Technische Hochschule Wildau [FH]
Technical University of Applied Sciences

Maschinelles Lernen

Modellselektion und -validierung

Prof. Dr. Rainer Stollhoff

Übersicht

Technische Hochschule Wildau
Technical University of Applied Sciences

- Motivation
- Bias-Varianz-Zerlegung
- Regularisierung in der Modellselektion
- Resampling in der Modellvalidierung

Motivation: Polynominterpolation

Aufgabe: Regression, d.h. Vorhersage $\hat{y} = \hat{y}(x) = f(x)$

Erfahrung: Datensatz $(x_i, y_i)_{i=1}^{n}$

Qualität: Verlustfunktion: $L(y, \hat{y}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - f(x_i;\theta))^2 = L(\theta)$

Maschine: Regression mit $f(x;\ \theta_1, \cdots, \theta_m) = \theta_0 + \theta_1 \cdot x + \theta_2 \cdot x^2 + \cdots + \theta_m \cdot x^m$

Lernen: Finde Werte für $\theta = (\theta_1, \cdots, \theta_m)$, die die quadratische Verlustfunktion minimieren

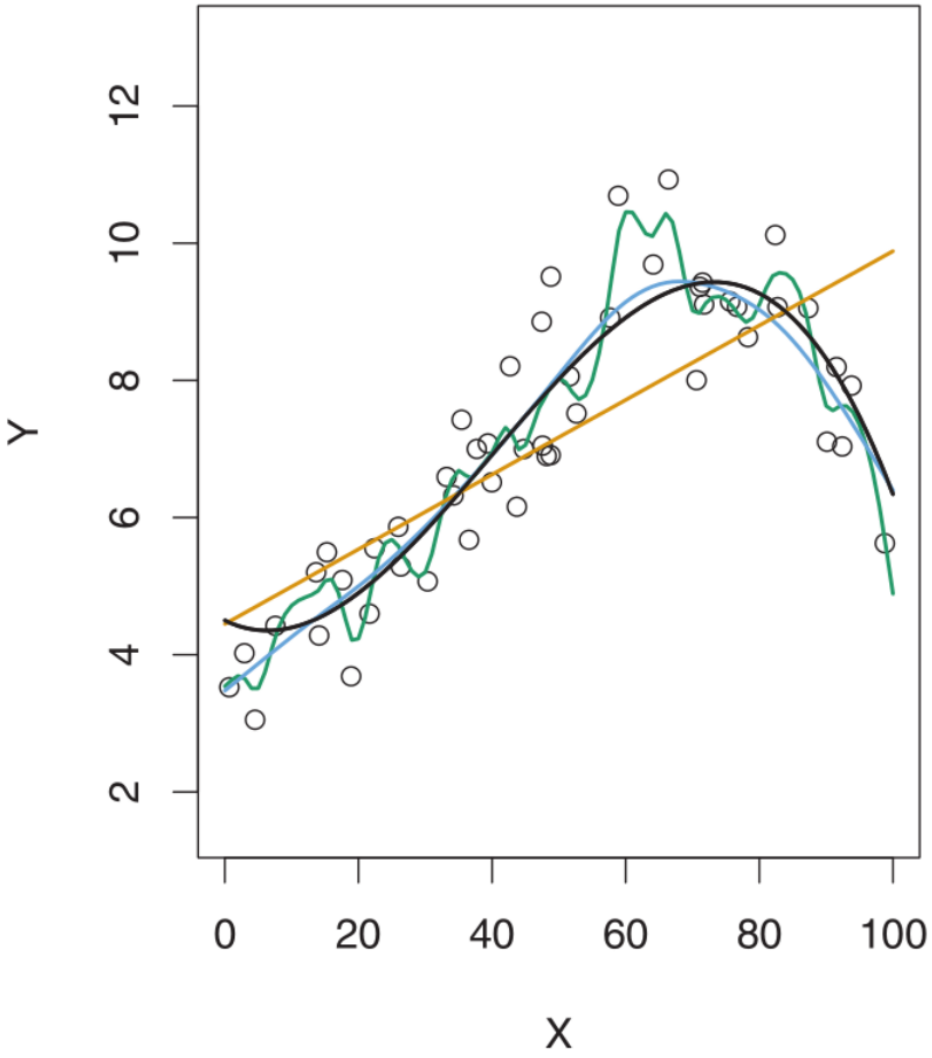
Polynominterpolation:

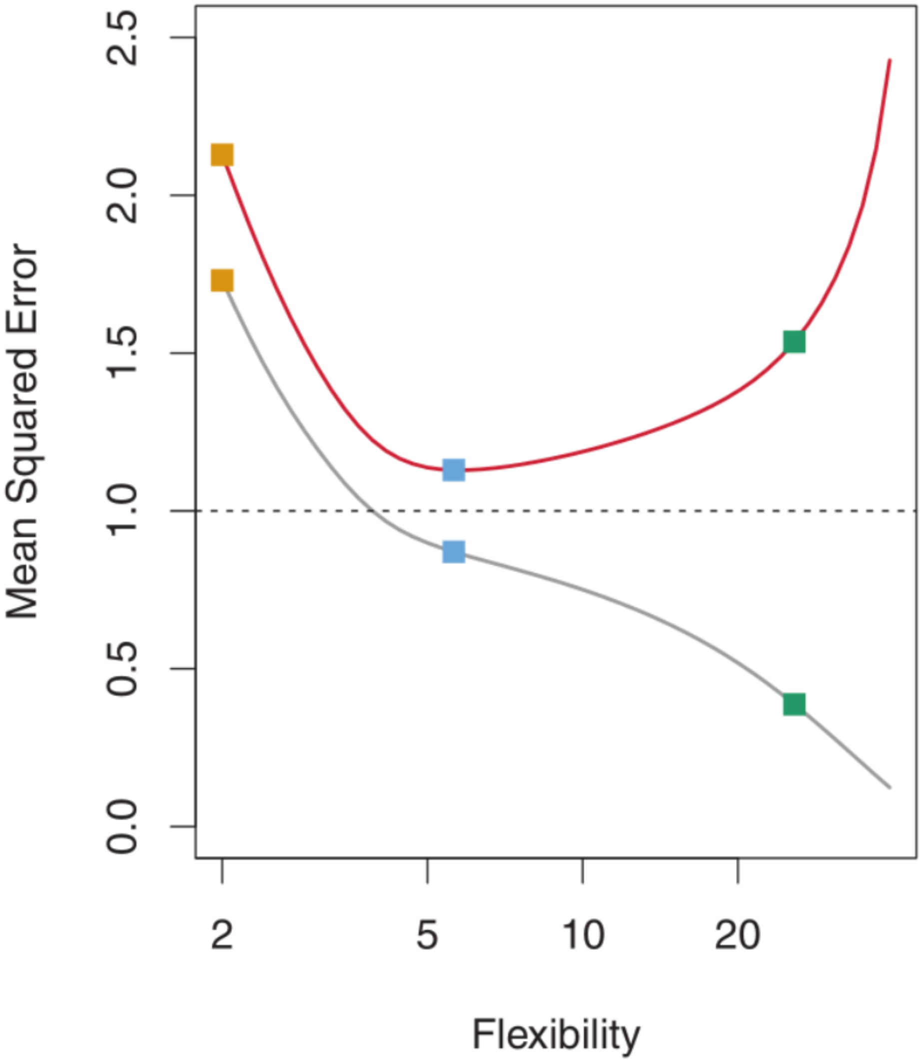
Falls $m \geq n$ gibt es immer Werte für θ so dass $f(x_i;\theta) = y_i$ und damit $L(\theta) = 0$

mit $$\ell_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} = \frac{x - x_0}{x_i - x_0} \cdots \frac{x - x_{i-1}}{x_i - x_{i-1}} \cdot \frac{x - x_{i+1}}{x_i - x_{i+1}} \cdots \frac{x - x_n}{x_i - x_n}$$

definiere $$f(x) = \sum_{i=1}^{n} y_i\ \ell_i(x)$$

Motivation: Overfitting der Trainingsdaten





James et al., *An Introduction to Statistical Learning : with Applications in R. New York :Springer, 2013.*

Bias-Varianz-Zerlegung des Vorhersagefehlers

Technische Hochschule Wildau
Technical University of Applied Sciences

- Bestmögliche Vorhersage
  - Vorhersage $y = f(x) + \epsilon$ dabei $f(x)$ als bestmögliche Vorhersage und ϵ als echter Zufallswert
  - Unvermeidbarer Fehler $E_{y,x}\left[[(y - f(x))^2\right]$
- Bestmögliches geschätztes Modell
  - Vorhersage $\hat{f}(x)$ minimiert $\left(f(x) - \hat{f}(x)\right)^2$ über x
  - Bias des Modells $E_x\left[[(f(x) - \hat{f}(x))^2\right]$
- Bestmögliches auf einem Trainingsdatensatz geschätztes Modell
  - Trainingsdatensatz $\mathrm{T} = (x_i, y_i)$ für $i = 1, \cdots, n$
  - Vorhersage $\hat{y} = \hat{f}(x; T)$ minimiert $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$
  - Varianz der Vorhersagen $E_{x,T}\left[[(\hat{f}(x) - \hat{f}(x;T))^2\right]$
- Bias-Varianz-Zerlegung

$$E_{y,x}\left[[(y - \hat{f}(x))^2\right] = E_{y,x}\left[[(y - f(x))^2\right] + E_x\left[[(f(x) - \hat{f}(x))^2\right] + E_{x,T}\left[[(\hat{f}(x) - \hat{f}(x;T))^2\right]$$

Vorhersagefehler = Unvermeidbarer Fehler + Bias + Varianz

Übersicht

- Motivation
- Bias-Varianz-Zerlegung
- **Regularisierung in der Modellselektion**
  - Ziel: Reduktion der Modellkomplexität und damit Verringerung der Varianz
- Resampling in der Modellvalidierung

Bias-Varianz-Zerlegung des Vorhersagefehlers

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

Unvermeidbarer Fehler

$$E_x\left[(y - f(x))^2\right]$$

Bias des Modells

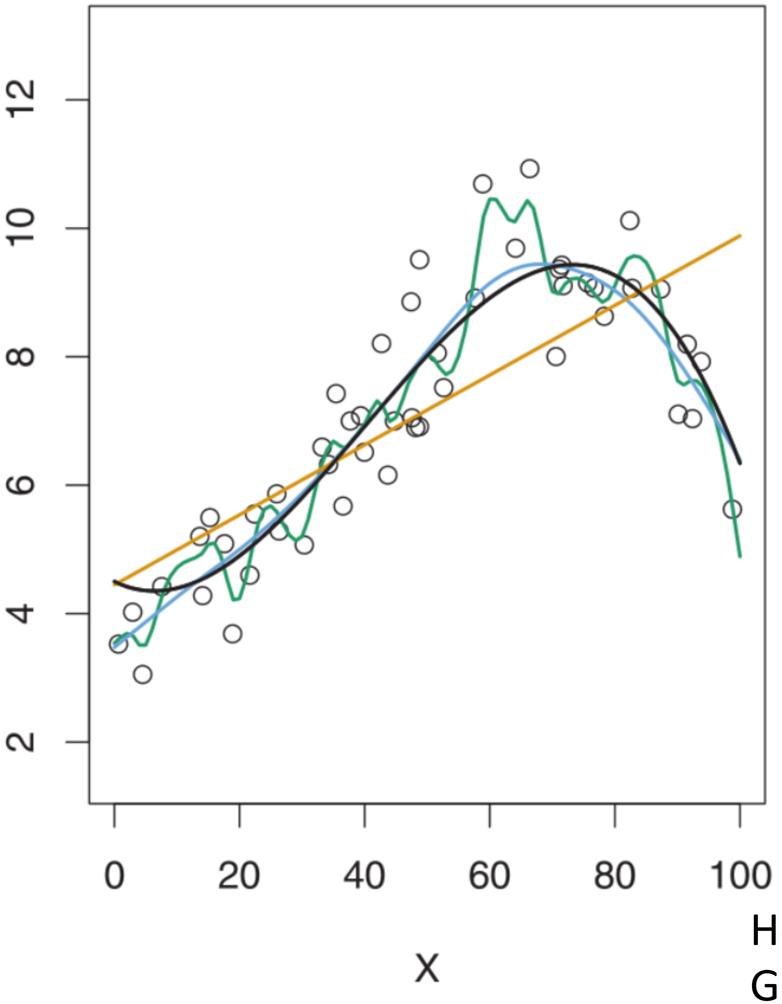
$$E_x\left[(f(x) - \hat{f}(x))^2\right]$$

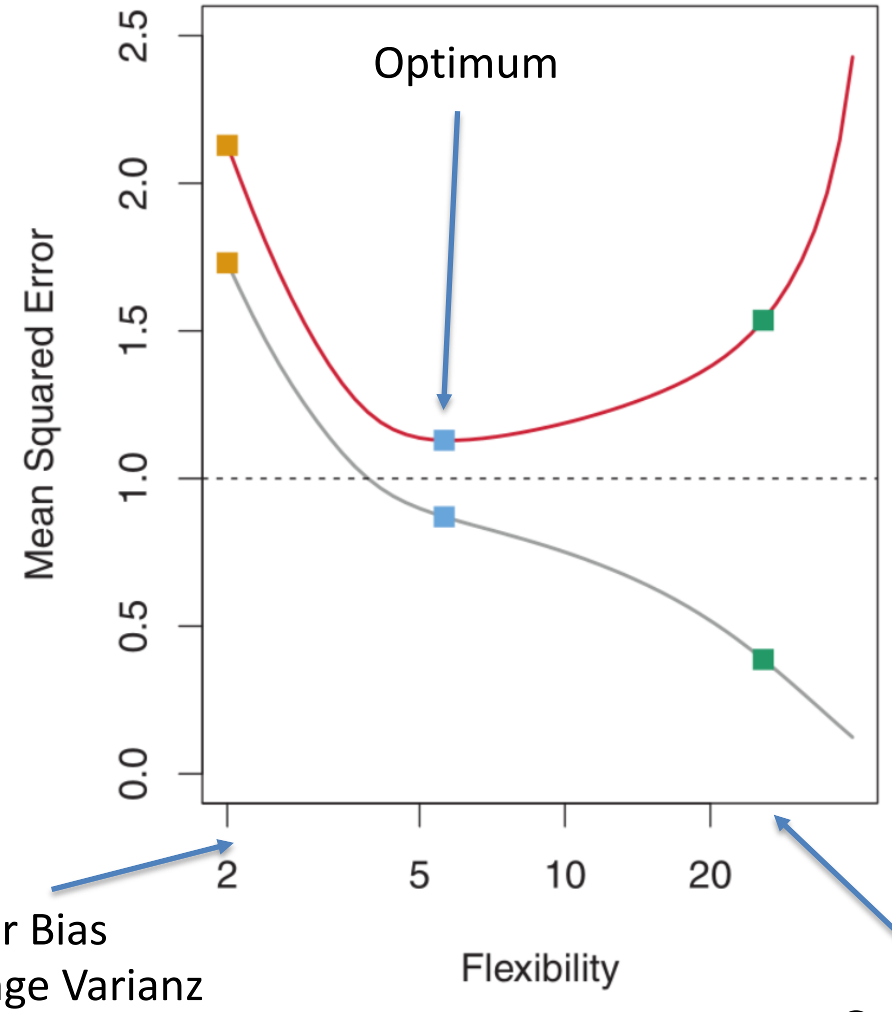
Varianz der Vorhersagen

$$E_{x,T}\left[(\hat{f}(x) - \hat{f}(x;T))^2\right]$$

Vorhersagefehler

$$E_x\left[(y - \hat{f}(x))^2\right]$$





Hoher Bias
Geringe Varianz

Geringer Bias
Hohe Varianz

James et al., *An Introduction to Statistical Learning : with Applications in R. New York :Springer, 2013.*

Methoden der Regularisierung in der Modellselektion

- Regularisierungsterm in der Verlustminimierung für parametrisches Modell mit $\theta = (\theta_1, \cdots, \theta_m)$
  - Tikhonov Regularization / Ridge Regression:
  - $L(\theta; \lambda) = \sum_{i=1}^{n} (y_i - f(x_i; \theta))^2 + \lambda \sum_{j=1}^{m} \left(\theta_j\right)^2$
  - Akaike Information Criterium:
  - $AIC(\theta) = -2 \left(\sum_i \log \hat{p}(y_i; x_i, \theta)\right) + 2 \cdot m$
  - Bayesian Information Criterium:
  - $BIC(\theta) = -2 \left(\sum_i \log \hat{p}(y_i; x_i, \theta)\right) + \ln(n) \cdot m$
- Early-Stopping, d.h. vorzeitiges Abbrechen der Optimierungsiterationen zum Beispiel
  - Begrenzen der Anzahl der aufeinanderfolgenden Splits bei Klassifikationsbäumen oder
  - Begrenzen der Anzahl der Boosting-Iterationen
- Feste Einschränkung der Modellkomplexität zum Beispiel
  - Beschränken des maximalen Grades einer polynomialen Funktion
  - Beschränken auf lineare Funktionen in der multivariaten Regression

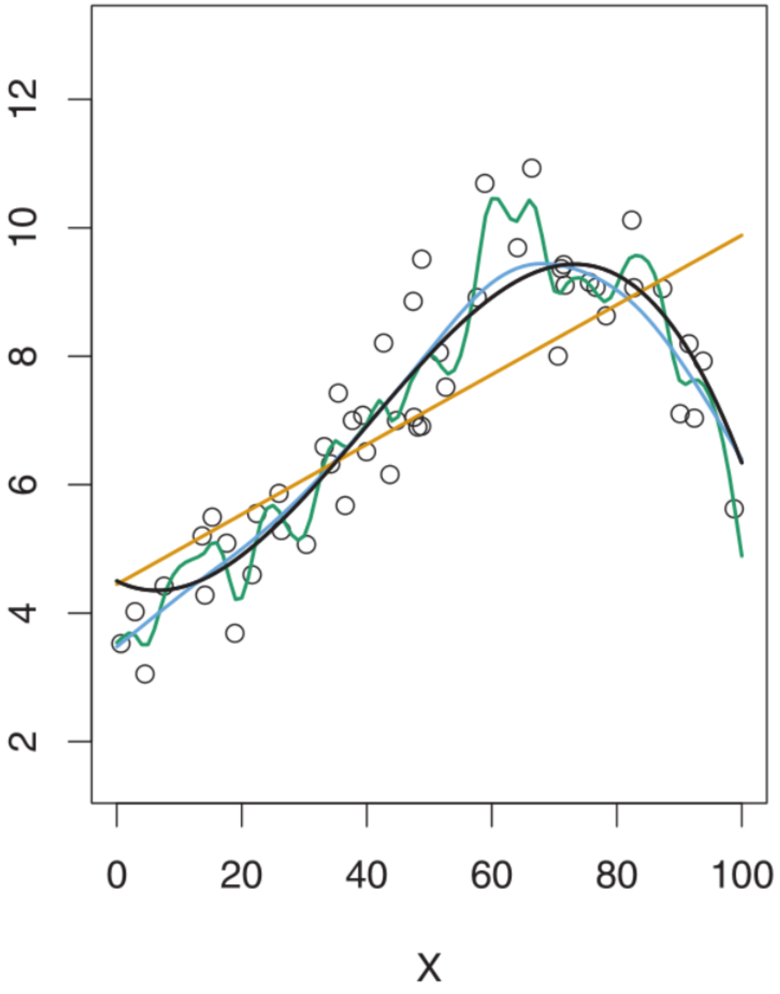
Übersicht

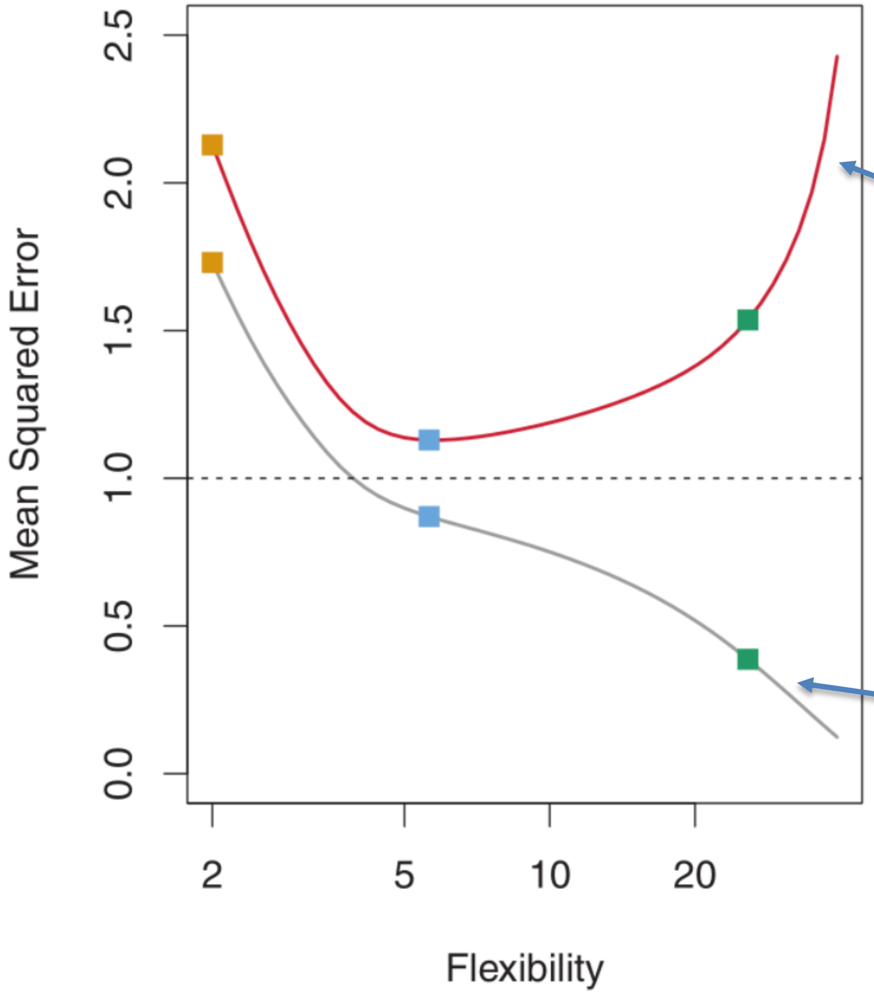
Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

- Motivation
- Bias-Varianz-Zerlegung
- Regularisierung in der Modellselektion
- **Resampling in der Modellvalidierung**
  - Ziel: Verlässlichere Schätzung des Vorhersagefehlers

Bias-Varianz-Zerlegung des Vorhersagefehlers

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU





Echter Fehler unbekannt

Trainingsfehler
zu optimistisch

James et al., *An Introduction to Statistical Learning : with Applications in R. New York :Springer, 2013.*

Technische Hochschule Wildau
Technical University of Applied Sciences
WILDAU

Sampling zur Modellvalidierung

Festes Holdout

Aufteilen

Train
Train
Train
Train
Test

Zufälliges Holdout

Permutieren (Mischen)

Aufteilen

Train
Train
Train
Train
Test

Kreuzvalidierung

Permutieren (Mischen)

Aufteilen

Train	Train	Train	Train	Test
Train	Train	Train	Test	Train
Train	Train	Test	Train	Train
Train	Test	Train	Train	Train
Test	Train	Train	Train	Train

Wiederholtes zufälliges Holdout

Permutieren (Mischen)

Aufteilen

Train
Train
Train
Train
Test

Permutieren (Mischen)

Aufteilen

Train
Train
Train
Train
Test

...

Wiederholte Kreuzvalidierung

Permutieren (Mischen)

Aufteilen

Train	Train	Train	Train	Test
Train	Train	Train	Test	Train
Train	Train	Test	Train	Train
Train	Test	Train	Train	Train
Test	Train	Train	Train	Train

Permutieren (Mischen)

Aufteilen

Train	Train	Train	Train	Test
Train	Train	Train	Test	Train
Train	Train	Test	Train	Train
Train	Test	Train	Train	Train
Test	Train	Train	Train	Train

...

Leave-One-Out Kreuzvalidierung

Training auf n-1 Daten, Test auf Restbeobachtung