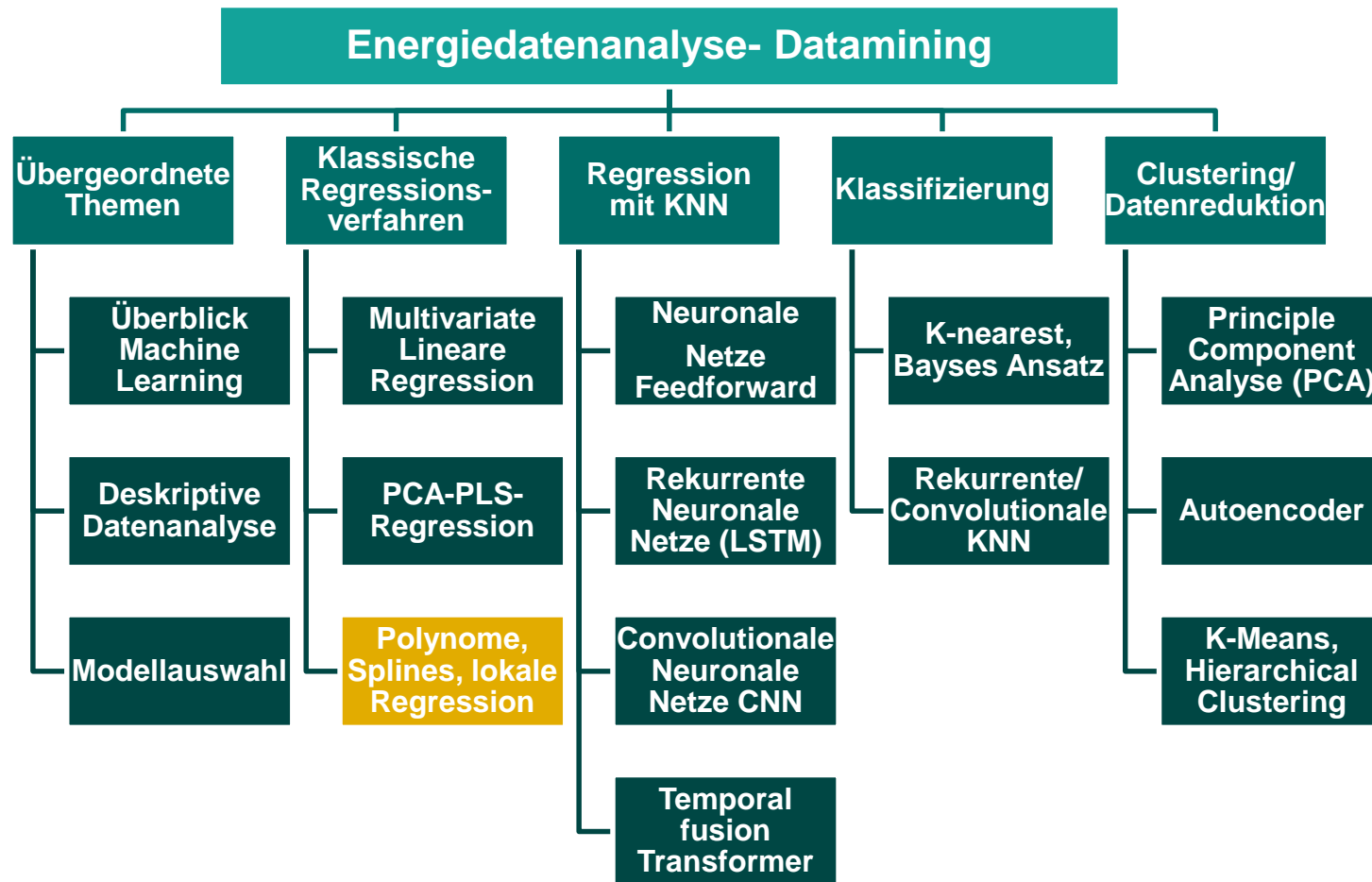


Splines und lokale Regression

Energiedatenanalyse - Datamining

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



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

Thema

Überblick über Verfahren: Polynome, Splines und lokale Regression

Aufbau der heutigen Vorlesung:

Lernziel:

1

Methoden: Polynome,
Splines, lokale Regression

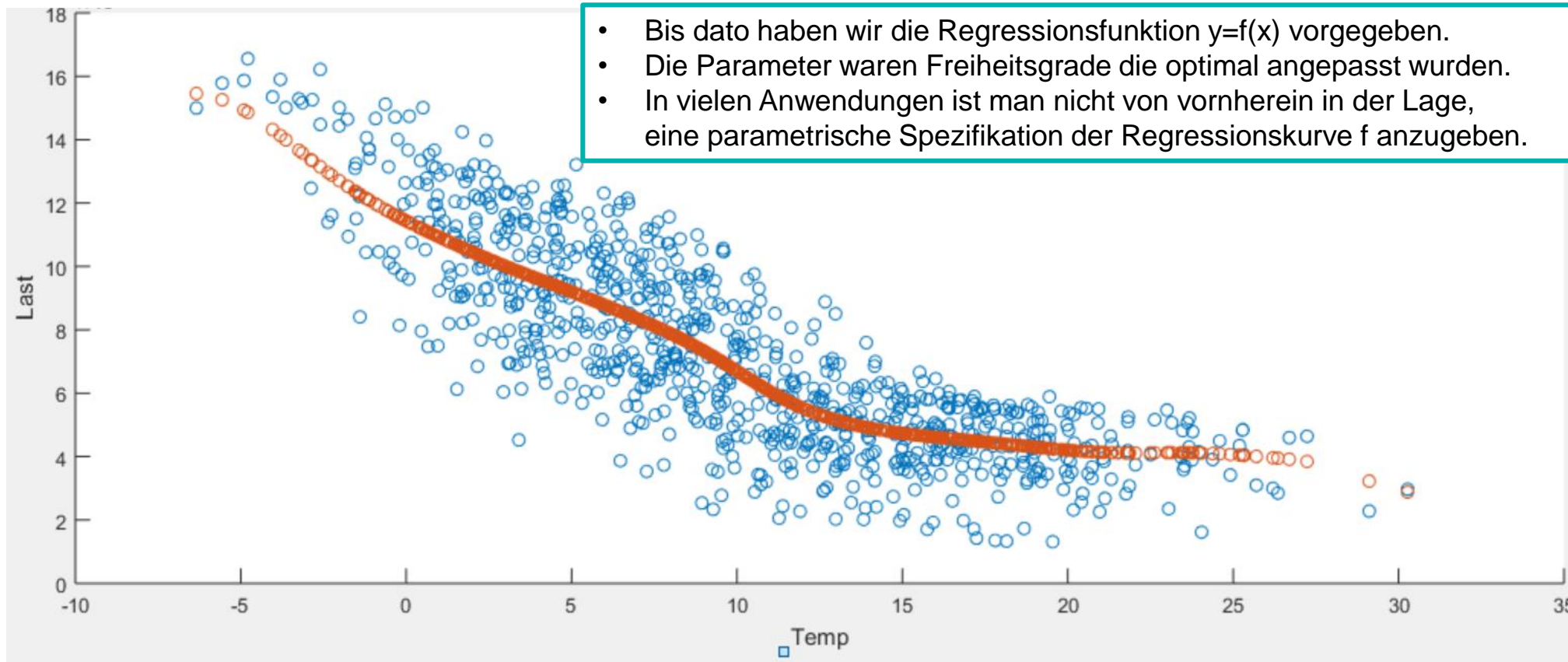
Nachvollziehen der mathematischen
Grundlagen

2

Energiewirtschaftliche Anwendung

Temperatur und Gasnachfrage

Häufig sind in der Energiewirtschaft nichtlineare Zusammenhänge zu beobachten

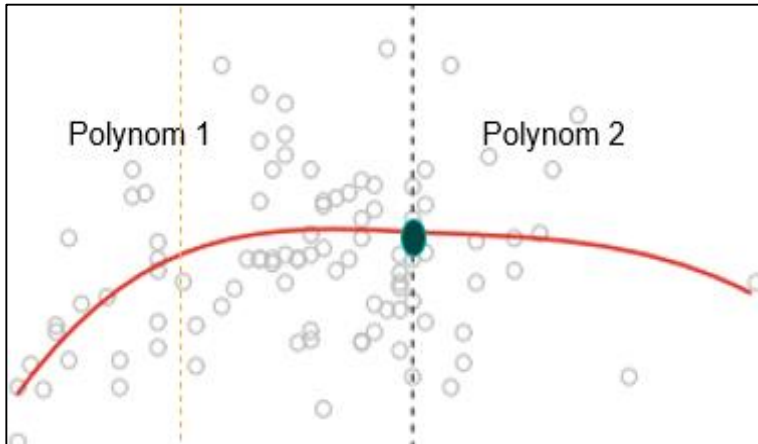


Nichtparametrische Regressionsmethoden kommen ohne konkrete Ausgestaltung von $f(x)$ aus.
Das Ziel der nichtparametrischen Regression besteht in der Schätzung der Funktion f .

In der linearen Regression beeinflussen die Parameter β linear y . Nichtlinear hingegen ist z.B. folgender Ansatz: $Y_i = e^{\beta} X_i$

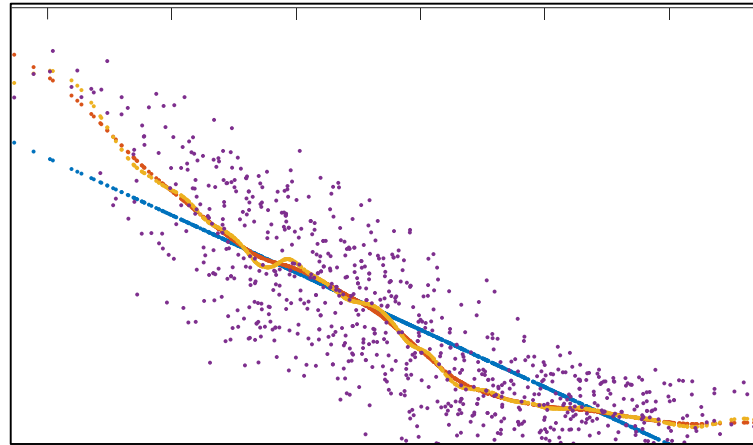
Für die Betrachtung nichtlinearer Zusammenhänge kommen unterschiedliche Verfahren in Frage

Regression-Splines



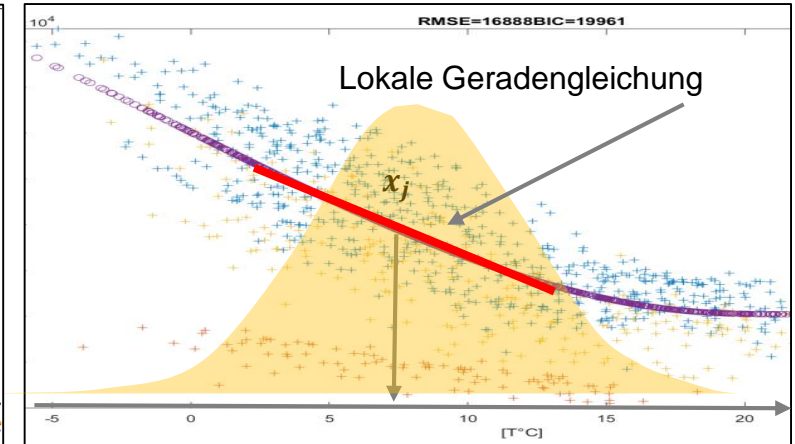
Regressions-Splines versuchen anhand einer Verkettung (Stützstellen) einfacher parametrischer Funktionen den Funktionsverlauf abzubilden.

Natürliche Splines



Natürliche Splines sind eine Weiterentwicklung von Regressions-Splines und besitzen eine extrem hohe Flexibilität nichtlineare Verläufe abzubilden.

Lokale Regression



Die lokale Regression ist ein auf der Methode der Kerndichteschätzung aufbauendes Verfahren und besitzt ebenfalls eine sehr hohe Flexibilität.

1

Polynome und Splines

2

Lokale Regression

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.

Als erste Idee zur Abbildung eines nichtlinearen Zusammenhangs zwischen X und Y kann eine Treppenfunktion verwendet werden

Stückweise Linearisierung

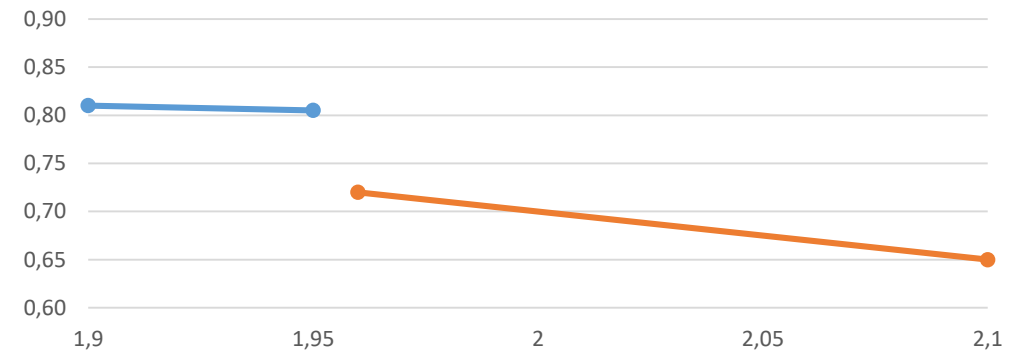
$$y = \begin{cases} a_1 + b_1x & \text{if } -\infty < x < c_1 \\ a_2 + b_2x & \text{if } c_1 \leq x < c_2 \\ a_3 + b_3x & \text{if } c_2 \leq x < c_3 \end{cases}$$

Modell mit 2 lin. Funktionen und Intervallwechsel bei $c = 1,96$

Parameter		$c = 1,96$		
a	b	X	bx	$Y=a+bx$
1	-0,1	1,9	-0,19	0,81
1	-0,1	1,95	-0,20	0,81
1,7	-0,5	1,96	-0,98	0,72
1,7	-0,5	2,1	-1,05	0,65

Erläuterung

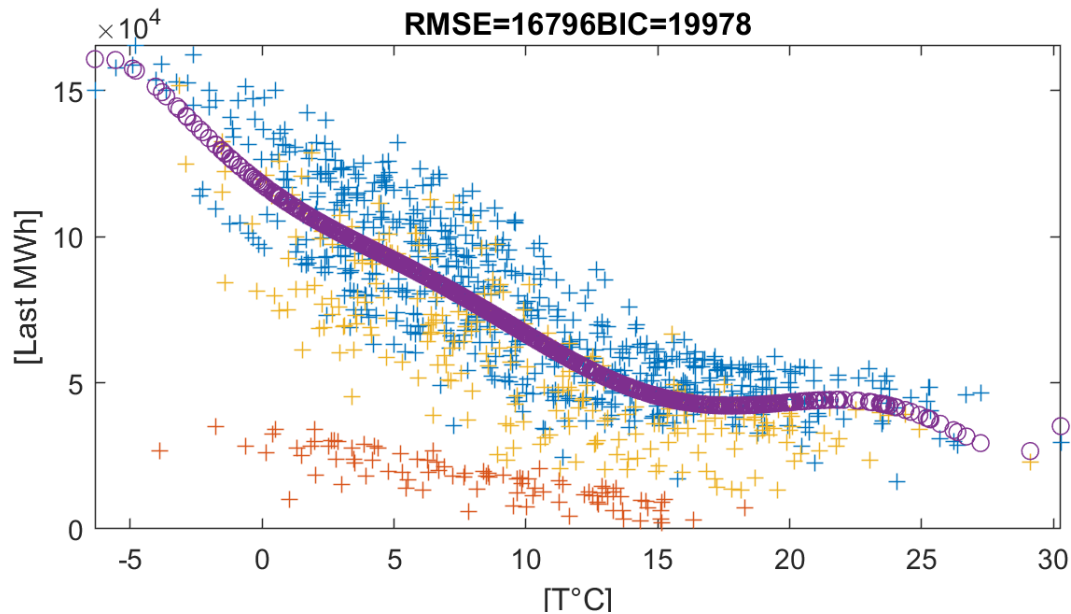
- Bei einer stückweisen Linearisierung des Zusammenhangs zwischen Y und X werden einzelne Bereiche c_i für X definiert und in diesen Bereichen erfolgt separat eine Parameterschätzung
- Diese Vorgehensweise benötigt einen kompletten Parametersatz pro Intervall und führt zu Sprungstellen an c



Als klassische Alternative bietet sich ein Polynom zur Abbildung von Y an

Abbildung des Zusammenhangs

$$y = \beta_0 + \sum_{i=1}^n \beta_i x^i + e$$



Erläuterung

- Polynome können nichtlineare Abhängigkeiten abgebildet.
- Das Regressionsverfahren funktioniert analog zur linearen Regression; alle Schätzverfahren besitzen weiter ihre Gültigkeit.

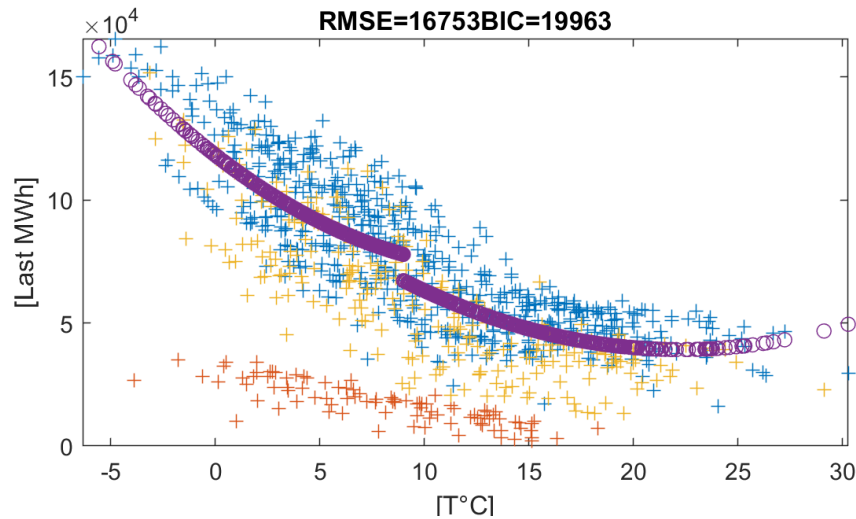
Nachteile:

- Polynomen streben gegen $+\infty/-\infty$
- Polynome reagieren an Rändern sensitiv ggü. Stichprobenänderungen
- Parameter sind häufig bei einzelnen Gliedern insignifikant (hoher p-Wert).

Alternativ zum Einsatz von Polynomen mit hoher Ordnung lassen sich stückweise kubische Polynome bilden

Stückweise kubische Polynome

$$y = \begin{cases} a_{01} + b_{01}x + b_{02}x^2 & \text{if } -\infty < x < c_1 \\ a_{11} + b_{11}x + b_{12}x^2 & \text{if } c_1 \leq x < c_2 \\ a_{21} + b_{21}x + b_{22}x^2 & \text{if } c_2 \leq x < c_3 \end{cases}$$

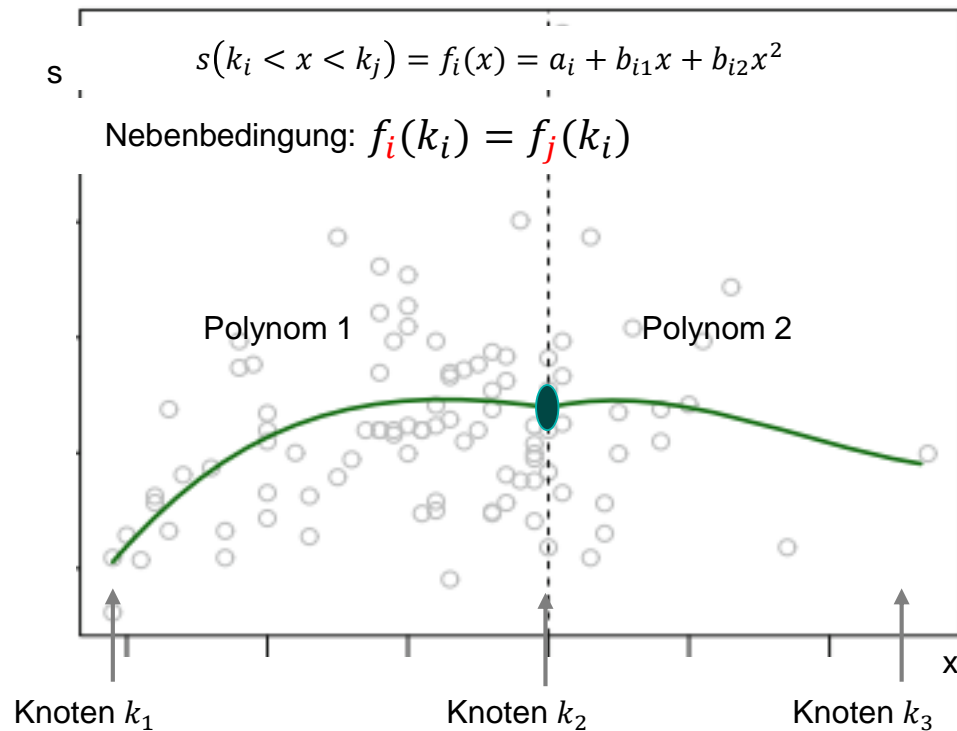


Erläuterung

- Stückweise Polynomfunktionen können mit niedrigerer Ordnung den Zusammenhang in der Regel besser abbilden als Polynome höherer Ordnung.
- Die Parameter der einzelnen Polynome mit niedriger Ordnung sind häufiger signifikant.
- Hier dargestellt Polynom zweiter Ordnung und einem Knoten.
- **Problem:** resultierende Funktion an den Abschlusstellen c ist weiterhin nicht stetig!

Ein **Spline n-ten Grades** ist eine Funktion, die **stückweise aus Polynomen höchstens n-ten Grades zusammengesetzt** ist.

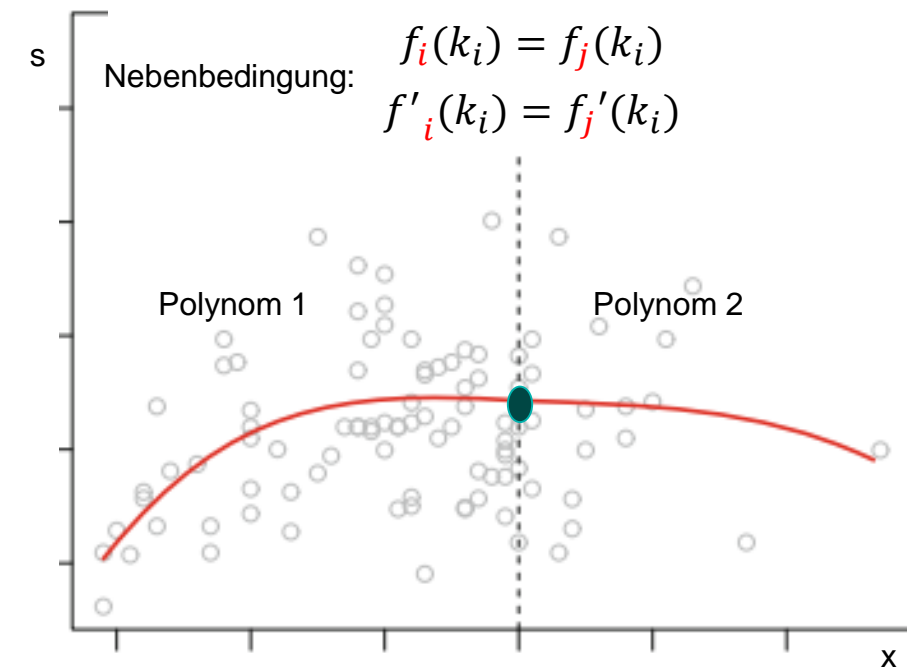
Stückweise Kontinuität



Nebenbedingung:

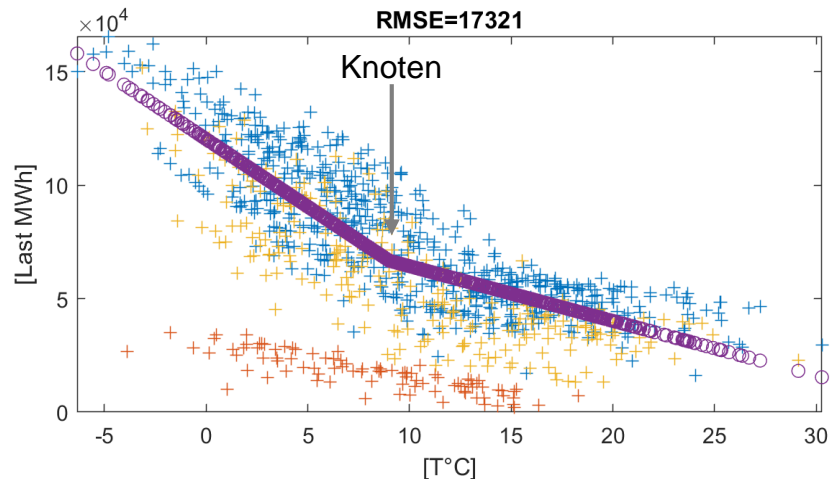
der kontinuierliche Verlauf des Splines wird gewährleistet, wenn dieser an den Knoten (n-1)-mal stetig differenzierbar ist.

Kontinuität zweite Ableitung (1.te und 2.te Ableitung müssen stetig sein)



Ein linearer Spline verbindet am Knoten zwei stückweise lineare Funktionen und kann wie folgt mathematisch formuliert werden

Darstellung



Regressionsfunktion in Matlab (Matrix):

Parameter	
a1	3
b1	-0,5
b2	0,7

Es ergeben sich $2+n$ Parameter
(n steht für die Anzahl der Knickstellen)



Verallgemeinerung

- Der lineare Spline besitzt im Knoten für beide Teilstücke den gleichen Wert*.
- Dies erfolgt durch das schrittweise „anschalten“ weiterer Terme, die erst rechts vom Knoten einen Wert größer Null annehmen.

$$y = a_1 + b_1x + b_2 \cdot \max(0, x - c_1) + \dots b_{n+2} \max(0, x - c_n)$$

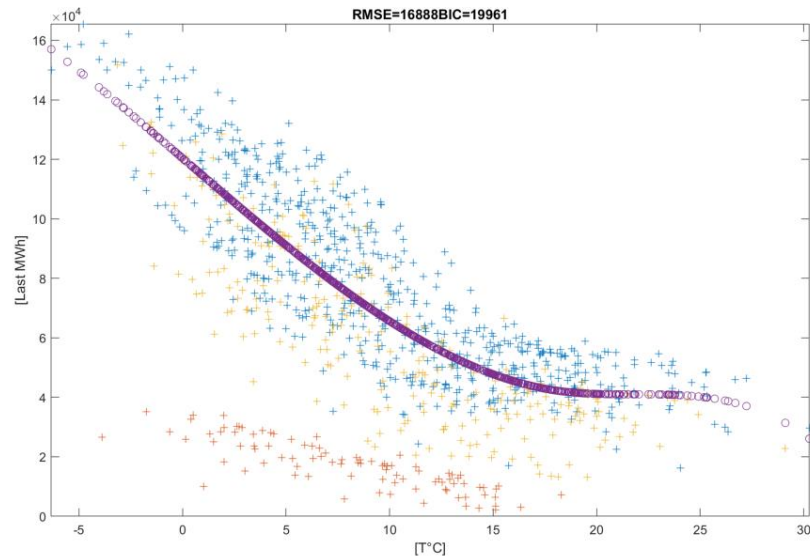
Regression

X	a	b_1x	$b_2 \cdot \max(0, x - c_1)$	Y
1,9	3	-0,95	0	2,05
1,95	3	-0,98	0	2,03
2	3	-1	0	2
2,1	3	-1,05	0,07	2,02

* Im Knoten besitzen beide stückweise linearen Funktionen nicht die gleiche Steigung. Deshalb existiert eine Knickstelle

Ein kubischer Spline besitzt für jedes Segment ein Polynom dritten Grades

Darstellung



Eigenschaften eines kubischen Splines s

1. f ist ein stückweises Polynom 3 Grades
2. $f(x_i) = y_i$
3. f ist in den Knoten zweifach differenzierbar

Mathematischer Aufbau

- Aufbau eines kubischen Splines mit K Stützstellen

$$y = a_1 + b_1x + b_2x^2 + b_3x^3 + \sum_{i=1}^K b_{3+i} \cdot \max(0, x - c_i)^3$$

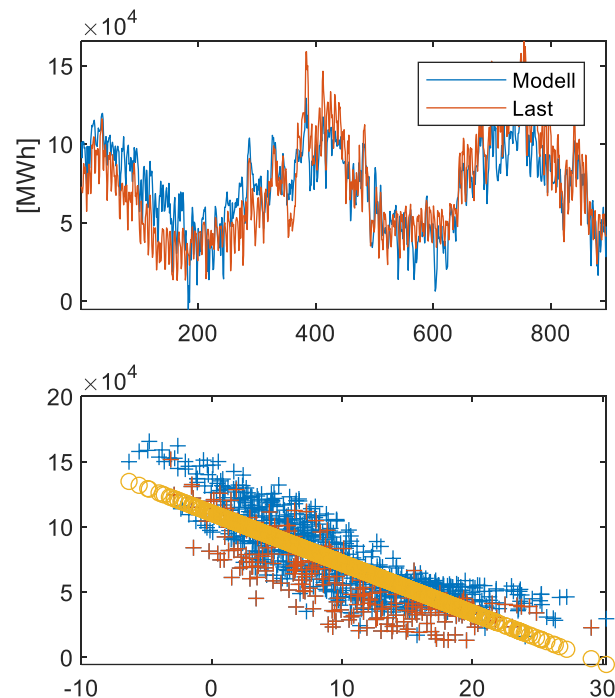
- An jedem Knoten i ist ein auf Null begrenztes Polynom hinzugefügt:

$$b_{3+i} \cdot \max(0, x - c_i)^3$$

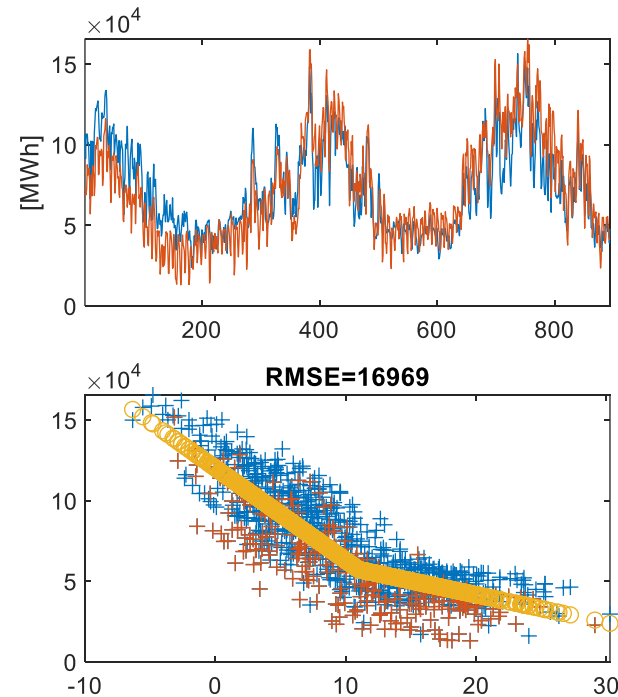
- Es werden somit bei einem kubischen Spline 4 (a_1, b_1, b_2, b_3) + K Parameter geschätzt

Mit Hilfe einer geringen mathematischen Umstellung lässt sich der Modellfit verbessern und die Anzahl der Parameter deutlich reduzieren

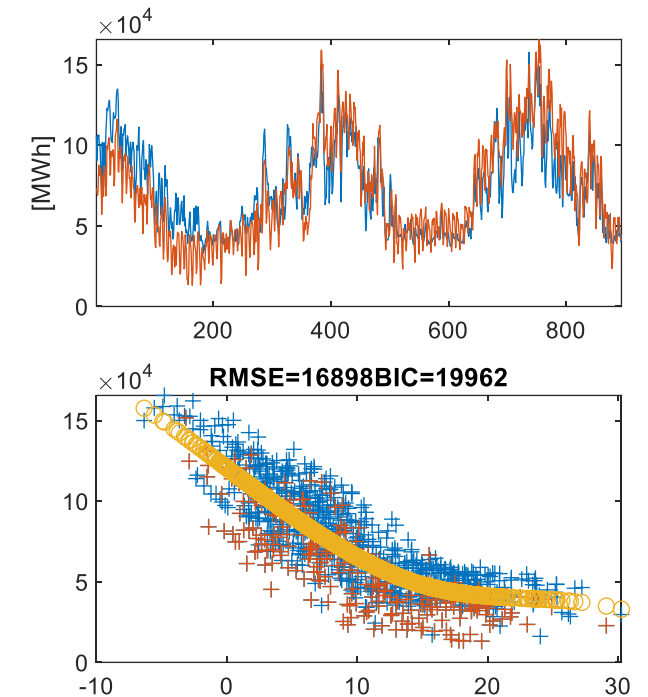
Einfache lineare Regression



Einfacher linearer Spline



Kubischer Spline



Hinweis: „Natürliche Splines“ fügen vor dem ersten und letzten Knoten ein lineares Segment ein, um die Varianz an den Rändern zu begrenzen

Glättung-Splines

- Generelles Ziel: Auffinden einer Funktion **g (cubic Spline)**, welche die Summe der quadrierten Residuen **RSS** minimiert

$$MIN RSS = \sum_{i=1}^n (y_i - g(x_i))^2$$

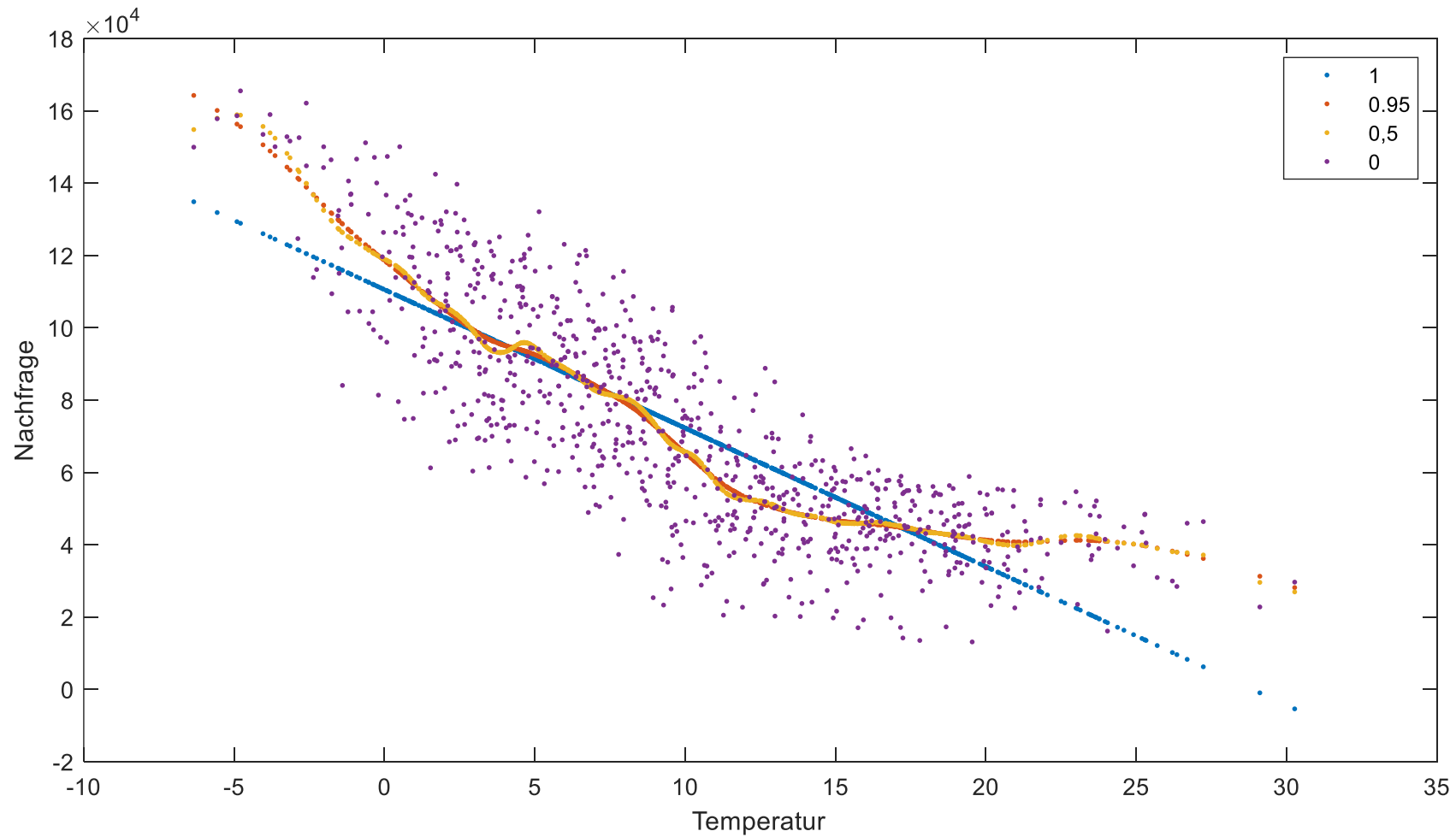
- Die Funktion **g** ist ein kubischer Spline mit **n** Stützstellen (Anzahl der unique Datenpunkte)
- Ohne weitere Restriktionen würde g so gewählt werden, dass RSS Null ergibt (n Freiheitsgrade). -> Overfitting

$$\sum_{i=1}^n \underbrace{(y_i - g(x_i))^2}_{\text{Fehler}} + \lambda \underbrace{\int g''(t)^2 dt}_{\text{Straffunktion}}$$

- Zur Minimierung der Freiheitsgrade wird ein Strafterm eingeführt, welcher die zweite Ableitung der Funktion g aufnimmt. Hierdurch werden Funktionen „bevorzugt“, welche möglichst glatt sind.
 - Die zweite Ableitung ist ein Maß für die Glattheit
 - λ ist der Gewichtungsfaktor des Strafterms
 - Das Integral des Strafterms bedeutet, dass die „Glattheit“ über den gesamten Wertebereich gemessen wird (addiert)

RSS = Residual Squared Sum

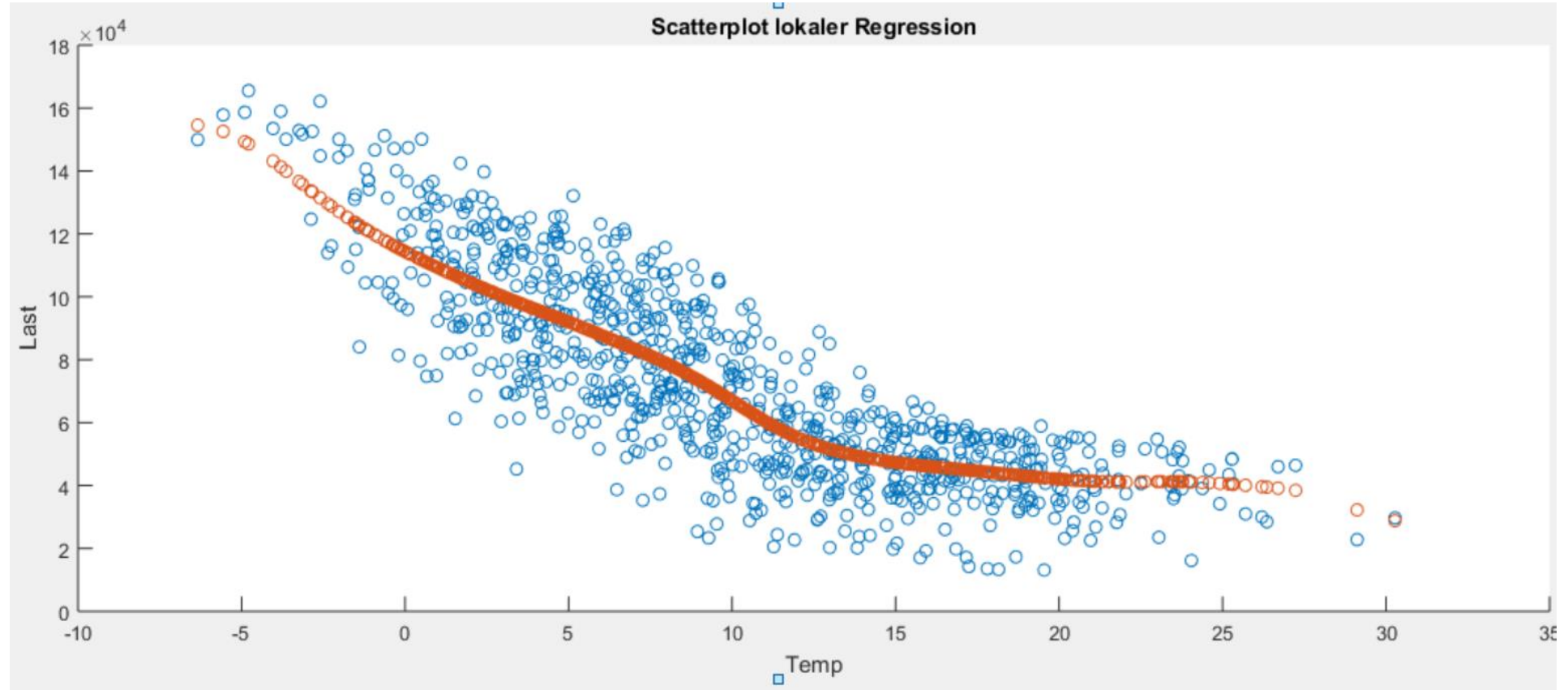
Wirkungsweise des Glättungsparameter λ



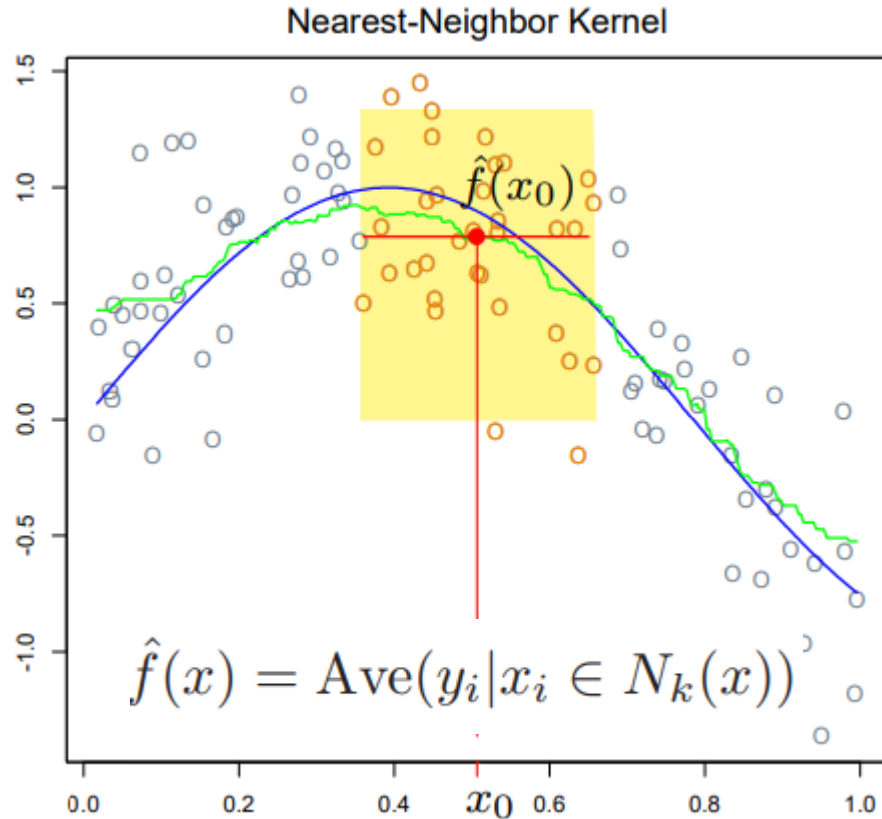
1	Polynome und Splines
2	Lokale Regression

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.

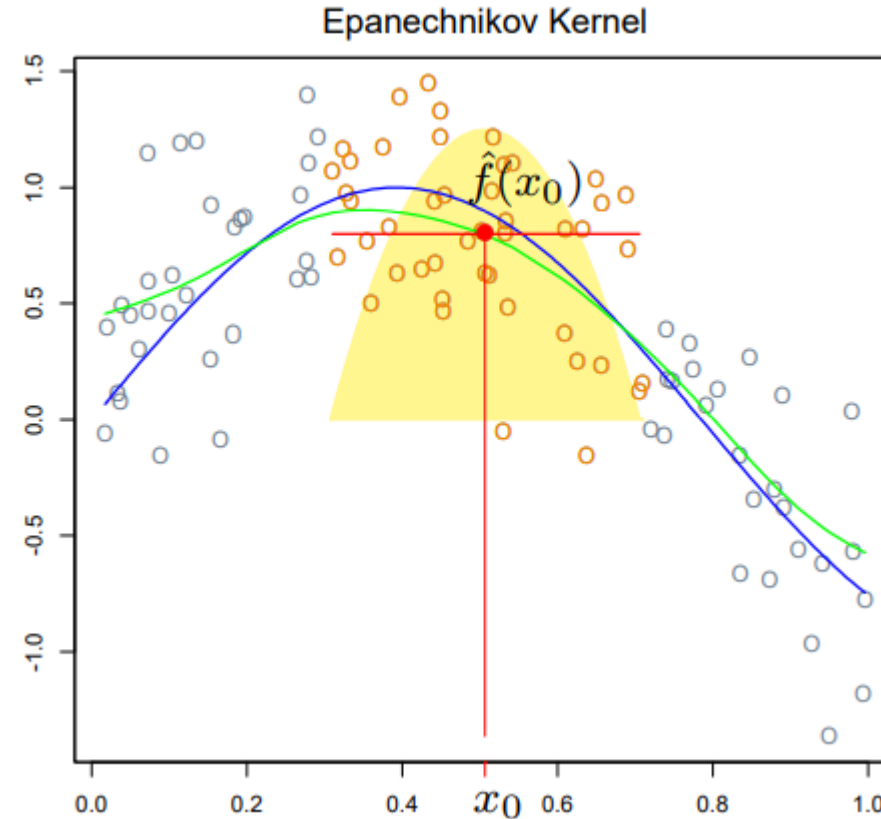
Einführung in die lokale Regression



Als Grundidee wird eine Gewichtungsfunktion in die Regression eingeführt



- $\hat{y}(x_0)$ wird aus dem Mittelwert aller y-Werte im gelben Fenster gebildet.
- Bei Variation des Fensters kommt es zu Sprüngen von \hat{y} .



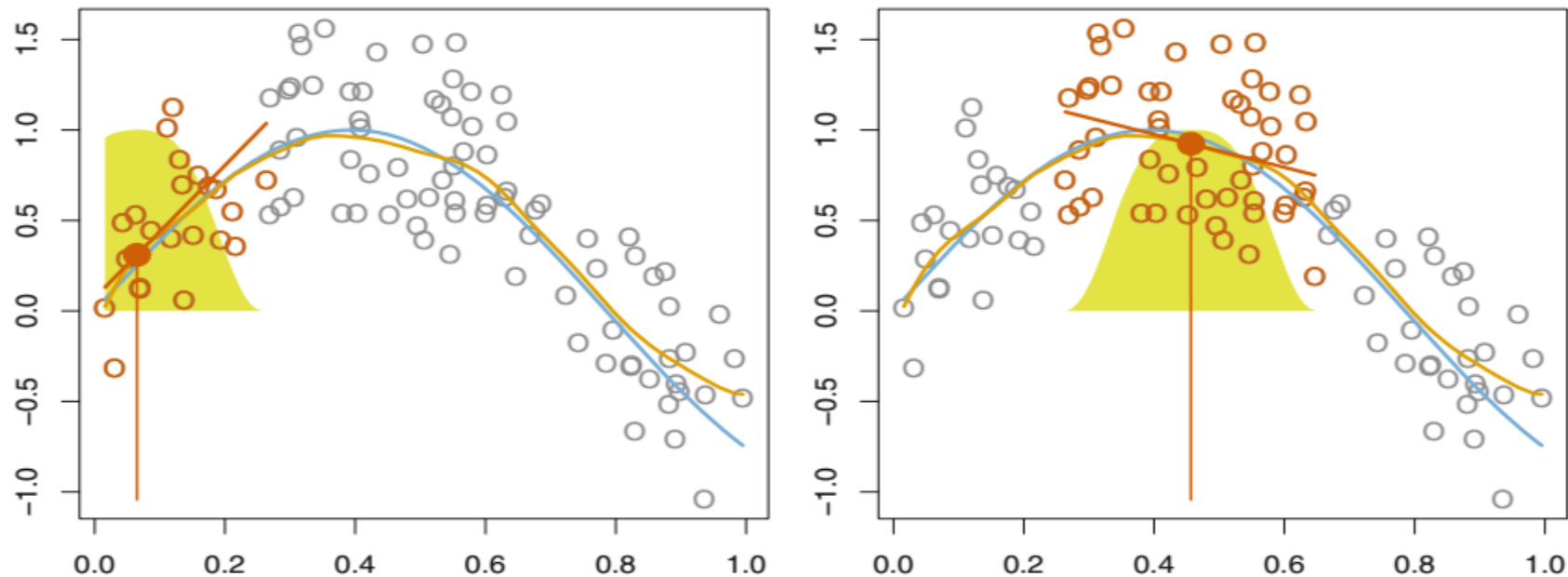
- $\hat{y}(x_0)$ wird aus dem gewichteten Mittelwert aller y-Werte im gelben Fenster gebildet.
- Der Verlauf von \hat{y} ist glatt.

Quelle: The Elements of Statistical Learning, Hastie et.al.

Die Lokale Regression kann eine stetige nichtlineare Funktion sehr genau approximieren

- Die lokale Regression ist die Anwendung einer Regressionsfunktion, wobei an jedem einzelnen Punkt x_0 insb. die unmittelbar umliegenden Datenpunkte in der Regression verwendet werden.
- Die umliegenden Datenpunkte werden mit einem Kernel (Dichtefunktion = gelbes Fenster) gewichtet.
- An jedem Punkt entspricht die Schätzung für \hat{y} der Auswertung der Funktion $f(x_0)$.

Local Regression



Quelle: The Elements of Statistical Learning, Hastie et.al.

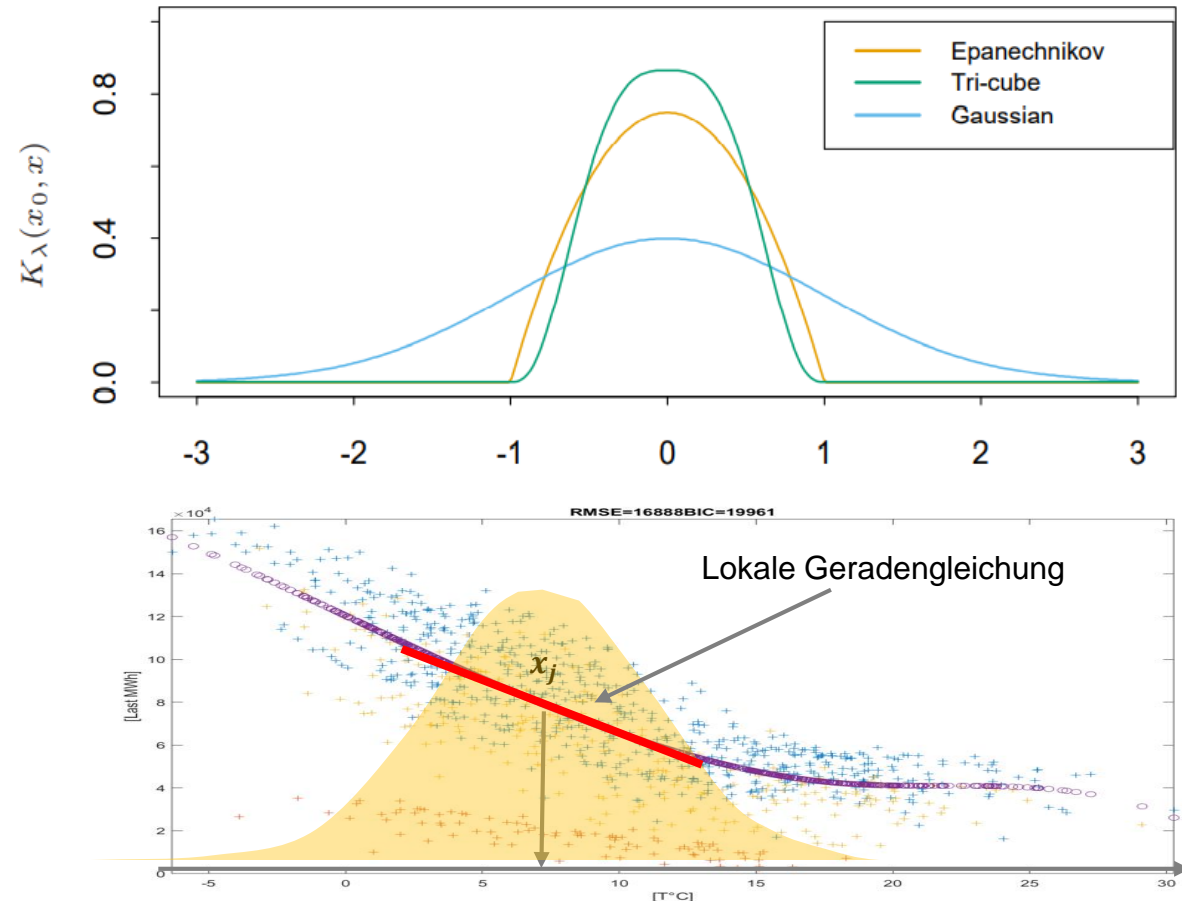
Für die Spezifikation der lokalen Regression müssen zwei grundsätzliche Spezifikationen vorgenommen werden

■ Wahl der Gewichtungsfunktion:

- Die Gewichtungsfunktion K bestimmt mit welchem Einfluss die Datenpunkte in der Zielfunktion zur Anpassung der lokalen Regression herangezogen werden.
- Auf der X-Achse ist ein Abstand aller Datenpunkte zum aktuellen Punkt x_0 betrachtet.
- Dieser Abstand kann beispielsweise die zeitliche saisonale Entfernung von x zu x_0 darstellen.

■ Wahl der Regressionsform:

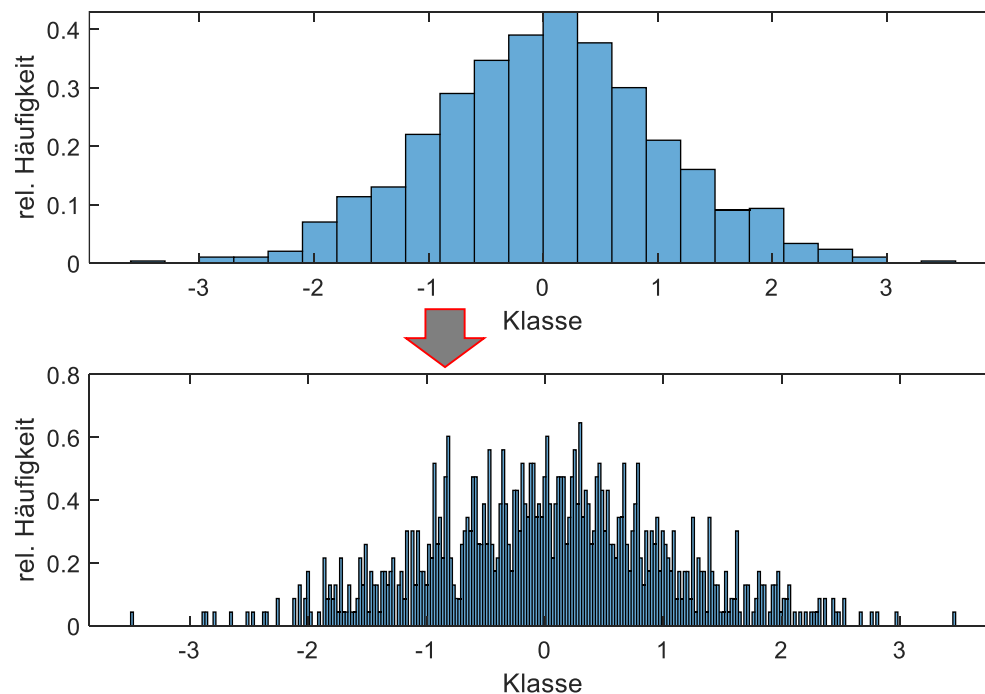
- Grundsätzlich kommen hier die klassischen Polynome nullter bis n-ter Ordnung in Frage.



Hintergrund: mit Hilfe eines Kerndichteschätzers kann aus einer Stichprobe eine stetige Dichtefunktion abgeleitet werden (nähere Erläuterungen siehe Anhang)

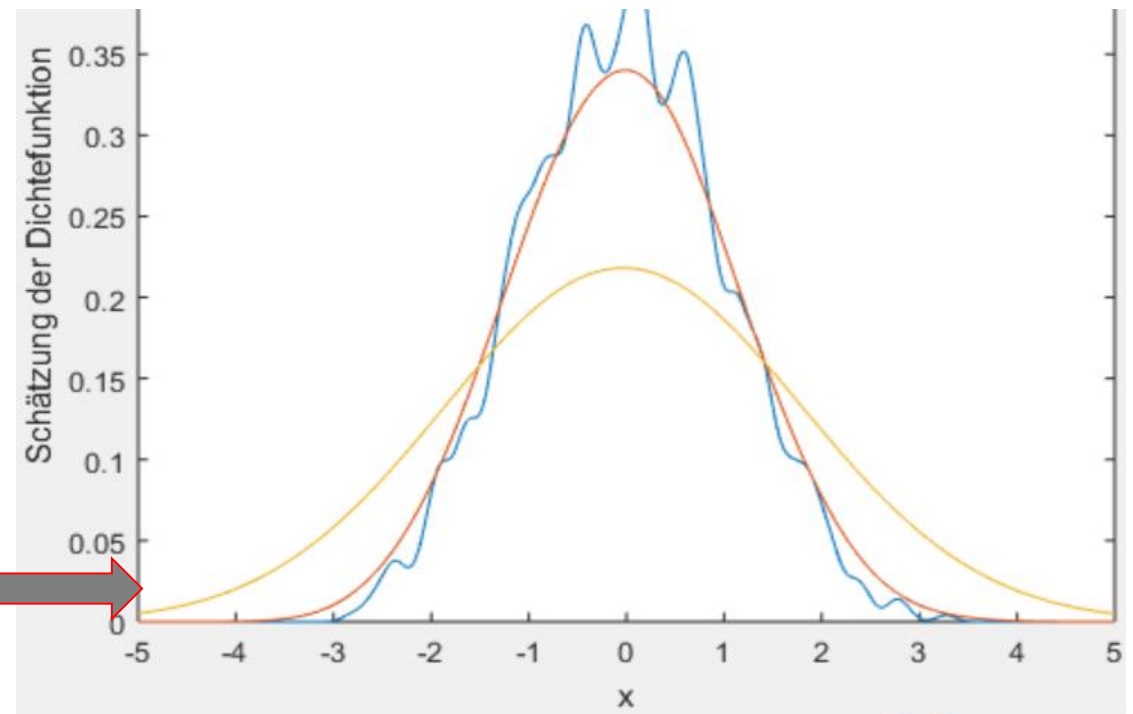
Histogramm (diskrete relative Häufigkeiten)

Ausgangsproblem: mit Hilfe eines Histogramms kann bei Verkleinerung der Klassenbreite keine stetige Dichtefunktion abgeleitet werden. Die Verkleinerung des Intervalls bewirkt im Gegenteil größere Unstetigkeit.



Stetige Dichtefunktion abgeleitet mit einem Kerndichteschätzer und Bandbreite $h = 0.1, 0.5, 1$

Mit Hilfe eines Kerndichteschätzers kann dieses Problem behoben werden. Der Parameter „Bandbreite“ beeinflusst die Form der Dichtefunktion. Ziel der Wahl sollte ein möglichst „glatter“ Verlauf der Dichtefunktion sein.



Die mathematische Eigenschaften eines Kerndichteschätzers werden auch für die Gewichtung der Datenpunkte im Rahmen der Regression verwendet

Eigenschaften des Kernels

- Der Kernel K stellt eine Gewichtungsfunktion der Beobachtungen mit folgenden Eigenschaften dar:
 - $\int K(u) du = 1$
 - $K(u) \geq 0$
- Aus der Klasse von Funktionen, die diese Anforderung erfüllen, wählen wir beispielsweise die Dichtefunktion der Standardnormalverteilung:

unbeschränktem Träger	
Kern	$K(u)$
Gauß	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$

Es gibt noch viele weitere Kernels neben der Gaussfunktion

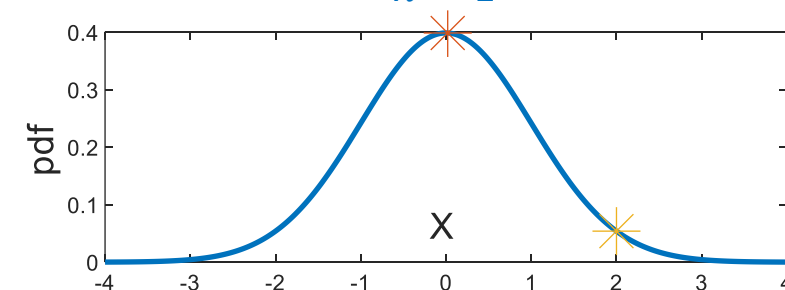
Kernelfunktionen

- Aufbau eines Kerndichteschätzers

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\underbrace{\frac{x - x_j}{h}}_u\right)$$

- Bandweite h : je größer h gewählt wird, desto mehr Gewicht erhalten auch Daten welche deutlich von x_i abweichen

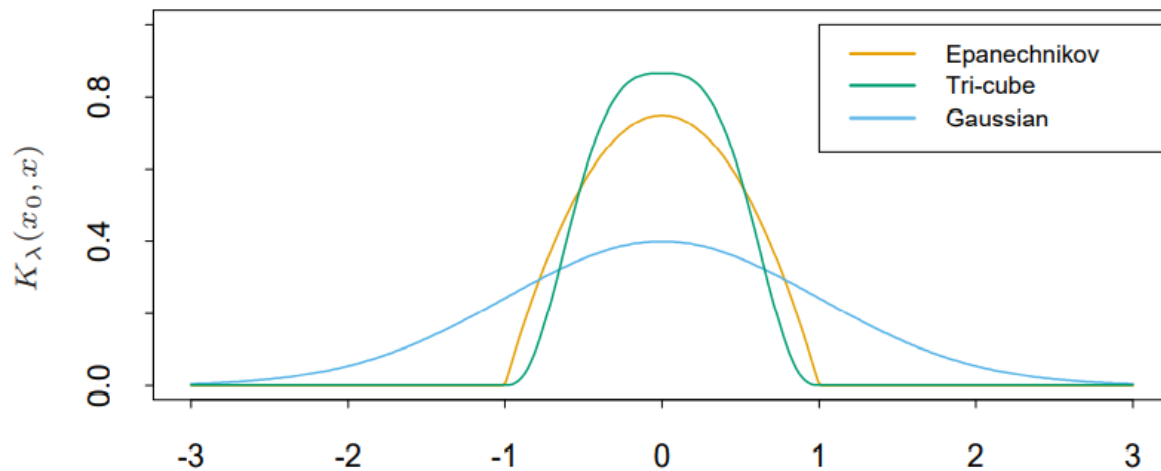
$$\frac{x - x_j}{h = 100} = 0,02 \quad \frac{x - x_j}{h = 1} = 2$$



h wird auch Smoothing-Parameter genannt

Neben dem Gauss-Kernel gibt es weitere Alternativen, die eine deutliche Einengung der zu betrachtenden Datenpunkte vornehmen

Kernelfunktionen



- In einem ersten Schritt wird für jeden Punkt eine normierte Variable u gebildet:

$$u = \frac{x - x_j}{h}$$

- Ein großes h bedeutet eine geringere Varianz (Mittelwertbildung über mehr Beobachtungen), aber auch eine höhere Verzerrung (wir gehen im Wesentlichen davon aus, dass die wahre Funktion innerhalb des Fensters konstant ist).

- Alternative Kernelfunktionen K :

- Epanechnikov:

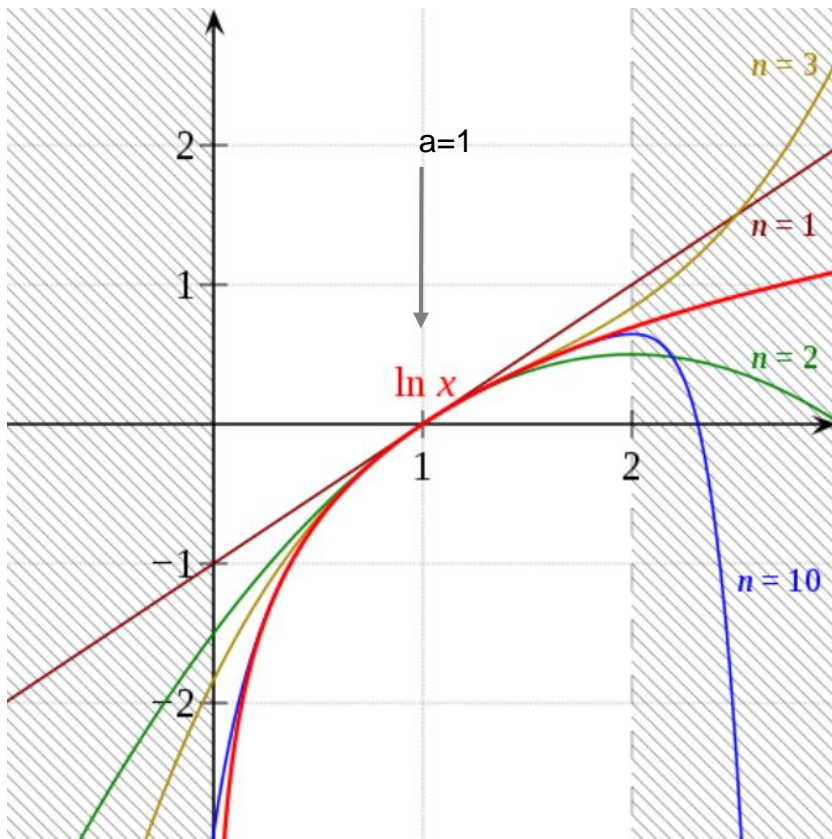
$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{wenn } |u| \leq 1 \\ 0, & \text{sonst} \end{cases}$$

- Tri-Cube:

$$K(u) = \begin{cases} (1 - |u|^3)^3, & \text{wenn } |u| \leq 1 \\ 0, & \text{sonst} \end{cases}$$

Die lokale Regression setzt auf dem Gedankenkonstrukt der Taylorreihenapproximation auf

Taylorreihenapproximation



Quelle: wikipedia

Erläuterung

Zielsetzung Taylorreihenapproximation:

- Entwicklung der Funktion um eine Stelle a mit Ableitungen bis zur n -ten Ordnung

$$Tf(x; a) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$$

- Taylorreihe 1. Ordnung

$$f(x, a) = f(a) + \underbrace{f'(a)(x - a)}_{\text{Lineare Regression}} + \text{Rest}(x, a)$$

Lineare Regression

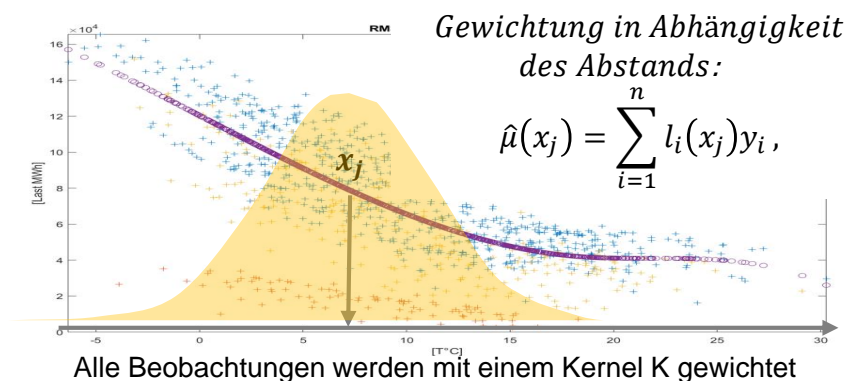
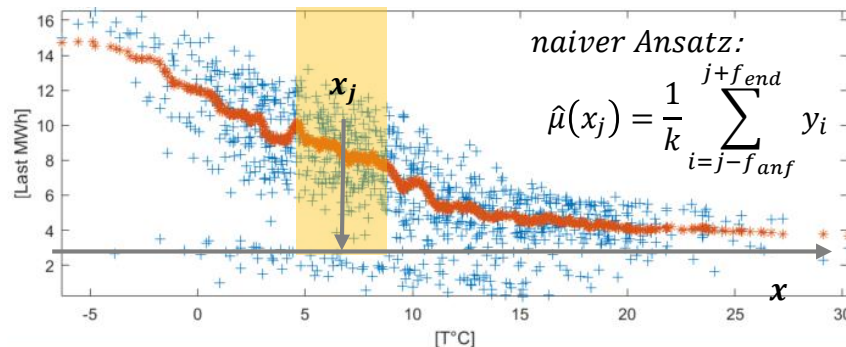
- Problem: bei nichtlinearen Funktionsverläufen ist die Linearisierung schnell mit großen Abweichungen verbunden!

Bei nichtlinearen Funktionsverläufen ist die Linearisierung schnell mit großen Abweichungen verbunden!

Der Nadaraya Watson Schätzer ist eine Taylorreihenapproximation 0-ter Ordnung $f(x, a) = f(a)$

Der Kernel hat gegenüber einem „rollierenden“ Fenster den Vorteil der „smootheren“ Hereinnahme neuer Datenpunkte

Es werden alle Datenpunkte im gelben Fenster um x_j verwendet, um $E[y_j] = \hat{\mu}(x_j)$ zu bestimmen. Das Fenster rolliert über den Datenbereich x .



- Der Schätzer für $\hat{\mu}(x_j)$ an jeder Stelle x_j wird durch folgenden Zusammenhang dargestellt:

Zielfunktion:

$$\hat{\mu}(x_j) = \text{Min} \sum_{i=1}^n \underbrace{(y_i - \beta_{1,j})^2}_{\text{Fehler}} \cdot \underbrace{K_h(x_j - x_i)}_{\text{Gewichtung des Datenpunktes}}$$

Lösung*:

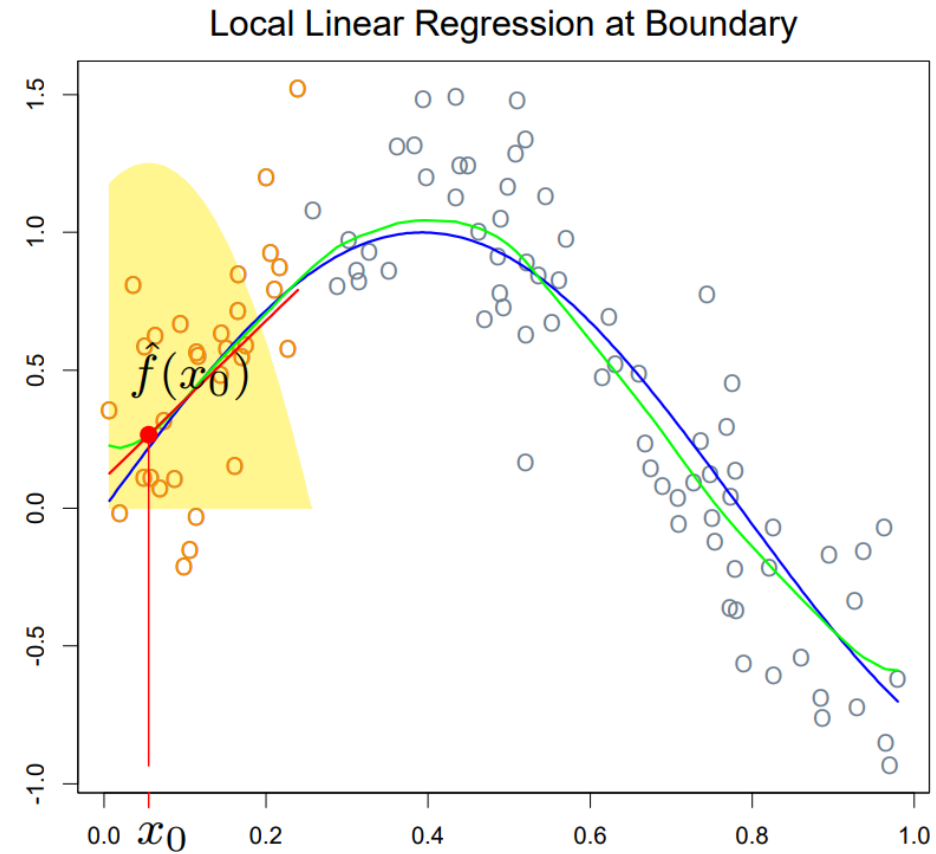
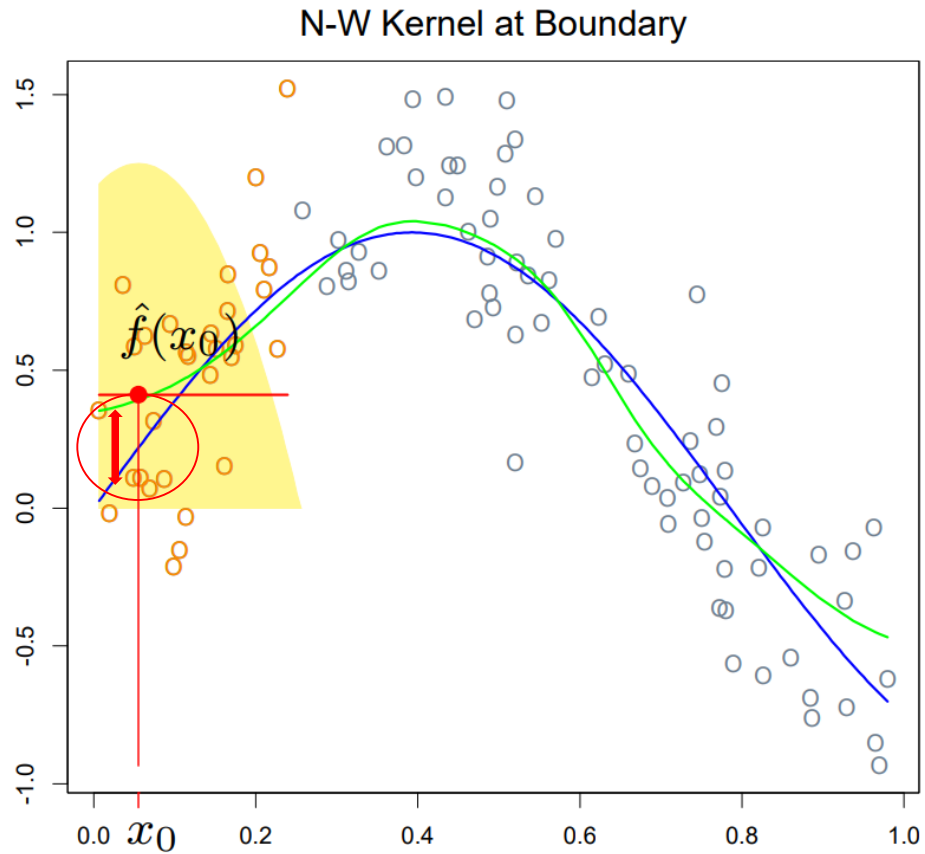
$$\hat{\mu}(x_j) = \beta_{1,j} = \frac{\sum_{i=1}^n y_i \cdot K_h(x_j - x_i)}{\sum_{i=1}^n K_h(x_j - x_i)}$$

Kernel:

$$K_h\left(\frac{x_j - x_i}{h}\right) K_h(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u)^2\right)$$

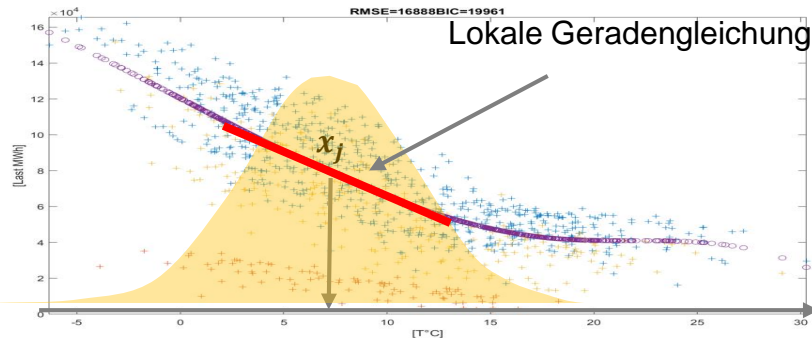
*Zielfunktion ableiten nach $\beta_{1,j}$ und Nullsetzen und auflösen nach $\beta_{1,j}$ führt zu der dargestellten Lösung

Der Nadaraya Watson Schätzer kann einen Bias in den Rändern von x aufweisen!



Die Taylorreihe erster Ordnung kann den Schätzer deutlich robuster ggü. Veränderungen der Bandbreite h machen

Darstellung Schätzung



$$W = \begin{pmatrix} K_h(x_j - x_n) & 0 & 0 \\ 0 & K_h(x_j - x_i) & 0 \\ 0 & 0 & K_h(x_j - x_n) \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_j - x_1 \\ \vdots & \vdots \\ 1 & x_j - x_n \end{pmatrix}$$

- Übertragung der Taylorreihenapproximation:

$$\underbrace{y_i = \mu(x_j) + \frac{\partial \mu}{\partial x_j} (x_i - x_j)}_{\text{Geradengleichung}}$$

- Zielfunktion an jeder Stelle x_j

$$\text{Min} \sum_{i=1}^n (y_i - \beta_{1,j} - \beta_{2,j}(x_i - x_j))^2 K_{j,i,h}$$

- Lösung in Matrixschreibweise:

$$\begin{pmatrix} \hat{\beta}_{1,j} \\ \hat{\beta}_{2,j} \end{pmatrix} = (X^T W X)^{-1} X^T W Y$$

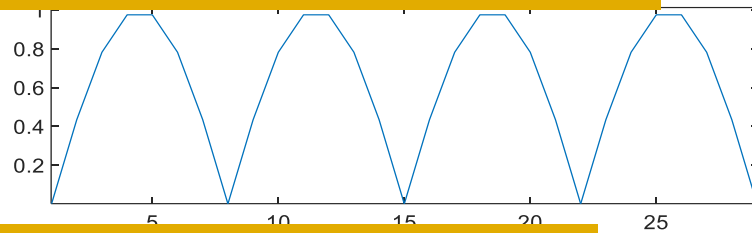
- Gesuchter Erwartungswert:

$$\hat{\mu}(x_j) = \hat{\beta}_{1,j}$$

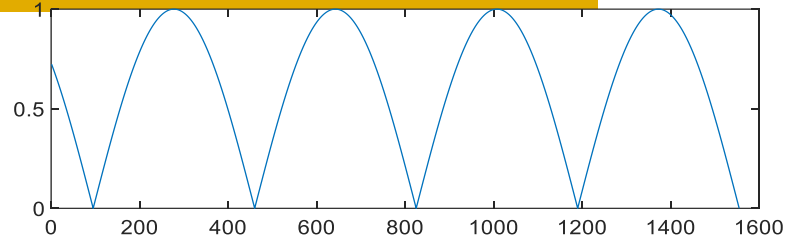
$$\text{Ableitung der Zielfunktion ergibt: } X^T W \left(Y - X \begin{pmatrix} \hat{\beta}_{1,j} \\ \hat{\beta}_{2,j} \end{pmatrix} \right) = 0$$

Im Rahmen einer mehrdimensionalen lokalen Regression muss in einem ersten Schritt die euklidische Norm des Vektors u gebildet werden

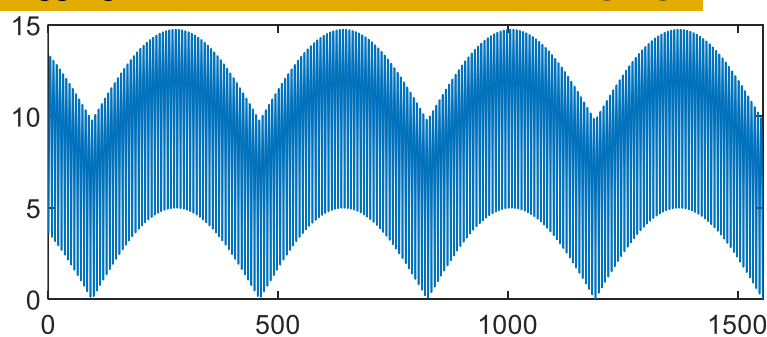
erstes Distanzmaß: Wochentagssaisonalität



zweites Distanzmaß: Jahressaisonalität



Aggregation zu Grund – und Oberschwingung



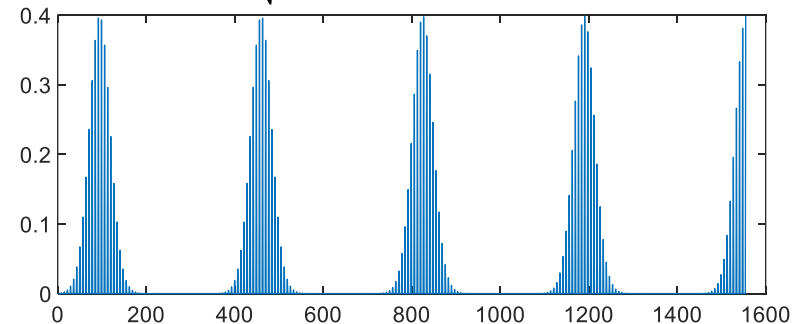
■ Regressionsmodell (dargestellt mit zwei Regressoren)

$$\min_{\beta} \sum_{i=1}^n (y_i - \beta_{1,j} - \beta_{2,j}(x_{i1} - x_{j1}) - \beta_{3,j}(x_{i2} - x_{j2}))^2 K_{j,i,h}$$

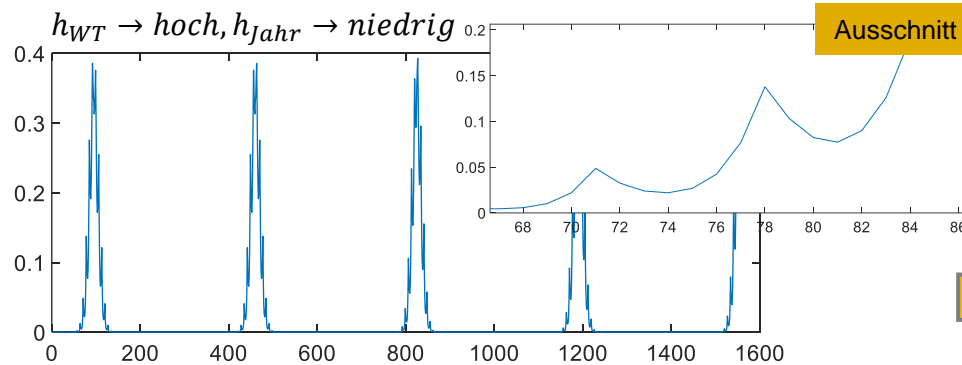
■ Die Kerndichtefunktion verarbeitet nun einen multidimensionalen Vektor mit Hilfe der euklidischen Norm

$$K_{j,i,h}(u) = K\left(\frac{\|x_{norm,j} - x_{norm,i}\|}{h}\right)$$

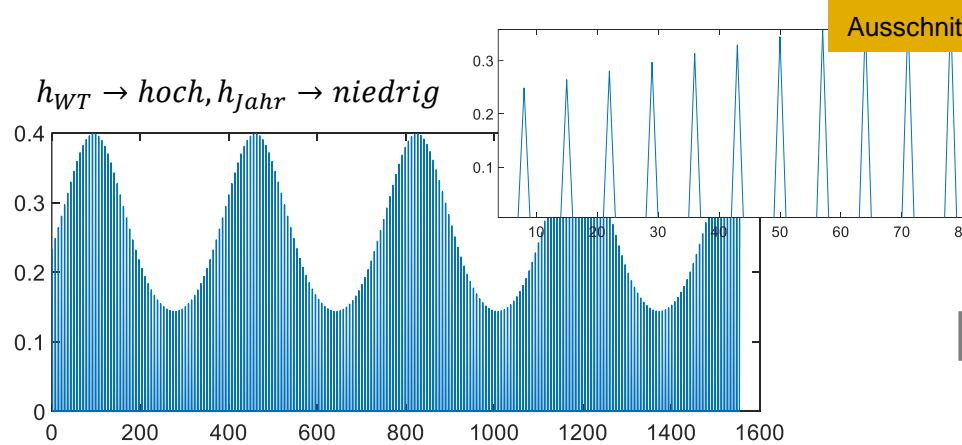
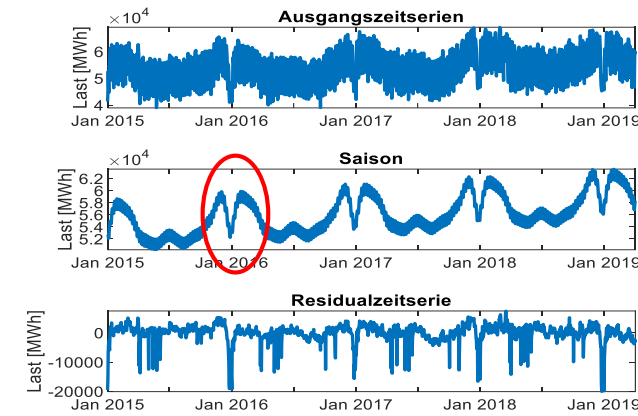
■ mit $\|v_i\| = \sqrt{\sum_{k=1}^p (x_{norm,i,k} - x_{norm,i,k})^2}$



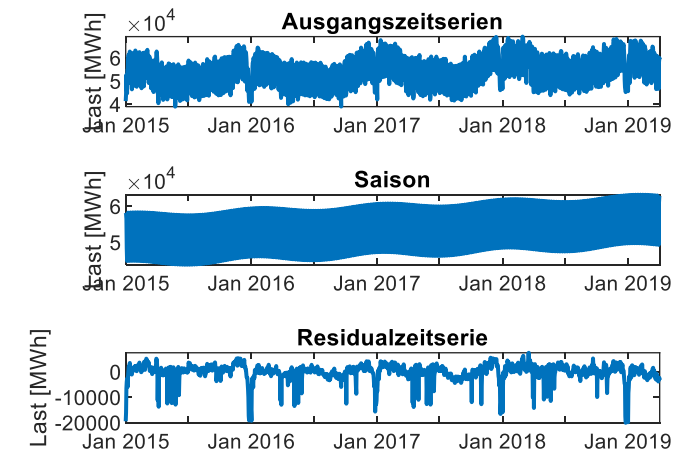
Einfluss der Bandbreite auf die Gewichtung der Datenpunkte im Kerndichteschätzer K



Auswirkung: ein gleicher Wochentag 365 Tage entfernt hat ein höheres Gewicht als 1 Woche entfernt



Ein gleicher Wochentag 365 Tage entfernt hat ein höheres Gewicht als 1 Woche entfernt



Weitergehende Einflussgrößen

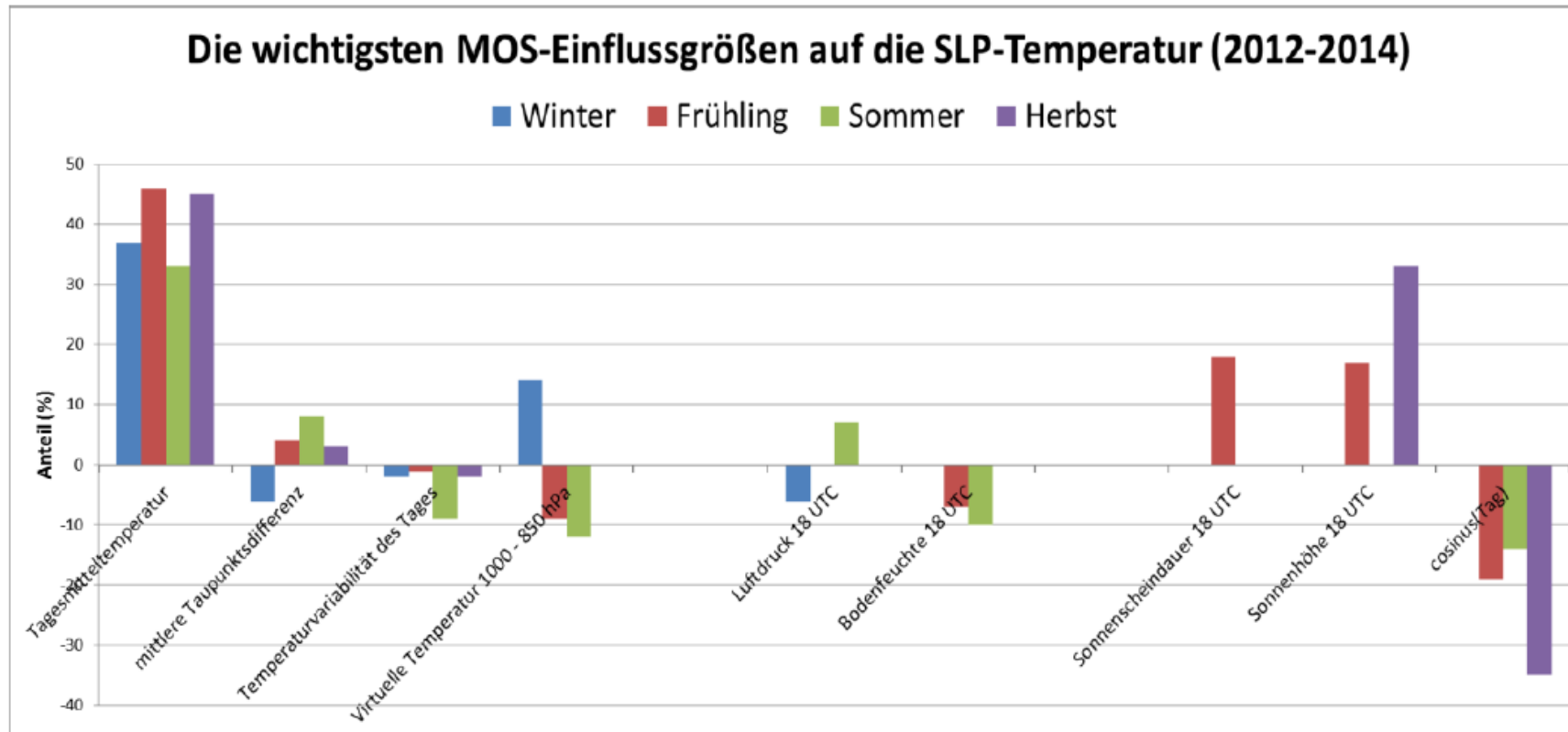
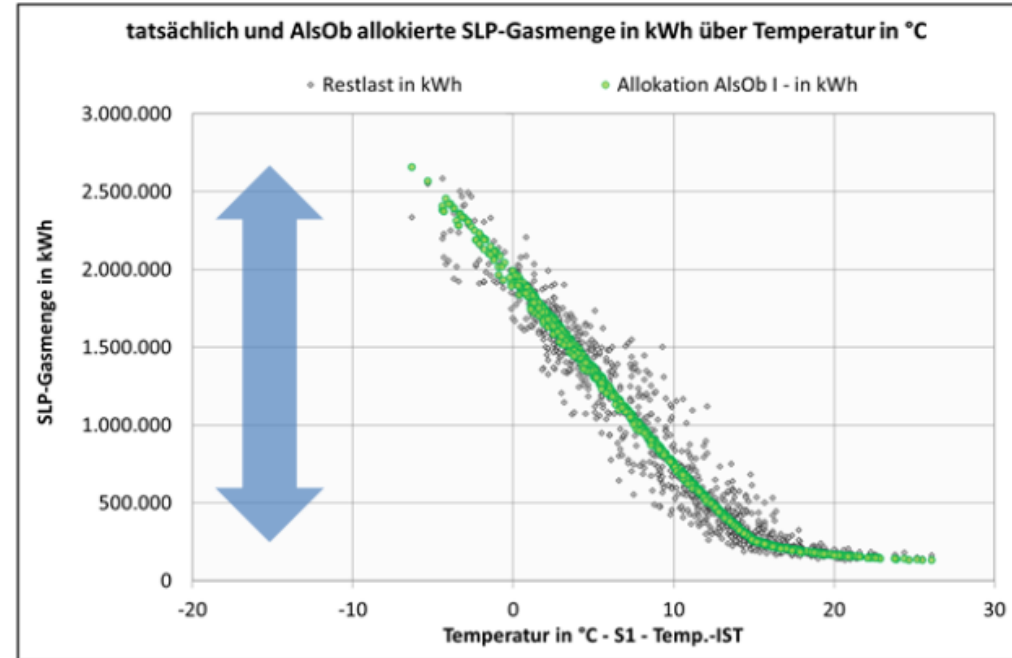


Abbildung 2 Einflussgrößen auf das Gasverbrauchsverhalten. Quelle: DWD

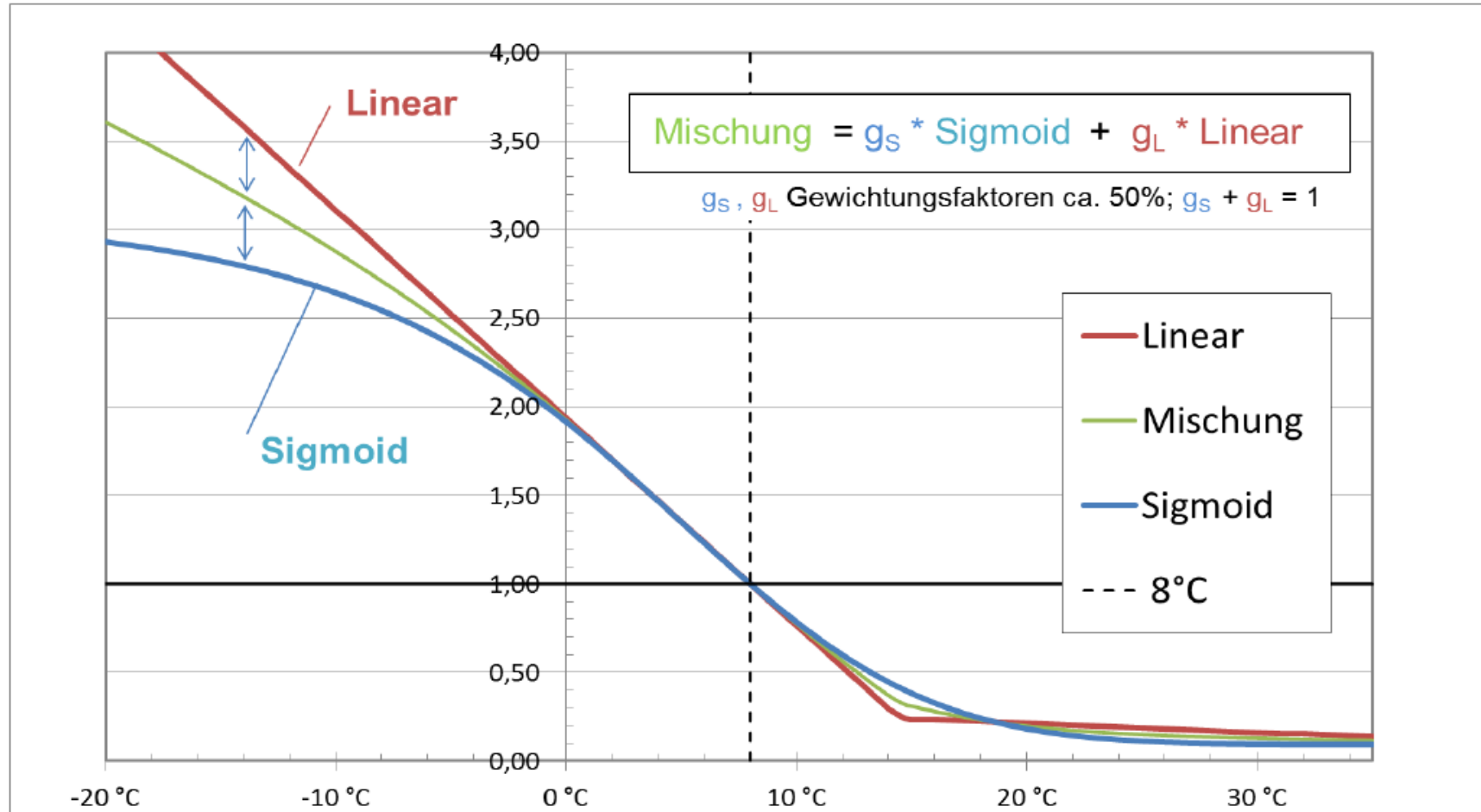
Allokationen von SLP-Profilen

Kundenwerte zur Abstimmung der Allokationsmenge mit der tatsächlichen Kundenmenge (Jahresmenge / Tagesleistung); Wert zur Skalierung der y-Achse

Knickpunkt-Temperatur zur Abstimmung des Übergangspunktes Heizbereich / Warmwasserbereich der Profilfunktion mit dem tatsächlichen Kundenverhalten im Netzgebiet; Wert zur Skalierung der x-Achse



Neues Verfahren zur Bestimmung der SLP-Profile

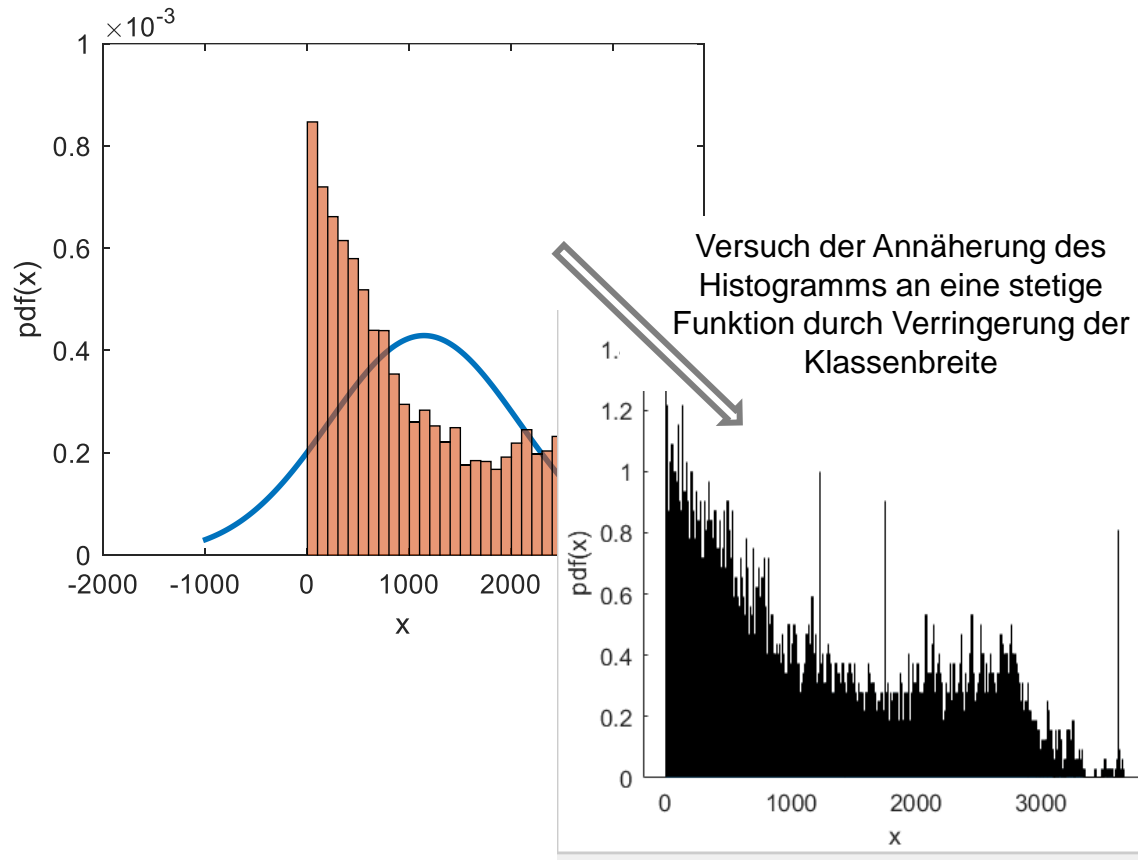


Anhang: Kerndichteschätzer



Bestimmte empirische Verteilungen lassen sich mit den bekannten parametrischen Dichtefunktionen nur unzureichend annähern.

Verteilung der Wind-Offshore-Einspeisung ggü. angepasster Normalverteilung

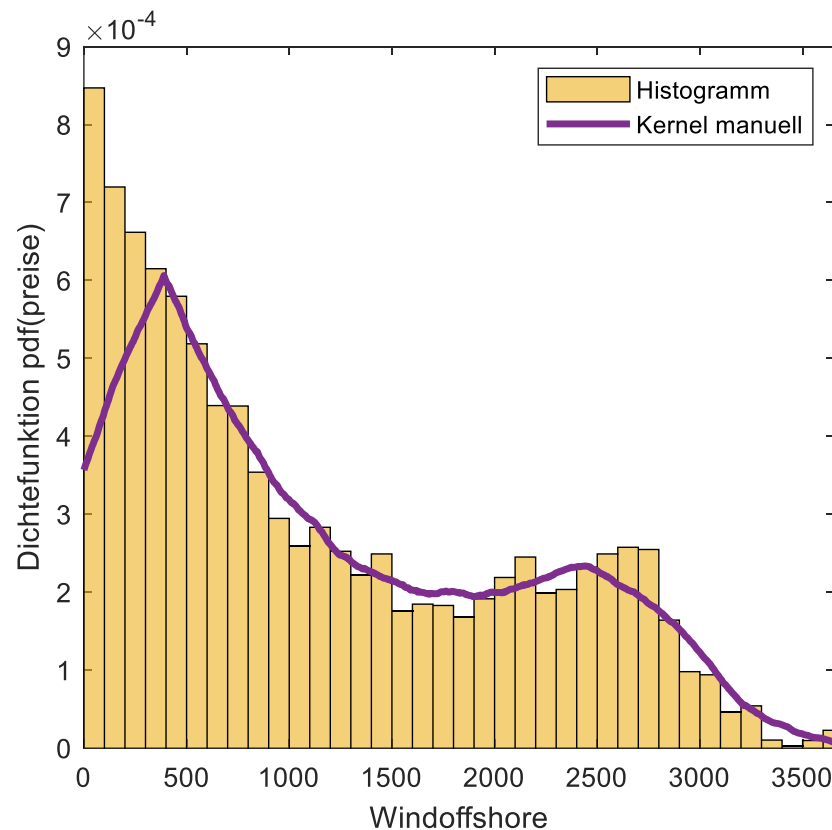


Erläuterung

- In bestimmten Fällen sind die zur Verfügung stehenden parametrischen Dichtefunktionen nicht gut an die empirische Verteilung angepasst.
 - Die Schwäche des Histogramms als Approximation einer stetigen Dichtekurve $\text{pdf}(x)$ liegt in der „treppenartigen“ Gestalt.
 - Eine Verkleinerung der Klassenbreite wird in der Regel das Problem der „Stetigkeit“ nicht abmildern.
- In bestimmten Situationen sind einzelne Klassen unbesetzt oder überdurchschnittlich stark besetzt.

Idee eines Kerndichteschätzers: die Klassen werden überlappend um x gebildet

Idee: Konstruktion eines gleitendes Histogramm



Erläuterung

1. Bildung Intervall $[2h]$ um beliebigen x -Wert: $[x - h, x + h)$
2. Approximation der Dichtekurve an der Stelle x durch:

$$\widehat{pdf}(x) = \frac{\frac{1}{n} \cdot \text{Anzahl } x_i \text{ in } [x - h, x + h)}{2h}$$

- Mathematische Umsetzung von „Anzahl ...“ durch:

$$\frac{1}{h} K\left(\frac{x - x_i}{h}\right) = \begin{cases} \frac{1}{2h} & x - h \leq x_i \leq x + h \\ 0 & \text{sonst} \end{cases}$$

- Mathematische Abbildung der Dichtefunktion

$$\widehat{pdf}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - x_i}{h}\right)$$

- Nachteil: Funktion ist immer noch unstetig

Die mathematische Eigenschaften eines Kerndichteschätzers werden auch für die Gewichtung der Datenpunkte im Rahmen der Regression verwendet

Kernelfunktionen

- Aufbau eines Kerndichteschätzers

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^n K_h(x - x_j) :$$

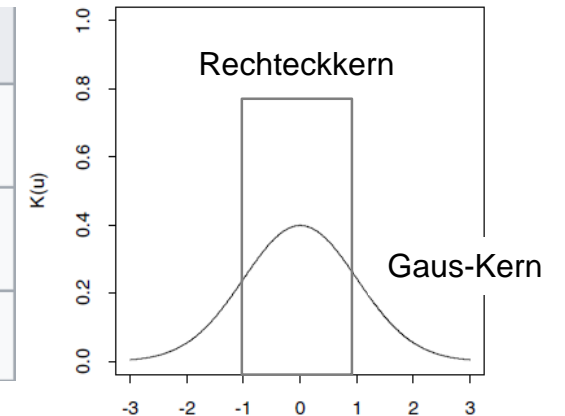
$$\frac{1}{nh} \sum_{j=1}^n K \left(\underbrace{\frac{x - x_j}{h}}_u \right)$$

- h stellt die Bandweite dar; je größer h gewählt wird, desto mehr Gewicht erhalten auch Daten welche deutlich von x_i abweichen

Eigenschaften des Kernels

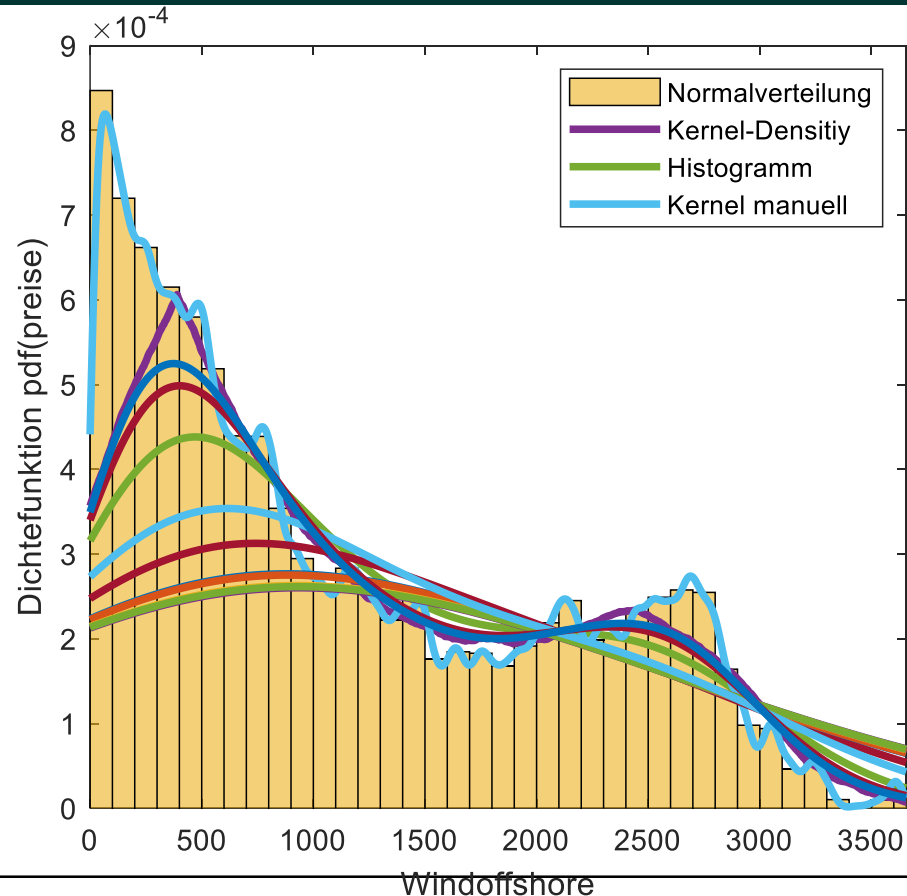
- Der Kernel stellt eine Gewichtungsfunktion der Beobachtungen dar mit folgenden Eigenschaften
 - $\int K(u) du = 1,$
 - $K(u) \geq 0$
 - Es gibt eine große Klasse von Funktionen, welche die oben geforderten Eigenschaften erfüllen:

Kern	$K(u)$
Gauß	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$
Cauchy	$\frac{1}{\pi(1+u^2)}$
Picard	$\frac{1}{2} \exp(- u)$



Die resultierende Dichtefunktion ist abhängig von der Wahl der Schrittweite

Abhängigkeit der Dichte von h



Zusammenfassung

- Im Vergleich zum Rechteckkern werden durch die Form von Epanechnikov-, Bisquare- und Gauß-Kern näher bei x liegende Datenpunkte stärker gewichtet, als weiter von x entfernte.
- Für großes h sind die Fenster weit, die Kurve wird sehr glatt, aber wichtige Details können verschluckt werden.
- Ist umgekehrt h zu klein, wird die Kurve rauer, es treten jedoch möglicherweise Details hervor, die nur mit der Zufälligkeit der Stichprobe etwas zu tun haben.

→ *Overfitting*