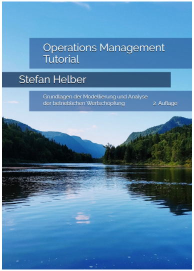
Prognoserechnungen

Problemaspekte

Prof. Dr. Stefan Helber



Gegenstand von Prognosen

Beispiele von Prognosen im Operations Management

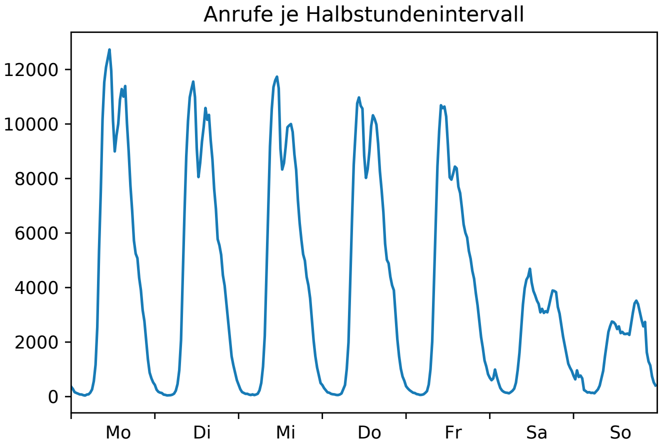
- Bedarfe
- Preise
- Wartezeiten
- Fahrzeiten
- Standzeiten

Individuelle Zufälligkeit, kollektive Vorhersehbarkeit I

Beispiel: Anrufaufkommen im Call Center

- Mikromodelle und Aggregation?
- Makromodelle unkoordinierten Verhaltens?

Individuelle Zufälligkeit, kollektive Vorhersehbarkeit II



Grundidee von Prognoserechnungen: Stabilität datenerzeugender Prozesse I

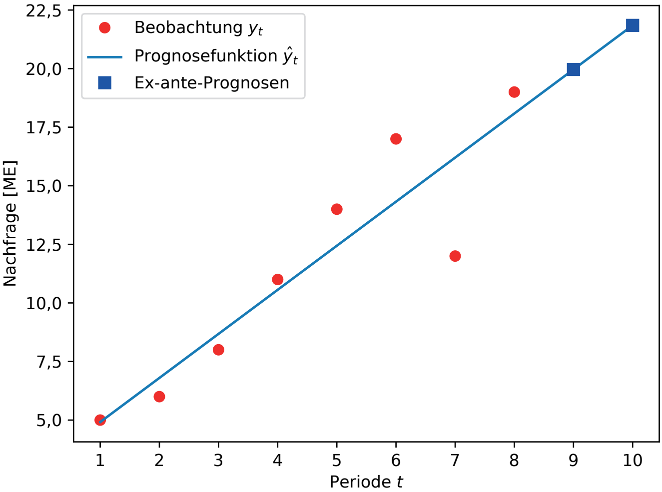
Annahmen

1. Existenz eines **datenerzeugenden Prozesses**
2. **Gesetzmäßigkeit** kann **identifiziert** und durch Prognosemodell beschrieben werden
3. Prozess ist **stabil**, kein Strukturbruch

Prognoserechnung

- Ex-post-“Prognose”
- Ex-ante-Prognose

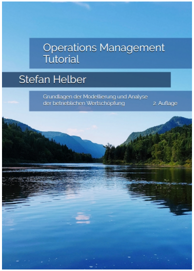
Grundidee von Prognoserechnungen: Stabilität datenerzeugender Prozesse II



Prognoserechnungen

Verwendungszwecke und Voraussetzungen von Prognoserechnungen

Prof. Dr. Stefan Helber

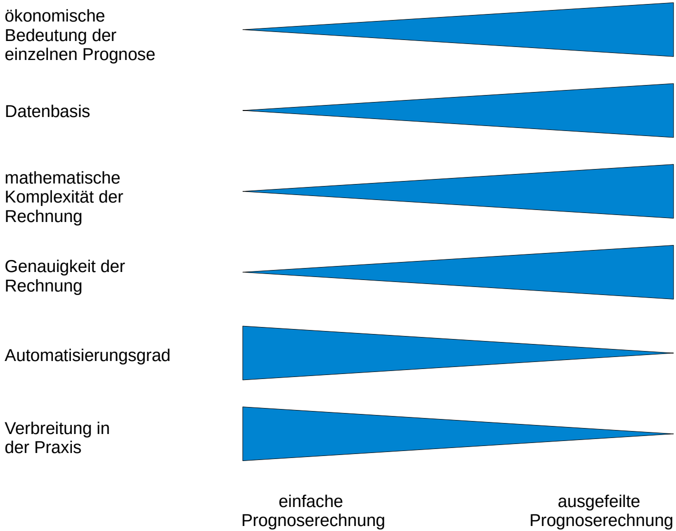


Verwendungszwecke

Grundlegende Fragen

- Zweck?
- Genauigkeit?
- Datenbasis?

Fallkonstellationen



Einfache Prognoserechnungen

Lagerhaltung geringwertiger Produkte über

- Bestellmengen und
- Bestellzeitpunkte

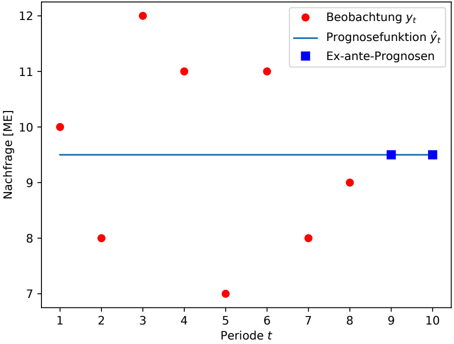
Bedarfsprognose

- computergestützt und automatisiert
- einfache Prognosemethoden
- geringe Anforderungen an die Datenbasis

Einfache Prognoserechnungen

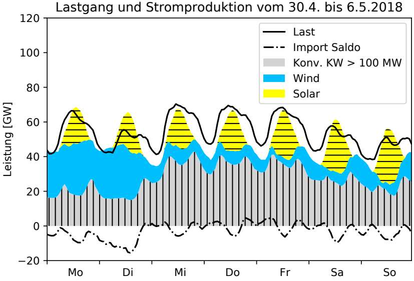
Beispiel: arithmetische Mittelwerte

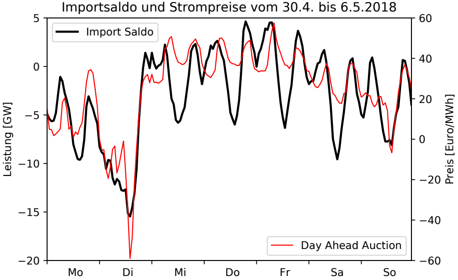
$$\hat{y}_{t_a+1} = \hat{y}_{t_a+2} = ... = \frac{1}{t_a} \sum_{t=1}^{t_a} y_t \qquad (1)$$



Vielfach verwendet in der Praxis!

Ausgefeilte Prognoserechnungen für Energiemärkte





Ausgefeilte Prognoserechnungen

Datenbasis für das Beispiel der Energiemärkte

- Wetterdaten und -prognosen
- Kalenderdaten
- Verfügbarkeit von Kraftwerken

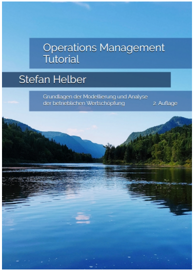
Instrumente

- Regressionsrechnungen
- Machine-Learning-Methoden
- ...

Prognoserechnungen

Elementare Prognosemodelle für Zeitreihen mit einem näherungsweise konstanten Niveau

Prof. Dr. Stefan Helber



Zeitreihen mit konstantem Niveau I

Arithmetischer Mittelwert

$$\hat{y}_{t_a+1} = \frac{1}{t_a} \sum_{t=1}^{t_a} y_t \tag{1}$$

Gleitender arithmetischer Mittelwert

$$\hat{y}_{t_a+1} = \frac{1}{N} \sum_{t=t_a-N+1}^{t_a} y_t \tag{2}$$

Exponentielle Glättung 1. Ordnung, Gewichtungsfaktor α mit $0 < \alpha < 1$

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1-\alpha) \cdot \hat{y}_t \tag{3}$$

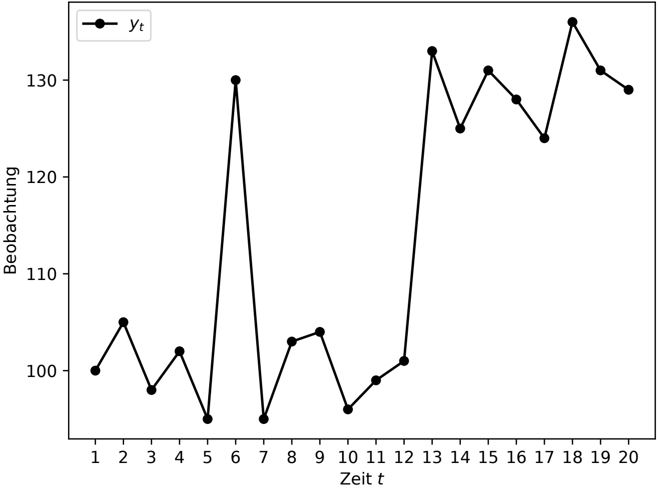
Zeitreihen mit konstantem Niveau II

Exponentielle Glättung 1. Ordnung, Gewichtungsfaktor α mit $0 < \alpha < 1$

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_t \tag{4}$$

$$\begin{aligned}
\hat{y}_{t+1} &= 0{,}2 \cdot y_t + 0{,}8 \cdot \hat{y}_t \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot (0{,}2 \cdot y_{t-1} + 0{,}8 \cdot \hat{y}_{t-1}) \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot 0{,}2 \cdot y_{t-1} + 0{,}8^2 \cdot \hat{y}_{t-1} \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot 0{,}2 \cdot y_{t-1} + 0{,}8^2 \cdot (0{,}2 \cdot y_{t-2} + 0{,}8 \cdot \hat{y}_{t-2}) \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot 0{,}2 \cdot y_{t-1} + 0{,}8^2 \cdot 0{,}2 \cdot y_{t-2} + 0{,}8^3 \cdot \hat{y}_{t-2} \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot 0{,}2 \cdot y_{t-1} + 0{,}8^2 \cdot 0{,}2 \cdot y_{t-2} + 0{,}8^3 \cdot (0{,}2 \cdot y_{t-3} + 0{,}8 \cdot \hat{y}_{t-3}) \\
&= 0{,}2 \cdot y_t + 0{,}8 \cdot 0{,}2 \cdot y_{t-1} + 0{,}8^2 \cdot 0{,}2 \cdot y_{t-2} + 0{,}8^3 \cdot 0{,}2 \cdot y_{t-3} + 0{,}8^4 \cdot \hat{y}_{t-3} \\
&= 0{,}2 \cdot y_t + 0{,}16 \cdot y_{t-1} + 0{,}128 \cdot y_{t-2} + 0{,}1024 \cdot y_{t-3} + 0{,}4096 \cdot \hat{y}_{t-3} \\
&= \ldots
\end{aligned}$$

Beispiel



y_t
130
120
110
100
Beobachtung
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Zeit t

Berechnung in Tabellenkalkulation I

	A	B	C	D	E	F	G	H	I	J
1				**(Gleitende) Durchschnitte**				**Exp. Glättung 1. Ordnung**		
2			**Perioden**	alle	6	3	**Alpha**	0,1	0,3	0,5
3										
4	t	$y(t)$		$p_{gda}(t)$	$p_{gd6}(t)$	$p_{gd3}(t)$		$p_{eg0.1}(t)$	$p_{eg0.3}(t)$	$p_{eg0.5}(t)$
5	1	100		NaN	NaN	NaN		NaN	NaN	NaN
6	2	105		100,00	NaN	NaN		100,00	100,00	100,00
7	3	98		102,50	NaN	NaN		100,50	101,50	102,50
8	4	102		101,00	NaN	101,00		100,25	100,45	100,25
9	5	95		101,25	NaN	101,67		100,43	100,92	101,13
10	6	130		100,00	NaN	98,33		99,88	99,14	98,06
11	7	95		105,00	105,00	109,00		102,89	108,40	114,03
12	8	103		103,57	104,17	106,67		102,10	104,38	104,52
13	9	104		103,50	103,83	109,33		102,19	103,97	103,76
14	10	96		103,56	104,83	100,67		102,37	103,98	103,88
15	11	99		102,80	103,83	101,00		101,74	101,58	99,94
16	12	101		102,45	104,50	99,67		101,46	100,81	99,47
17	13	133		102,33	99,67	98,67		101,42	100,87	100,23
18	14	125		104,69	106,00	111,00		104,58	110,51	116,62
19	15	131		106,14	109,67	119,67		106,62	114,85	120,81
20	16	128		107,80	114,17	129,67		109,06	119,70	125,90
21	17	124		109,06	119,50	128,00		110,95	122,19	126,95
22	18	136		109,94	123,67	127,67		112,26	122,73	125,48
23	19	131		111,39	129,50	129,33		114,63	126,71	130,74
24	20	129		112,42	129,17	130,33		116,27	128,00	130,87

Berechnung in Tabellenkalkulation II

	A	B	C	D	E	F
1				**(Gleitende) Durchschnitte**		
2			**Perioden**	alle	6	3
3						
4	*t*	*y(t)*		$p_{gda}(t)$	$p_{gd6}(t)$	$p_{gd3}(t)$
5	1	100		NaN	NaN	NaN
6	=A5+1	105		=MITTELWERT($B$5:B5)	NaN	NaN
7	=A6+1	98		=MITTELWERT($B$5:B6)	NaN	NaN
8	=A7+1	102		=MITTELWERT($B$5:B7)	NaN	=MITTELWERT($B5:$B7)
9	=A8+1	95		=MITTELWERT($B$5:B8)	NaN	=MITTELWERT($B6:$B8)
10	=A9+1	130		=MITTELWERT($B$5:B9)	NaN	=MITTELWERT($B7:$B9)
11	=A10+1	95		=MITTELWERT($B$5:B10)	=MITTELWERT($B5:$B10)	=MITTELWERT($B8:$B10)
12	=A11+1	103		=MITTELWERT($B$5:B11)	=MITTELWERT($B6:$B11)	=MITTELWERT($B9:$B11)
13	=A12+1	104		=MITTELWERT($B$5:B12)	=MITTELWERT($B7:$B12)	=MITTELWERT($B10:$B12)

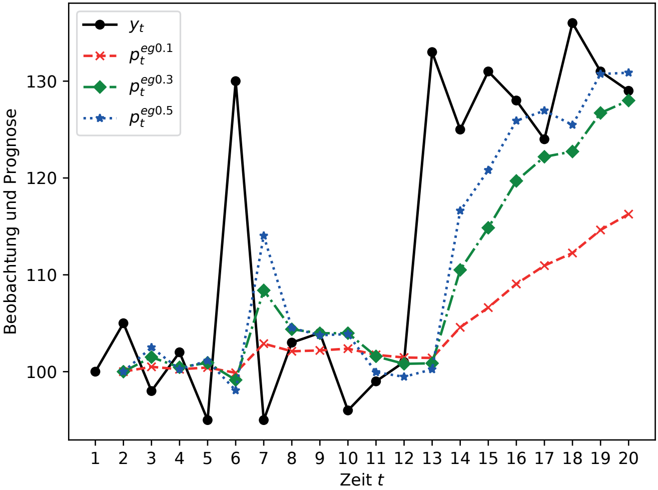
Berechnung in Tabellenkalkulation III

	A	B	G	H	I	J
1				**Exp. Glättung 1. Ordnung**		
2			**Alpha**	0,1	0,3	0,5
3						
4	*t*	*y(t)*		$p_{eg0.1}(t)$	$p_{eg0.3}(t)$	$p_{eg0.5}(t)$
5	1	100		NaN	NaN	NaN
6	=A5+1	105		=B5	=B5	=B5
7	=A6+1	98		=H$2*$B6+(1-H$2)*H6	=I$2*$B6+(1-I$2)*I6	=J$2*$B6+(1-J$2)*J6
8	=A7+1	102		=H$2*$B7+(1-H$2)*H7	=I$2*$B7+(1-I$2)*I7	=J$2*$B7+(1-J$2)*J7
9	=A8+1	95		=H$2*$B8+(1-H$2)*H8	=I$2*$B8+(1-I$2)*I8	=J$2*$B8+(1-J$2)*J8
10	=A9+1	130		=H$2*$B9+(1-H$2)*H9	=I$2*$B9+(1-I$2)*I9	=J$2*$B9+(1-J$2)*J9
11	=A10+1	95		=H$2*$B10+(1-H$2)*H10	=I$2*$B10+(1-I$2)*I10	=J$2*$B10+(1-J$2)*J10

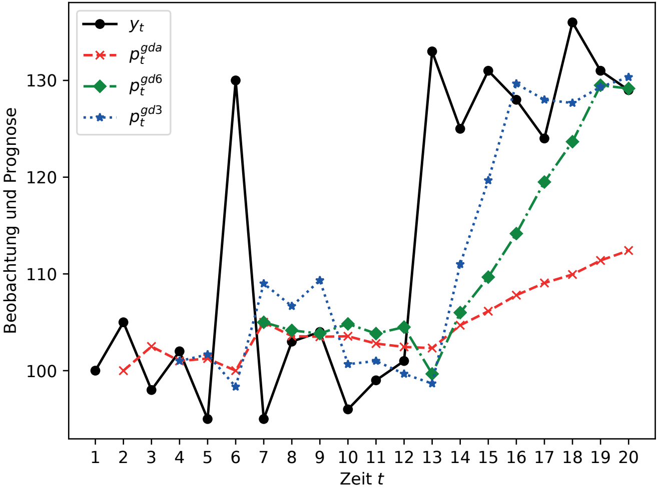
Berechnung in Tabellenkalkulation IV

	A	B	C	D	E	F	G	H	I	J
1				**(Gleitende) Durchschnitte**				**Exp. Glättung 1. Ordnung**		
2			**Perioden**	alle	6	3	**Alpha**	0,1	0,3	0,5
3										
4	*t*	*y(t)*		$p_{gda}(t)$	$p_{gd6}(t)$	$p_{gd3}(t)$		$p_{eg0.1}(t)$	$p_{eg0.3}(t)$	$p_{eg0.5}(t)$
5	1	100		NaN	NaN	NaN		NaN	NaN	NaN
6	2	105		100,00	NaN	NaN		100,00	100,00	100,00
7	3	98		102,50	NaN	NaN		100,50	101,50	102,50
8	4	102		101,00	NaN	101,00		100,25	100,45	100,25
9	5	95		101,25	NaN	101,67		100,43	100,92	101,13
10	6	130		100,00	NaN	98,33		99,88	99,14	98,06
11	7	95		105,00	105,00	109,00		102,89	108,40	114,03
12	8	103		103,57	104,17	106,67		102,10	104,38	104,52
13	9	104		103,50	103,83	109,33		102,19	103,97	103,76
14	10	96		103,56	104,83	100,67		102,37	103,98	103,88
15	11	99		102,80	103,83	101,00		101,74	101,58	99,94
16	12	101		102,45	104,50	99,67		101,46	100,81	99,47
17	13	133		102,33	99,67	98,67		101,42	100,87	100,23
18	14	125		104,69	106,00	111,00		104,58	110,51	116,62
19	15	131		106,14	109,67	119,67		106,62	114,85	120,81
20	16	128		107,80	114,17	129,67		109,06	119,70	125,90
21	17	124		109,06	119,50	128,00		110,95	122,19	126,95
22	18	136		109,94	123,67	127,67		112,26	122,73	125,48
23	19	131		111,39	129,50	129,33		114,63	126,71	130,74
24	20	129		112,42	129,17	130,33		116,27	128,00	130,87

Berechnung in Tabellenkalkulation V



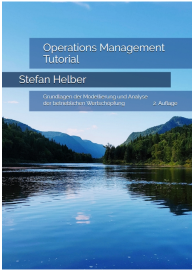
Berechnung in Tabellenkalkulation VI



Prognoserechnungen

Konstruktion komplexerer Prognosemodelle mittels Regressionsrechnung

Prof. Dr. Stefan Helber



Prognosefunktion, Prognosefehler, Verlustfunktion

Datenpunkte $i = 1, ..., I$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_I \end{pmatrix} \quad \text{und} \quad X = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_I \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,j} & \dots & x_{1,J} \\ x_{2,1} & x_{2,2} & \dots & x_{2,j} & \dots & x_{2,J} \\ \vdots & & \ddots & & & \vdots \\ x_{i,1} & x_{i,2} & \dots & x_{i,j} & \dots & x_{i,J} \\ \vdots & & \ddots & & & \vdots \\ x_{I,1} & x_{I,2} & \dots & x_{I,j} & \dots & x_{I,J} \end{pmatrix} \tag{1}$$

Prognosemodell

$$\hat{y}_i = f(\mathbf{x}_i) = f(x_{i,1}, x_{i,2}, ..., x_{i,j}, ..., x_{i,J}) \tag{2}$$

liefert Schätzwert $\hat{y}_i$

Beispiel

Person	Gewicht [kg]	Alter [Jahre]	Größe [cm]	Geschlecht
1	80	45	185	m
2	55	42	170	w
3	50	17	172	w
4	65	16	180	m

Dummy-Variable für Geschlecht

$$Y = \begin{pmatrix} 80 \\ 55 \\ 50 \\ 65 \end{pmatrix} \quad \text{und} \quad X = \begin{pmatrix} 45 & 185 & 1 & 0 \\ 42 & 170 & 0 & 1 \\ 17 & 172 & 0 & 1 \\ 16 & 180 & 1 & 0 \end{pmatrix} \tag{3}$$

Beispiel: Zwei Prognosemodelle

$$Y = \begin{pmatrix} 80 \\ 55 \\ 50 \\ 65 \end{pmatrix} \quad \text{und} \quad X = \begin{pmatrix} 45 & 185 & 1 & 0 \\ 42 & 170 & 0 & 1 \\ 17 & 172 & 0 & 1 \\ 16 & 180 & 1 & 0 \end{pmatrix} \tag{4}$$

$$f_A(\mathbf{x}_i) = \hat{y}_i^A = c_1 \cdot x_{i,1} + c_2 \cdot x_{i,2} + c_3 \cdot x_{i,3} + c_4 \cdot x_{i,4} \tag{5}$$

$$f_B(\mathbf{x}_i) = \hat{y}_i^B = c_0 + c_1 \cdot x_{i,1} + c_2 \cdot x_{i,2} + c_3 \cdot x_{i,3} \tag{6}$$

Modelle äquivalent

Prognosefehler $e_i = y_i - \hat{y}_i$ und Verlustfunktion

- mittlerer quadratischer Fehler, (*Mean Squared Error, MSE*)

$$L^{\text{MSE}} = \frac{1}{I}\sum_{i=1}^{I}(y_i - \hat{y}_i)^2 \tag{7}$$

Quadratwurzel daraus, (*Root Mean Squared Error, RMSE*)

$$L^{\text{RMSE}} = \sqrt{L^{\text{MSE}}} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(y_i - \hat{y}_i)^2}. \tag{8}$$

- mittlerer absolute Fehler, (*Mean Absolute Error, MAE*)

$$L^{\text{MAE}} = \frac{1}{I}\sum_{i=1}^{I}|y_i - \hat{y}_i|. \tag{9}$$

- mittlerer absoluter prozentuale Fehler, (*Mean Absolute Percentage Error, MAPE*)

$$L^{\text{MAPE}} = \frac{1}{I}\sum_{i=1}^{I}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{10}$$

Konstruktion eines elementaren Prognosemodells I

Schätzfunktion

$$\hat{y}_i = f(\mathbf{x}_i) = f(x_{i,1}, x_{i,2}, ..., x_{i,j}, ..., x_{i,J}) = c, \tag{11}$$

Verlustfunktion

$$L^{\text{MSE}} = \frac{1}{I} \sum_{i=1}^{I} (y_i - \hat{y}_i)^2 = \frac{1}{I} \sum_{i=1}^{I} (y_i - c)^2 \tag{12}$$

Ableitung der Verlustfunktion L^{MSE} nach c

$$\frac{\mathrm{d}L^{\text{MSE}}}{\mathrm{d}c} = (-2)\frac{1}{I} \sum_{i=1}^{I} (y_i - c) \stackrel{!}{=} 0 \tag{13}$$

Konstruktion eines elementaren Prognosemodells II

Nullsetzen und Auflösen

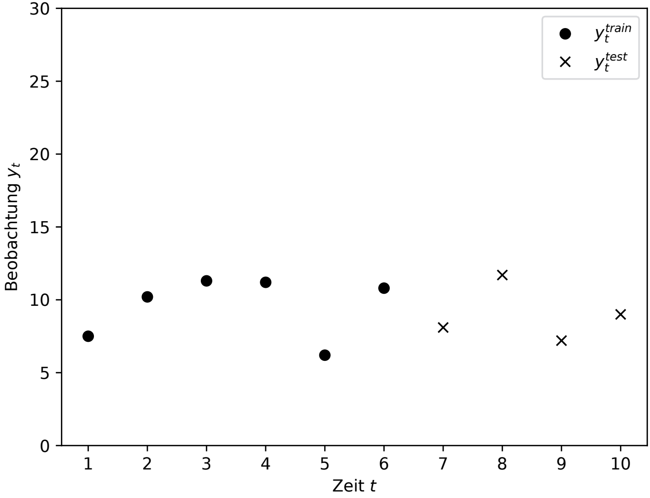
$$c = \frac{1}{I} \sum_{i=1}^{I} y_i \tag{14}$$

Prognosefunktion jetzt explizit

$$f(\mathbf{x}_i) = c = \frac{1}{I} \sum_{i=1}^{I} y_i \tag{15}$$

Ermittlung der Koeffizienten mit Software wie `R` oder `Python`

Trainings- vs. Validierungsdaten, Overfitting I



Trainings- vs. Validierungsdaten, Overfitting II

Zeit bzw. Polynome der Zeit als erklärende Variable

$$Y = \begin{pmatrix} 7{,}5 \\ 10{,}2 \\ 11{,}3 \\ 11{,}2 \\ 6{,}2 \\ 10{,}8 \\ 8{,}1 \\ 11{,}7 \\ 7{,}2 \\ 9{,}0 \end{pmatrix} \quad \text{und} \quad X = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{pmatrix} \tag{16}$$

Trainings- vs. Validierungsdaten, Overfitting III

lineare Regressionsrechnung für lineares Modell

$$\hat{y}_i^{(1)} = f_1(\mathbf{x}_i) = c_0 + c_1 \cdot x_i \tag{17}$$

mit Trainingsdatensatz

$$Y^{\text{train}} = \begin{pmatrix} 7{,}5 \\ 10{,}2 \\ 11{,}3 \\ 11{,}2 \\ 6{,}2 \\ 10{,}8 \end{pmatrix} \quad \text{und} \quad X^{\text{train}} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix}. \tag{18}$$

Trainings- vs. Validierungsdaten, Overfitting IV

lineare Regression für quadratisches Modell

$$\hat{y}_i^{(2)} = f_2(\mathbf{x}_i) = c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2 \tag{19}$$

mit Trainingsdaten

$$Y^{\text{train}} = \begin{pmatrix} 7{,}5 \\ 10{,}2 \\ 11{,}3 \\ 11{,}2 \\ 6{,}2 \\ 10{,}8 \end{pmatrix} \quad \text{und} \quad X^{\text{train}} = \begin{pmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \\ 5 & 25 \\ 6 & 36 \end{pmatrix} \tag{20}$$

Trainings- vs. Validierungsdaten, Overfitting V

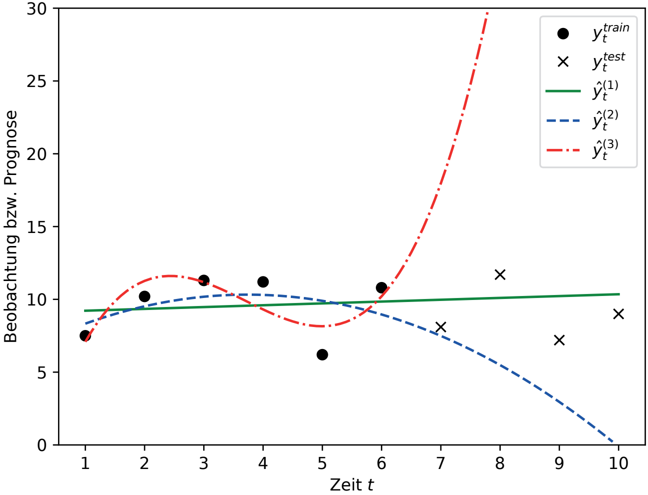
lineare Regression für kubisches Modell

$$\hat{y}_i^{(3)} = f_3(\mathbf{x}_i) = c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2 + c_3 \cdot x_i^3 \qquad (21)$$

mit Trainingsdaten

$$Y^{\text{train}} = \begin{pmatrix} 7{,}5 \\ 10{,}2 \\ 11{,}3 \\ 11{,}2 \\ 6{,}2 \\ 10{,}8 \end{pmatrix} \quad \text{und} \quad X^{\text{train}} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \\ 3 & 9 & 27 \\ 4 & 16 & 64 \\ 5 & 25 & 125 \\ 6 & 36 & 216 \end{pmatrix} \qquad (22)$$

Trainings- vs. Validierungsdaten, Overfitting VI



Trainings- vs. Validierungsdaten, Overfitting VII

Ergebnisse der drei Regressionsrechnungen

Nr.	Schätzfunktion	Koeffizienten	RMSE-Train	RMSE-Test
1	$\hat{y}_i^{(1)} = c_0 + c_1 \cdot x_i$	$c_0 = 9{,}09333333$ $c_1 = 0{,}12571429$	1,954	2,064
2	$\hat{y}_i^{(2)} = c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2$	$c_0 = 6{,}61$ $c_1 = 1{,}98821429$ $c_2 = -0{,}26607143$	1,838	5,918
3	$\hat{y}_i^{(3)} = c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2 + c_3 \cdot x_i^3$	$c_0 = -3{,}866$ $c_1 = 15{,}167$ $c_2 = -4{,}631$ $c_3 = 0{,}416$	1,230	54,305

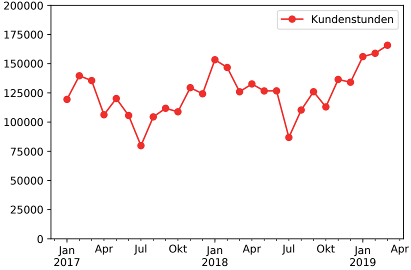
Beispiel: Besucherprognose im Fitnessstudio

Einflussgrößen der Besucherzahl je Stunde („Kundenstunden“)

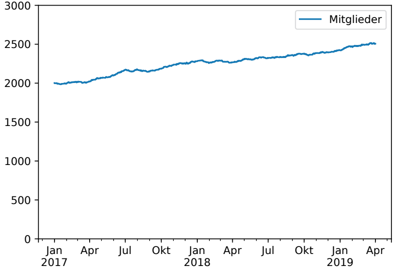
- Wochentag
- Stunde des Tages
- Monat
- Feiertag
- Schulferien
- Wetter

Prognose Basis für Personaleinsatzplan, Datenerfassung über Armbänder

Kundenstunden und Mitgliederbestand

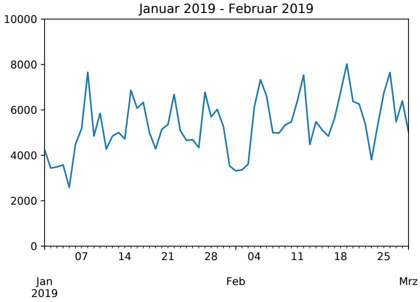


(a) Monatliche Kundenstunden

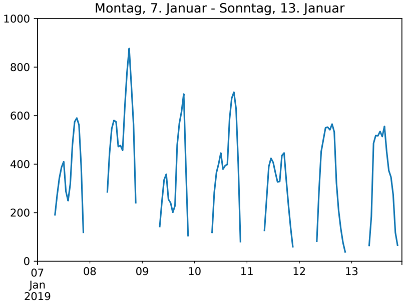


(b) Mitgliederbestand

Kundenstunden auf Tages- und Stundenbasis



(a) Kundenstunden auf Tagesbasis



(b) Kundenstunden auf Stundenbasis

Naives Modell

Regression auf Kundenanzahl und Uhrzeit

$$\hat{y}_i = c_{\text{Anz}} \cdot x_{i,\text{Anz}} + c_{8\text{h}} \cdot x_{i,8\text{h}} + c_{9\text{h}} \cdot x_{i,9\text{h}} + \ldots + c_{21\text{h}} \cdot x_{i,21\text{h}} \qquad (23)$$

Problem: Schwankungsbreite hängt dann nicht von der Kundenanzahl ab

Alternativer Ansatz: Multiplikative Modelle

Multiplikatives Modell A I

$$\hat{y}_i^A = x_{i,\text{Anz}}^{c_\text{Anz}} \cdot c_{8\text{h}}^{x_{i,8\text{h}}} \cdot c_{9\text{h}}^{x_{i,9\text{h}}} \cdot \ldots \cdot c_{21\text{h}}^{x_{i,21\text{h}}} \tag{24}$$

Logarithmieren zur Linearisierung

$$\begin{aligned}
\ln\left(\hat{y}_i^A\right) &= \ln\left(x_{i,\text{Anz}}^{c_\text{Anz}} \cdot c_{8\text{h}}^{x_{i,8\text{h}}} \cdot c_{9\text{h}}^{x_{i,9\text{h}}} \cdot \ldots \cdot c_{21\text{h}}^{x_{i,21\text{h}}}\right) \\
&= \ln\left(x_{i,\text{Anz}}^{c_\text{Anz}}\right) + \ln\left(c_{8\text{h}}^{x_{i,8\text{h}}}\right) + \ln\left(c_{9\text{h}}^{x_{i,9\text{h}}}\right) + \ldots + \ln\left(c_{21\text{h}}^{x_{i,21\text{h}}}\right) \\
&= c_\text{Anz} \cdot \ln\left(x_{i,\text{Anz}}\right) + \ln(c_{8\text{h}}) \cdot x_{i,8\text{h}} + \ln(c_{9\text{h}}) \cdot x_{i,9\text{h}} + \ldots + \ln(c_{21\text{h}}) \cdot x_{\text{i},21\text{h}}
\end{aligned} \tag{25}$$

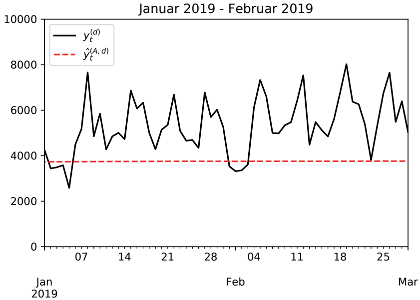
Kurzschreibweise

$$\begin{aligned}
\tilde{c}_{8\text{h}} &= \ln(c_{8\text{h}}) \\
\tilde{c}_{9\text{h}} &= \ln(c_{9\text{h}}) \\
&\vdots \\
\tilde{c}_{21\text{h}} &= \ln(c_{21\text{h}})
\end{aligned}$$

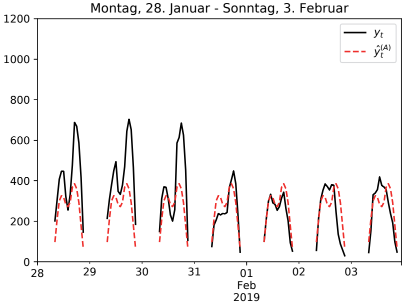
lineare Schätzfunktion

$$\ln\left(\hat{y}_i^A\right) = c_\text{Anz} \cdot \ln\left(x_{i,\text{Anz}}\right) + \tilde{c}_{8\text{h}} \cdot x_{\text{i},8\text{h}} + \tilde{c}_{8\text{h}} \cdot x_{i,9\text{h}} + \ldots + \tilde{c}_{21\text{h}} \cdot x_{i,21\text{h}} \tag{26}$$

Multiplikatives Modell A II



(a) Kundenstunden auf Tagesbasis

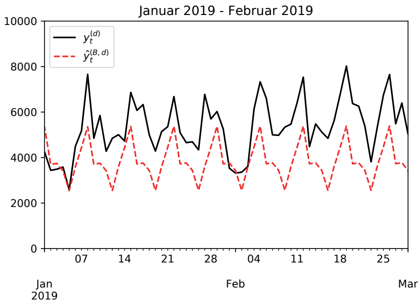


(b) Kundenstunden auf Stundenbasis

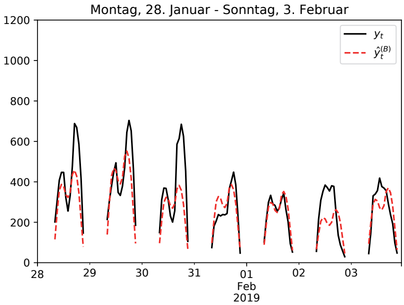
Problem: Alle Wochentage werden gleich behandelt

Multiplikatives Modell B

$$\ln\left(\hat{y}_i^B\right) = c_{\text{Anz}} \cdot \ln(x_{i,\text{Anz}}) + \tilde{c}_{8\text{h}} \cdot x_{\text{i},8\text{h}} + \tilde{c}_{8\text{h}} \cdot x_{i,9\text{h}} + \ldots + \tilde{c}_{21\text{h}} \cdot x_{i,21\text{h}} + \tilde{c}_{\text{Mo}} \cdot x_{i,\text{Mo}} + \tilde{c}_{\text{Di}} \cdot x_{i,\text{Di}} + \ldots + \tilde{c}_{\text{Sa}} \cdot x_{i,\text{Sa}} + \tilde{c}_{\text{So}} \cdot x_{i,\text{So}} \tag{27}$$



(a) Kundenstunden auf Tagesbasis

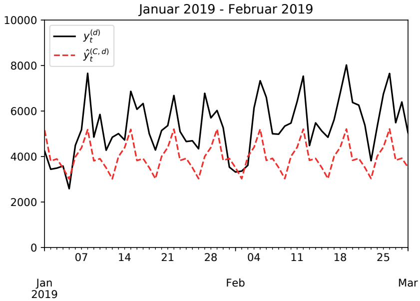


(b) Kundenstunden auf Stundenbasis

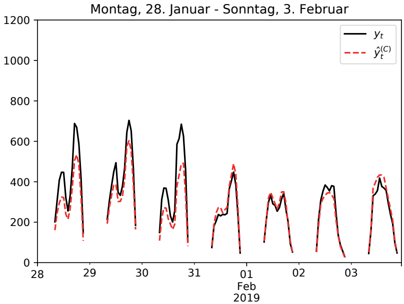
Problem: Unterschiedliche Muster innerhalb der Wochentage nicht erfasst

Multiplikatives Modell C

$$\ln\left(\hat{y}_i^C\right) = c_{\text{Anz}} \cdot \ln(x_{i,\text{Anz}}) + \tilde{c}_{\text{Mo8h}} \cdot x_{\text{i,Mo8h}} + \tilde{c}_{\text{Mo9h}} \cdot x_{i,\text{Mo9h}} + \ldots + \tilde{c}_{\text{So21h}} \cdot x_{i,\text{So21h}} \tag{28}$$



(a) Kundenstunden auf Tagesbasis

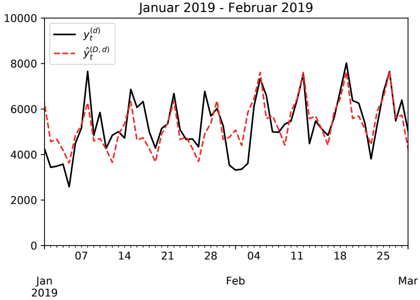


(b) Kundenstunden auf Stundenbasis

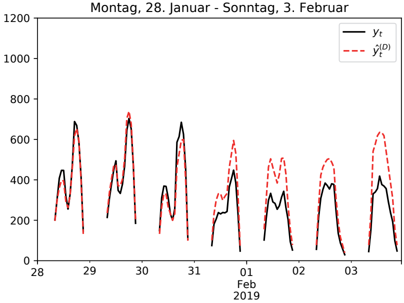
Problem: Unterschiede der Monate nicht erfasst

Multiplikatives Modell D

$$\ln\left(\hat{y}_i^D\right) = c_{\text{Anz}} \cdot \ln(x_{i,\text{Anz}}) + \tilde{c}_{\text{Mo8h}} \cdot x_{\text{i,Mo8h}} + \tilde{c}_{\text{Mo9h}} \cdot x_{i,\text{Mo9h}} + \ldots + \tilde{c}_{\text{So21h}} \cdot x_{i,\text{So21h}} + \tilde{c}_{\text{Jan}} \cdot x_{\text{i,Jan}} + \tilde{c}_{\text{Feb}} \cdot x_{i,\text{Feb}} + \ldots + \tilde{c}_{\text{Dez}} \cdot x_{i,\text{Dez}} \tag{29}$$



(a) Kundenstunden auf Tagesbasis

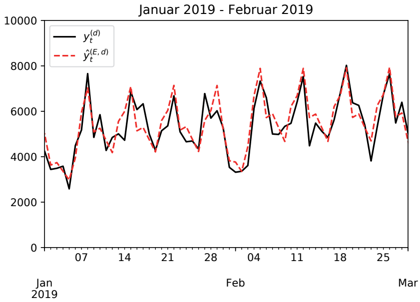


(b) Kundenstunden auf Stundenbasis

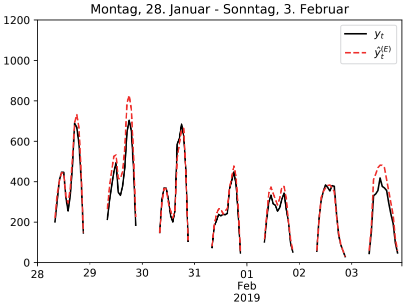
Problem: Schulferien und Feiertage nicht erfasst

Multiplikatives Modell E

$$\ln\left(\hat{y}_i^E\right) = c_{\text{Anz}} \cdot \ln(x_{i,\text{Anz}}) + \tilde{c}_{\text{Mo8h}} \cdot x_{\text{i,Mo8h}} + \tilde{c}_{\text{Mo9h}} \cdot x_{i,\text{Mo9h}} + \ldots + \tilde{c}_{\text{So21h}} \cdot x_{i,\text{So21h}} + \tilde{c}_{\text{Jan}} \cdot x_{\text{i,Jan}} + \tilde{c}_{\text{Feb}} \cdot x_{i,\text{Feb}} + \ldots + \tilde{c}_{\text{Dez}} \cdot x_{i,\text{Dez}} + \tilde{c}_{\text{SFFT}} \cdot x_{i,\text{SFFT}} \tag{30}$$



(a) Kundenstunden auf Tagesbasis



(b) Kundenstunden auf Stundenbasis

Vergleich der Modelle

Modell	erklärende Variablen	RMSE-Train	RMSE-Test
A	$x_{i,\text{Anz}},\ \ x_{i,8h},\ldots,x_{i,21h}$	118,96	181,56
B	$x_{i,\text{Anz}},\ \ x_{i,8h},\ldots,x_{i,21h},\ \ x_{i,Mo},\ldots,x_{i,So}$	110,55	168,06
C	$x_{i,\text{Anz}},\ \ x_{i,Mo8h},\ldots,x_{i,So21h}$	73,14	124,98
D	$x_{i,\text{Anz}},\ \ x_{i,Mo8h},\ldots,x_{i,So21h},\ \ x_{i,\text{Jan}},\ldots,x_{i,\text{Dez}}$	55,45	65,57
E	$x_{i,\text{Anz}},\ \ x_{i,Mo8h},\ldots,x_{i,So21h},\ \ x_{i,\text{Jan}},\ldots,x_{i,\text{Dez}},\ \ x_{i,\text{SFFT}}$	41,80	47,77

Erkenntnisse

- Modellbildung sukzessiver Prozess
- Test- oder Validierungsdaten wichtig
- Software wie `R` oder `Python` erforderlich
- Statistik-Kenntnisse und Programmierkenntnisse erforderlich