktionen.
- Grundsätzlich haben flexiblere Methoden zur Abbildung der Funktion f eine höhere Varianz.

Nebenrechnung: $MSE = E[(\hat{f}(\mathbf{x}) - f(x))^2] = \mathbf{E}[\hat{f}(\mathbf{x})^2] + E[f(x)^2] - 2\mathbf{E}[\hat{f}(\mathbf{x}) \cdot f(\mathbf{x})] = \mathbf{E}[\hat{f}(\mathbf{x})^2] - \left[E[\hat{f}(x)]\right]^2 + \left[E[\hat{f}(x)]\right]^2 + E[f(x)^2] - 2\mathbf{E}[\hat{f}(\mathbf{x}) \cdot f(x)]$

$E[(\hat{f}(\mathbf{x}) - \mathbf{E}[\hat{f}(x)])^2] = \mathbf{E}[\hat{f}(\mathbf{x})^2] + \left[E[\hat{f}(x)]\right]^2 - 2\mathbf{E}[\hat{f}(\mathbf{x}) \cdot \mathbf{E}[\hat{f}(x)]] = \mathbf{E}[\hat{f}(\mathbf{x})^2] - \left[E[\hat{f}(x)]\right]^2$

Der Trainings-MSE und der Test-MSE verhalten ab einen gewissen Punkt gegensätzlich -> Zustand Overfitting II

Grau MSE der Trainingsdaten
Rot MSE der Testdaten

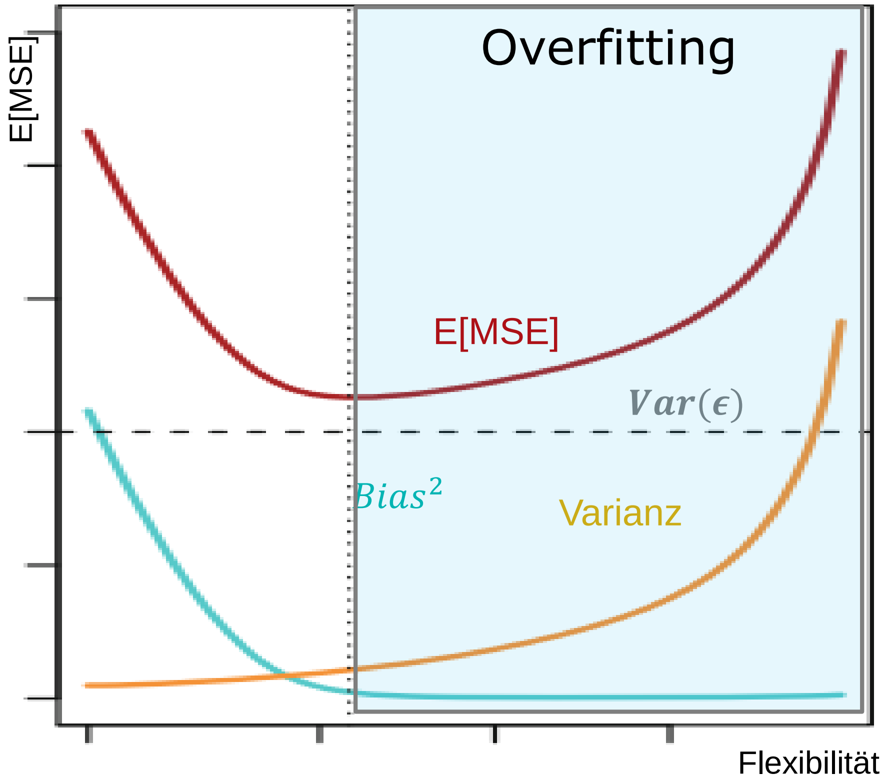


- Die Änderungsrate von Bias und Varianz bestimmt die Veränderung des erwarteten MSE.
- Wenn wir die Flexibilität einer Klasse von Methoden erhöhen, neigt der Bias dazu, zunächst schneller zu sinken als die Varianz steigt.
- Infolgedessen sinkt der erwartete MSE (approximiert durch den Test-MSE).
- Allerdings hat die Erhöhung der Flexibilität ab einen bestimmten Punkt nur noch geringen Einfluss auf den Bias.
- Die Varianz erhöht sich ab diesem Punkt deutlich.

Quelle: James, Gareth, An Introduction to Statistical Learning

Gesucht wird das Minimum des Test MSE

Zusammenhang zwischen Bias, Varianz und E[MSE]



<u>Erkenntnis:</u>

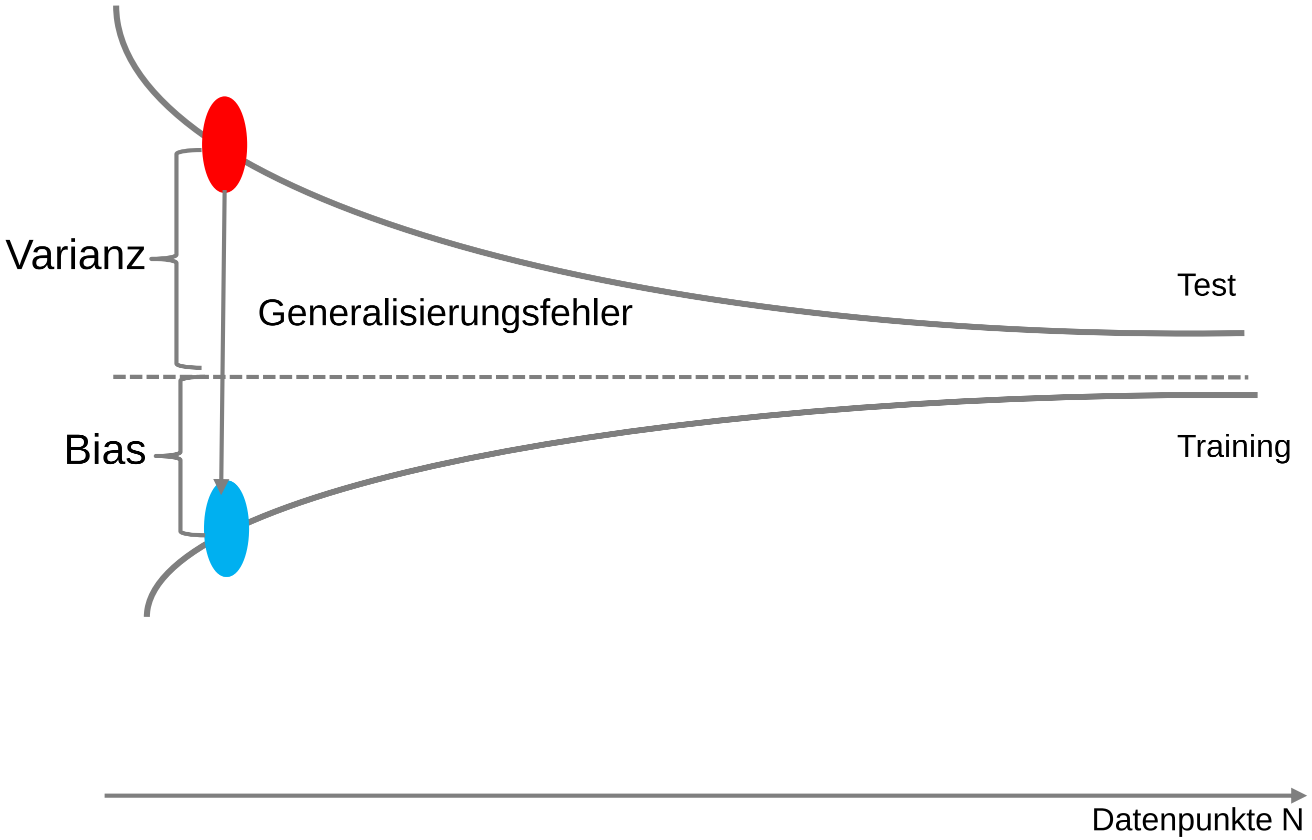
- Der Erwartungswert des MSE kann nicht unterhalb des systematischen Fehlers liegen.
- Bei höherer Flexibilität wird dieser zunächst positiv und dann negativ (Overfitting) beeinflusst.

<u>Zielsetzung:</u>

- Auffinden einer Funktion mit geringer Varianz und geringem Bias.
- In Realität ist der wahre funktionale Zusammenhang und damit auch die Varianz von f unbekannt, somit auch der <u>Erwartungswert</u> des Test-MSE.

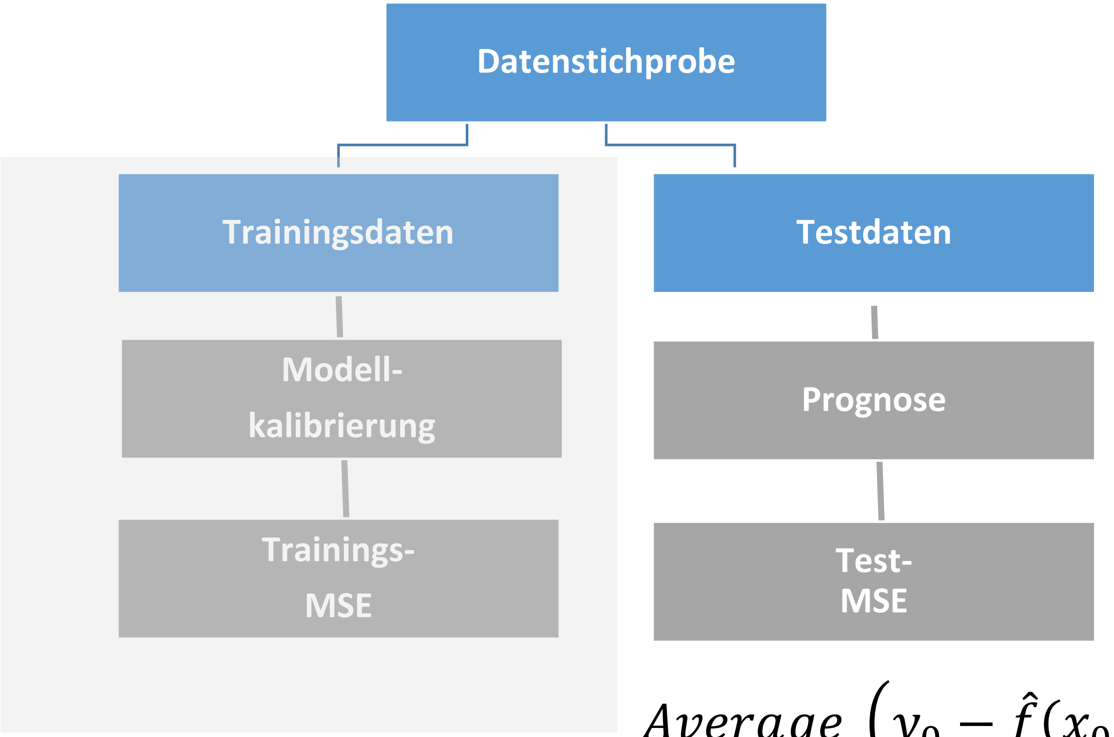
Quelle: James, Gareth, An Introduction to Statistical Learning

Der Stichprobenumfang hat einen bedeutenden Einfluss auf die Varianz



Für den Einsatz des Modells ist nicht der Trainings-MSE, sondern die nachträgliche „Güte“ bei „unbekannten Daten“ (Testdaten) entscheidend

Zusammenhang zwischen X und Y



$$Average\left(y_0 - \hat{f}(x_0)\right)^2$$

Erkenntnis:

- Für Modelleinsatz ist **nicht** entscheidend, wie gut das Modell die Daten wiedergibt, anhand dessen es kalibriert wurde! (Trainings-MSE)
- Entscheidend ist, wie gut das Modell mit „unbekannten“ Daten funktioniert (Test-MSE)
- Der Trainings-MSE ist kein guter Schätzer für den Test-MSE

Konsequenz:

- Der Trainings-MSE kann kein Kriterium für die Modellwahl darstellen.

Ziel bei Modellauswahl:

- **Modellwahl** anhand des geringsten ***Test-MSE***

1	Bedeutung des Bias-Variance Trade off
2	Samplingtechniken
3	Best Subset selection Stepwise selection

Literaturempfehlung und Quellennachweise der Abbildungen: James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics), Springer New York. Kindle-Version.

Für die Schätzung des MSE der Testdaten gibt es zwei mögliche Vorgehensweise

Modellauswahl

Indirekte Schätzung des erwarteten MSE	**Direkte Schätzung des erwarteten MSE**
Informationskriterien auf Basis Trainingsdaten	**Auf Basis Validation/ Cross-Validation der Testdaten**
Der Trainings-MSE wird mathematisch geeignet angepasst	Vom Trainingsdatensatz werden einzelne Datenpunkte der Parametrierung vorenthalten (Testdaten), um nach erfolgter Schätzung die geschätzte Funktion auf diese anzuwenden. Test-MSE

Die Modellauswahl kann durch direkter Betrachtung des Test-MSE oder Approximation des Test-MSE erfolgen

Modellauswahl

- **Auf Basis Validation/ Cross-Validation der Testdaten**
  - Validation mit Testdaten
  - Leave one out cross-validation (LOOC)
  - K-Fach
  - Walk-forward-selection
- **Informationskriterien auf Basis Trainingsdaten**
  - C
  - AIC
  - BIC
  - R^2

Der Einsatz von Informationskriterien kann bei bestimmter zugrunde liegender Verteilung hinreichende Rückschlüsse bieten

- Zielsetzung von Informationskriterien
  - Auffinden des optimalen Fits bzgl. Bias und Variance anhand eines **Informationskriterium**.
    - Als Anpassungsgüte wird i.d.R. die **Trainings- MSE oder RSS** verwendet (RSS=MSE*n)
    - Ein Strafterm für die Anzahl der Parameter d berücksichtigt, um Overfitting zu verhindern.
  - Es wird ein möglichst geringer Wert des Informationskriteriums angestrebt

Informationskriterien	Mathematische Formulierung
C_p	$C_p = \text{MSE} + 2d\hat{\sigma}^2/\text{n}$
Akaide Informationskriterium	$AIC = \frac{1}{\hat{\sigma}^2}(\text{MSE} + 2d\hat{\sigma}^2/\text{n})$
Bayessches Informationskriterium	$BIC = MSE + \log(n)d\hat{\sigma}^2/\text{n}$

d= Anzahl Regressoren, $\hat{\sigma}^2$ Schätzer der Varianz des unsystematischen Fehlers ϵ, für n > 7 -> log(n) > 2

RSS = Residual squared sum; n = Anzahl Datenpunkte; d = Anzahl Parameter

Wie generieren wir Testdaten: Resampling Methoden stellen eine Möglichkeit dar

Resampling Methods

Cross Validation

Zielsetzung:
Schätzung Erwartungswert Test MSE

- Schätzung des Out-of-Sample Fehler
- Kriterium für Modellperformance
- Kriterium für Modellauswahl

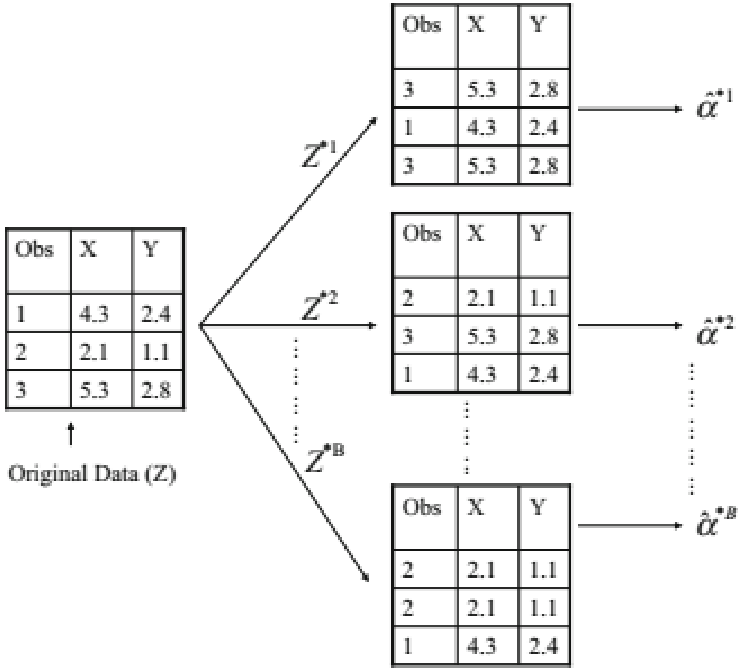
Bootstrapping

Zielsetzung:
Genauigkeit der Parameterschätzung

- Numerischer Ansatz, um Konfidenz und Quantile der Parameterschätzung zu bestimmen

Bootstrapping als allgemeingültiger Ansatz zur Berechnung von Unsicherheit

Aufbau zufälliger Datensamples: Ziehen aus Stichprobe mit zurücklegen



Zielsetzung und Vorgehensweise

Zielsetzung:

- Quantifizierung der Unsicherheit, die mit einer Parameterschätzung verbunden ist
  - Z.B. Standardabweichung der Parameterschätzung

Vorgehensweise

- Ziehen mit Zurücklegen aus der Datenstichprobe zur Generierung von n neuen Datensamples
- Parametrierung auf Basis eines jeden Datensamples

$$\mathrm{std}(\hat{\boldsymbol{\varphi}}) = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}\left(\hat{\varphi}_r - \overline{\hat{\varphi}_r}\right)^2}$$

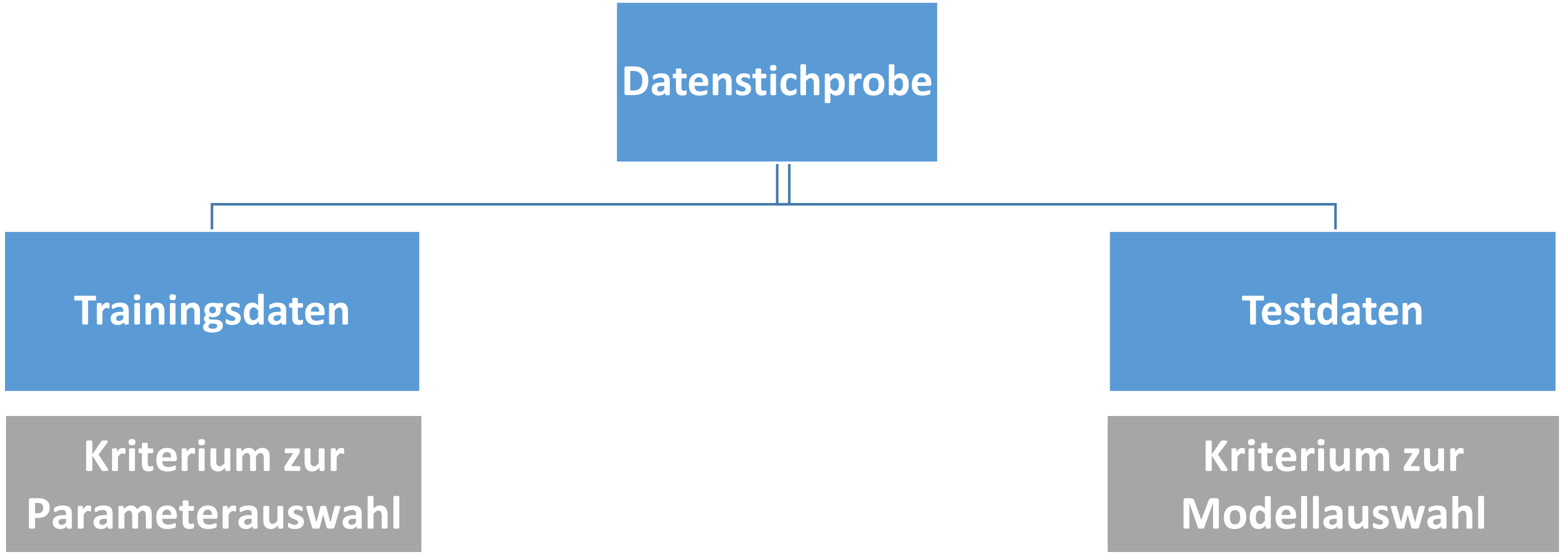
Quelle: James, Gareth, An Introduction to Statistical Learning

Die Modellauswahl kann durch direkte Betrachtung des Test-MSE oder Approximation des Test-MSE erfolgen

Modellauswahl

- **Auf Basis Validation/ Cross-Validation der Testdaten**
  - Validation mit Testdaten
  - Leave one out cross-validation (LOOC)
  - K-Fold
  - Walk-forward-selection
- **Informationskriterien auf Basis Trainingsdaten**
  - C
  - AIC
  - BIC
  - R^2

Durch die Aufteilung der Datenstichprobe in Trainings- und Testdaten sollen letztere als Schätzer für den Erwartungswert des MSE verwendet werden



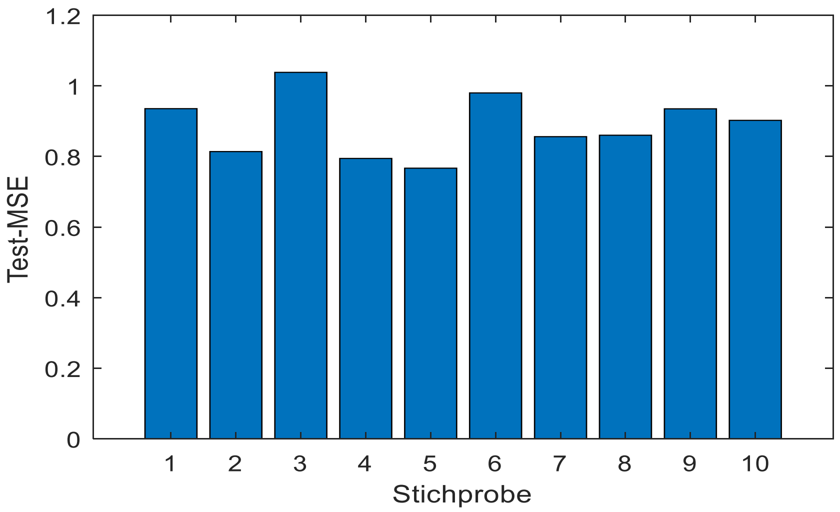
- Schätzung des Out-of-Sample Fehler

E[Test MSE]

$$Average\ \left(y_0 - \hat{f}(x_0)\right)^2$$

Der Validierungsset-Ansatz versucht anhand der Testdaten den Test-MSE zu schätzen; besitzt aber Nachteile insbesondere bei kleinen Stichproben

Ergebnis des Test-MSE nach 10-maliger Unterteilung von Trainings- und Testdaten



- Der <u>Validierungsset-Ansatz</u> unterteilt die Datenstichprobe in Trainings- und Testdaten

<u>Funktion der Trainingsdaten:</u>

- Schätzung funktionaler Zusammenhang

<u>Funktion der Testdaten:</u>

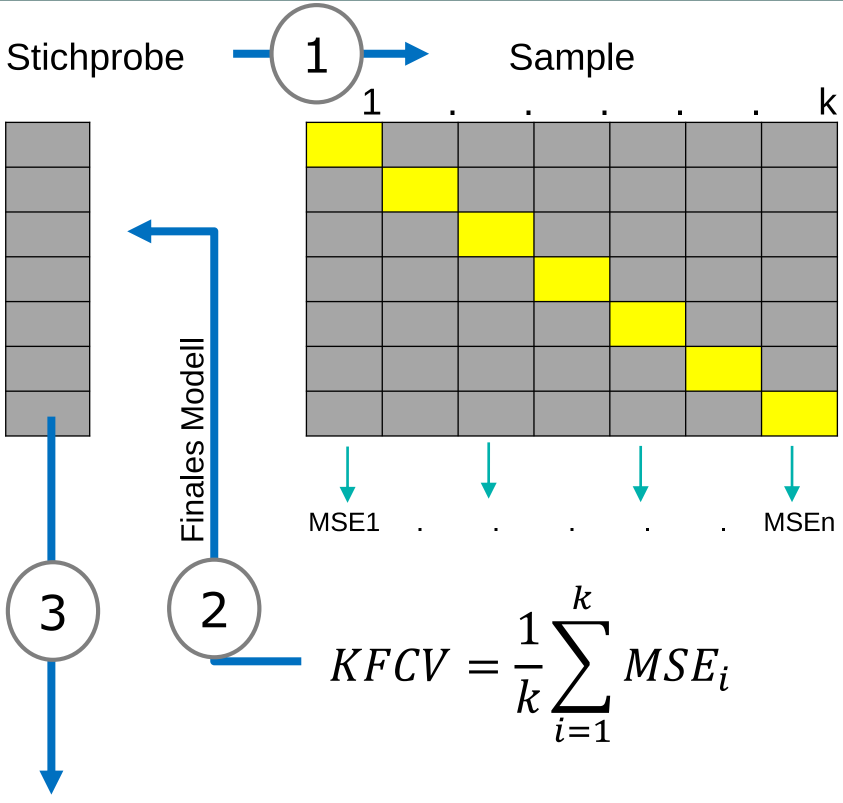
- Modellbewertung
- Schätzung des Test-MSE
- Modellauswahl

<u>Nachteile:</u>

- Schätzer für Test–MSE variiert in Abhängigkeit der zufällig gewählten Testdaten.
- Test- MSE wird bei kleiner Training-Stichprobe überschätzt.

Cross-Validierungs-Ansatz k-fold (KFCV) verringert den numerischen Aufwand ggü. LOOCV

Unterteilung der Datenstichprobe in k-Bereiche



$$KFCV = \frac{1}{k}\sum_{i=1}^{k} MSE_i$$

Funktionsweise:

- Erstellung k disjunkter zufälliger Segmente der Datenstichprobe

1. Kalibrierung Modell i exkl. i-ten Segment
2. Prognose i-ten Segment mit Modell i

$$\hat{y}_i = f_i(x_i)\ x_i \in Segment_i$$

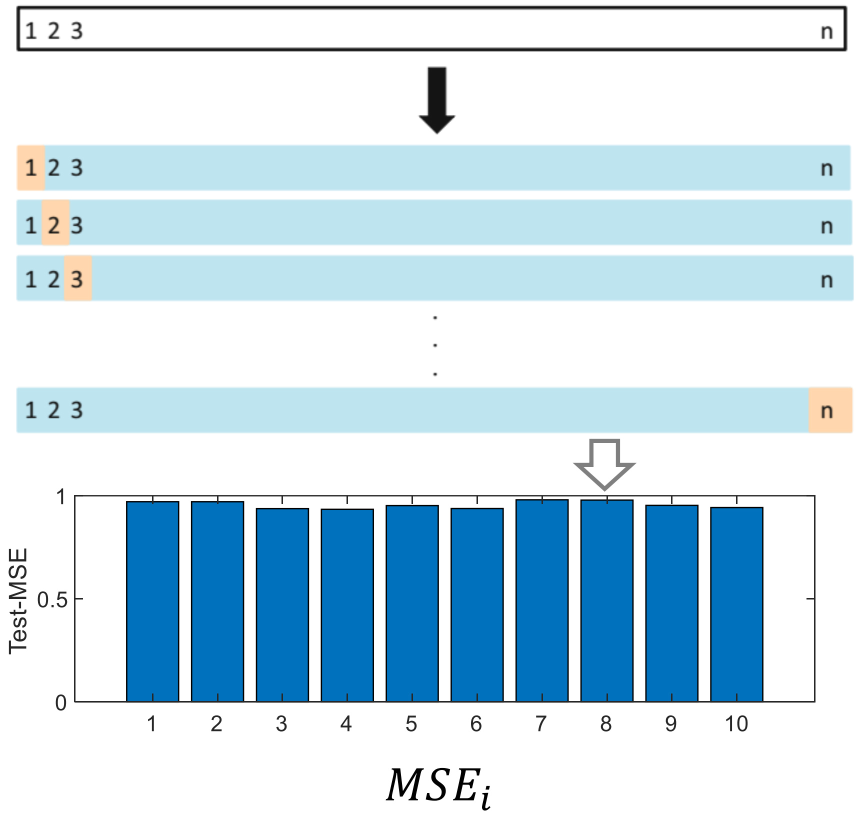
3. Berechnung des quadrierten Fehlers

$$MSE_i = \frac{1}{n_i}\sum_{i=1}^{n_i}(y_i - \hat{y}_i)^2$$

4. Berechnung des Cross Evaluation auf Basis aller durchlaufenden Testsets.
5. Wahl des Modells mit kleinsten KFCV

Leave one out Cross Validation LOOCV stellt der Parametrierung deutlich mehr Daten zur Verfügung

Unterteilung der Datenstichprobe



Funktionsweise:

1. Das Training Set =
   Stichprobe exklusive einer Beobachtung (x_i, y_i)
2. Prognose des Testdatenpunktes x_i
   $$\hat{y}_i = f_i(x_i)$$
3. Berechnung des quadrierten Fehlers
   $$MSE_i = (y_i - \hat{y}_i)^2$$

- Berechnung des Cross Evaluation auf Basis aller durchlaufenden Testsets.

$$LOOCV = \frac{1}{n}\sum_{i=1}^{n} MSE_i$$

Beurteilung:

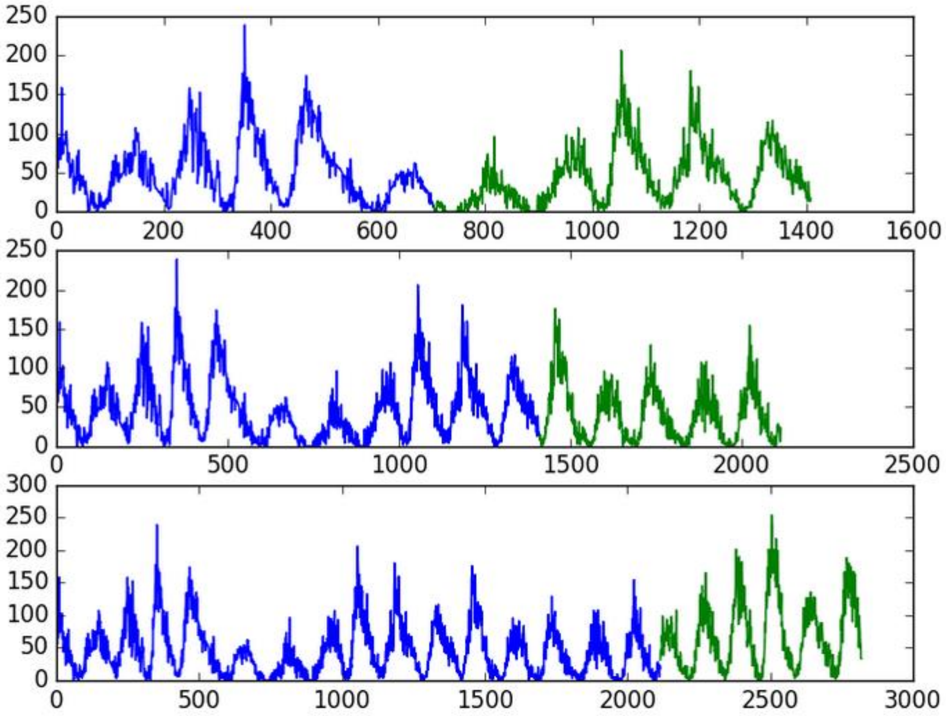
- Vorteil: große Trainingsstichprobe, stabiler Schätzer (LOOCV) des Test-MSE
- Nachteil: Performance

Quelle: James, Gareth, An Introduction to Statistical Learning

Für Zeitseriendaten sind die bisher besprochenen Unterteilungen ggf. nicht gut geeignet

Methoden für Zeitserien

TimeSeriesSplit: Verlängerung der Trainingsdaten bei konstant gehaltenem Testdatensatz



- Zeitseriendaten können zur Generierung von Training, Validierung und Testdaten nicht zufällig getrennt werden.
- Hierdurch würde die zeitliche Kopplung (autokorrelation) aufgehoben.
- Bisher betrachtete Lösungsmöglichkeit:
  - Teilgruppen müssen zeitliche Ordnung behalten (bei scikit learn mit dem Zusatz shuffle =‘None‘ erreichbar
  - Suksessive Verlängerung des Trainingsdatensatzes bei konstant gehaltenem Testdatensatz

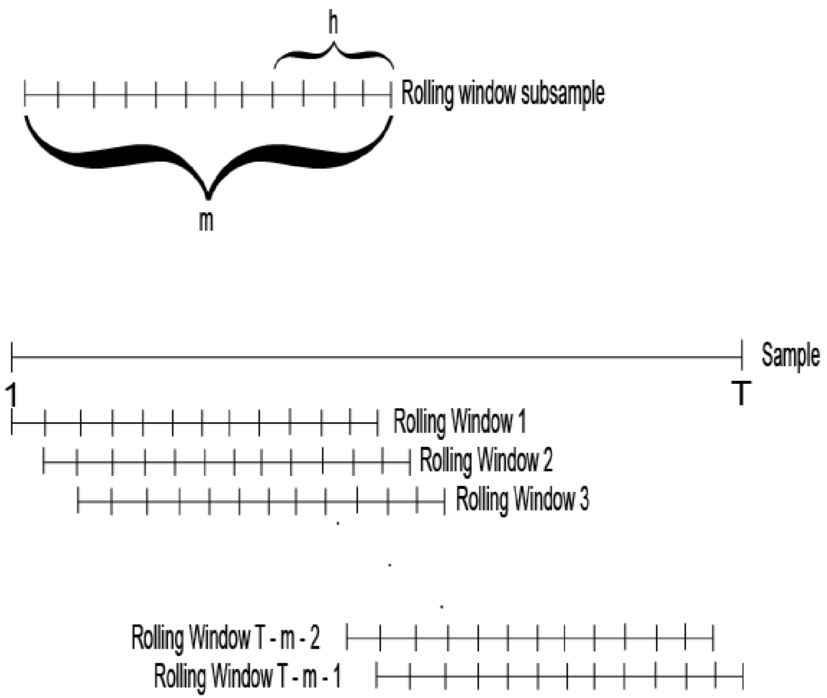
class sklearn.model_selection.TimeSeriesSplit(*n_splits=5*, *, *max_train_size=None*, *test_size=None*, *gap=0*)

Provides train/test indices to split time series data samples that are observed at fixed time intervals, in train/test sets. In each split, test indices must be higher than before, and thus shuffling in cross validator is inappropriate.

- Alternative Walk-forward-Validation
  - Hierbei erfolgt eine kontinuierliches Update des Modells sobald neue daten verfügbar sind.

Die walk-forward-Evaluation ist ein iterativer Prozess und ein guter Prozess in der Praxis zur Aktualisierung eines Modells

walk-forward evaluation



rolling window size, *m*, forecast horizon, *h*.

1. Es wird ein Fenster (Samplesize) m definiert.
2. Die erste Modellparametrierung basiert auf: $[x_1, x_m]$
3. Erstellung einer Prognose $\hat{y}_{m+h}$ für den Prognosehorizont h.
4. Berechnung des Fehlers: $e_1 = \hat{y}_{m+h} - y_{m+h}$.
5. Das Fenster wird um einen Zeitschritt verschoben und der Prozess wird wiederholt.
6. Berechnung des Test-RMSE aus den Einzelfehlern

$$RMSE = \sqrt{\sum_{i=1}^{T-m-h} e_i^2}$$

Notwendige Parameter:

- Minimale Anzahl an Beobachtungen zu Beginn m
- Prognose-Horizont h

1	Bedeutung des Bias-Variance Trade off
2	Generierung von Schätzern für den Test-RMSE
3	Best Subset selection Stepwise selection

Literaturempfehlung und Quellennachweise der Abbildungen: James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics), Springer New York. Kindle-Version.

<u>Best Subset Selection</u> ist ein standardisierter Suchalgorithmus zur Bestimmung des besten Modells im Fall einer multiplen Regression

Parametrierung

- Modelle mit 1 Parameter M_0
- Modelle mit 2 Parametern M_1
- …
- Modelle mit p Parametern M_p

1. Startpunkt: Modell $\boldsymbol{M_0}$ enthält keinen Regressor → $Prognose = \text{Mittelwert der Datenstichprobe}$
2. Für $k = 1, 2, \ldots, p$

   a) Parametrierung <u>aller möglichen</u> Modelle mit genau k Variablen (Regressoren)

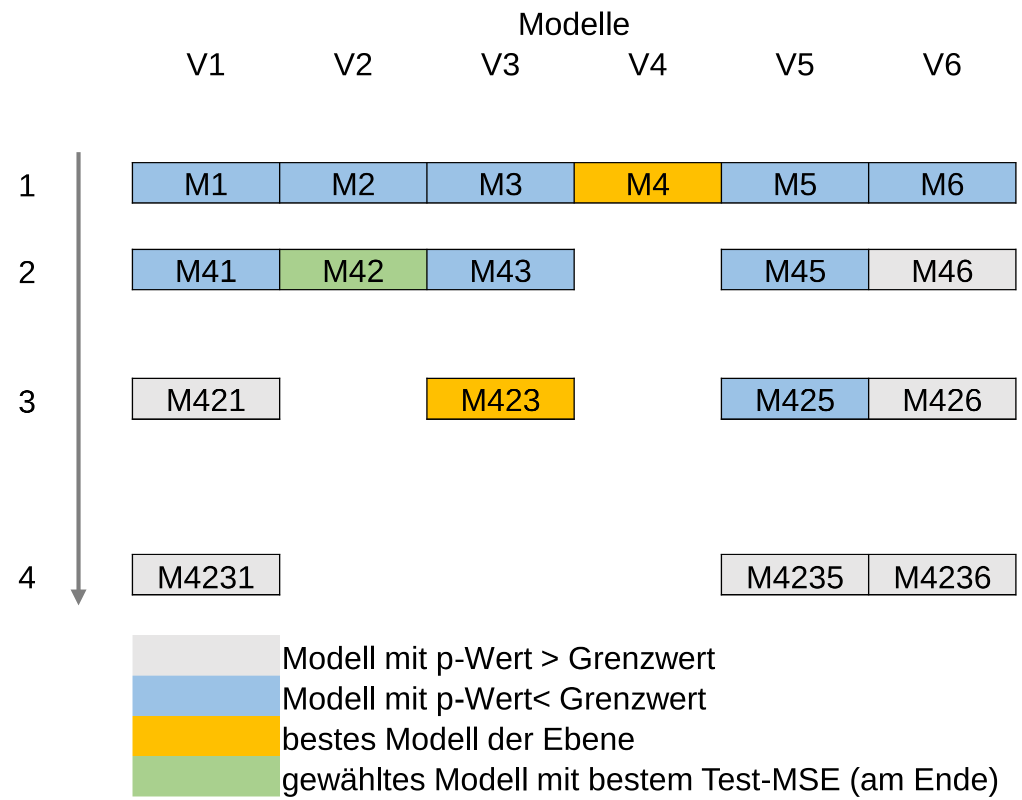
   b) Wahl des besten Modell M_k mit k Variablen entsprechend des kleinsten <u>Trainings-MSE</u> oder R^2
3. Wahl des besten Modells jeder Klasse aus $M_0, \ldots, M_p$ durch Verwendung Cross Validation.

<u>Erläuterung:</u>

Schritt 2 identifiziert das beste Modell (auf Basis der Trainingsdaten) für jede Teilmengengröße, um das Problem von $\binom{p}{2}$ möglichen Modelle auf eines der $p + 1$ möglichen Modelle zu reduzieren.
In Schritt 3 wird Cross Validation, Cp, BIC oder R^2 verwendet, um zwischen $M_0, M_1, \ldots, M_p$ auszuwählen.

Forward Stepwise Selection (FSS) ist ein inkrementelles Suchverfahren des besten Modells mit p Variablen

Abbildung der Modellsuche



- **FSS** betrachtet schrittweise einen kleineren Satz von Modellen.
- Auf der 1.ten Ebene werden p-Modelle mit jeweils 1 Regressor miteinander verglichen
- Die Variable des besten Modells wird beibehalten.
- In jedem Schritt wird die Variable, die die größte zusätzliche Verbesserung der Passform bewirkt, dem Modell hinzugefügt.

- Zur Wahl stehen nur Modelle mit signifikanten Parametern
- Am Ende wird mit Hilfe des Test-MSE das beste Modell gewählt

Problem: Auftreten von lokalen Minima

Suchmethoden helfen dabei den Modellauswahlprozess in die Schätzung der Parameter zu integrieren

	Modellhypothesen*	Modelle im CV
Klassische Modellauswahl		$p \cdot \frac{p+1}{2}$
Best subset Selection	$maximal$ $p \cdot \frac{p+1}{2}$	$p+1$
Forward Stepwise Selection		$maximal$ p

Grundsätzliche Schwierigkeit bei allen Ansätzen:

Es müssen alle in Frage kommenden Modellhypothesen separat geschätzt werden

Bildung Modellhypothese (**Vorauswahl** relevanter Parameter)

Risiko:

Modelle mit potentiell niedrigen E[Test-MSE] bleiben unberücksichtigt

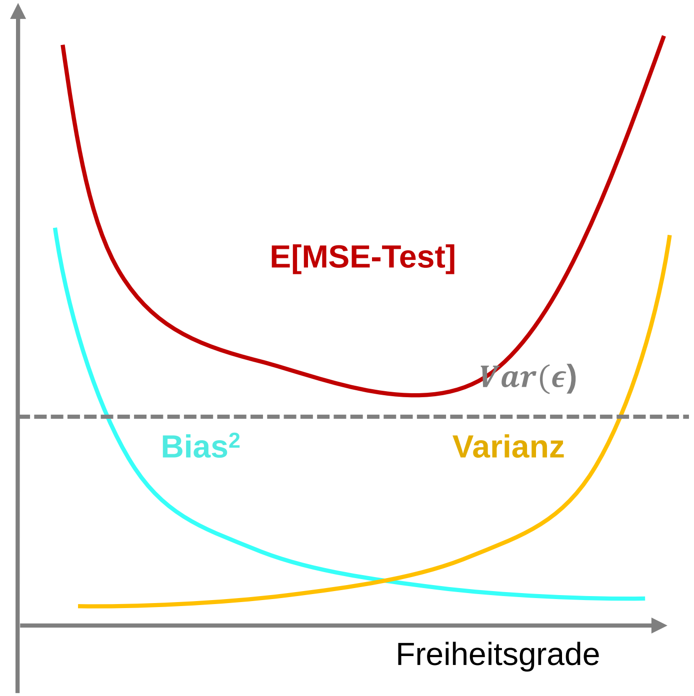
*Annahme jeder Regressor wird nur maximal einmal in das Modell mitaufgenommen

Zusammenfassung

Im Rahmen der Modell-kalibrierung existiert ein trade-off zwischen Bias und Varianz

Es existieren direkte und indirekte Methoden zur Schätzung des erwarteten Test- MSE

Best Subject und Forward stepwise Selection verringern den Suchraum



Modellauswahl

Validation/ Cross-Validation
- Training und Testdaten
- LOOC
- K-Fold

Informations-kriterien
- C
- AIC
- BIC
- R^2

Klassische Modellaus-wahl — $p \cdot \frac{p+1}{2}$

Best subset Selection — $p+1$

Forward Step-wise Selection — $maximal$ p