# Ausblick: Datenfit

## Arbeiten mit Listen und Matrizen

Thomas Heim

IMN, Hochschule für Technik, FHNW

Windisch, 29. Oktober 2016

# Allgemeine lineare Regression

- kombiniert lineare Algebra (Matrizenrechnung) und mehrdimensionale Analysis (partielle Ableitungen)
- Umsetzung mit Ti-nspire

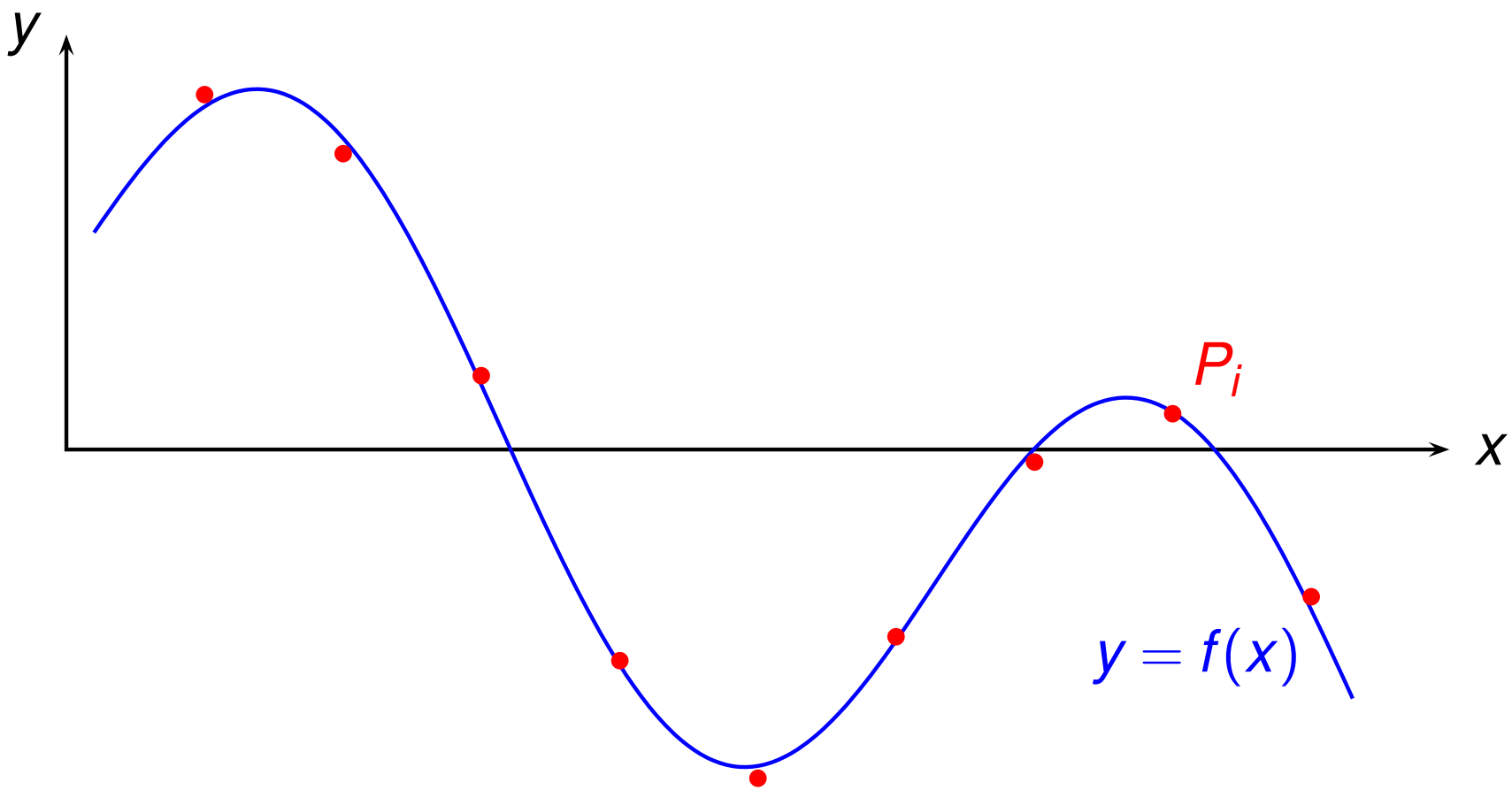
# Allgemeine lineare Regression

- kombiniert lineare Algebra (Matrizenrechnung) und mehrdimensionale Analysis (partielle Ableitungen)
- Umsetzung mit Ti-nspire

mit Tool wie Matlab gehts allerdings besser

# Problemstellung

Durch vorgegebene Punkte $P_1, \ldots, P_n$ soll eine optimale Kurve $y = f(x)$ gelegt werden.



# Problemstellung

- Fitfunktion $f(x)$ hängt ausser von $x$ auch von Fitparametern $c_1, \ldots, c_m$ ab

# Problemstellung

- Fitfunktion $f(x)$ hängt ausser von $x$ auch von Fitparametern $c_1, \ldots, c_m$ ab
- „optimale“ Fitfunktion liefert minimale quadratische Abweichung

$$S(c_1, \ldots, c_m) = \sum_{i=1}^{n} (f(x_i) - y_i)^2 \quad \rightarrow \quad \min$$

(least squares fitting)

# Problemstellung

- Fitfunktion $f(x)$ hängt ausser von $x$ auch von Fitparametern $c_1, \ldots, c_m$ ab
- „optimale“ Fitfunktion liefert minimale quadratische Abweichung

$$S(c_1, \ldots, c_m) = \sum_{i=1}^{n} (f(x_i) - y_i)^2 \quad \rightarrow \quad \min$$

(least squares fitting)

- Bedingungen für Fitparameter:

partielle Ableitungen $\frac{\partial S}{\partial c_k} = 0$ für $k = 1, \ldots, m$

# Lineare Regression

- Falls die Fitparameter nur linear in der Fitfunktion vorkommen,

$$f(x) = c_1 \cdot f_1(x) + c_2 \cdot f_2(x) + \ldots + c_m \cdot f_m(x),$$

  ergibt sich ein lineares Gleichungssystem für die $c_k$
- die Basisfunktionen $f_1(x)$, $f_2(x), \ldots, f_m(x)$ dürfen dabei beliebig kompliziert von $x$ abhängen

# Lineares Gleichungssystem

Gleichungssystem für die Fitparameter:

$$\begin{pmatrix} \sum_{i=1}^{n} f_1(x_i) f_1(x_i) & \cdots & \sum_{i=1}^{n} f_1(x_i) f_m(x_i) \\ \sum_{i=1}^{n} f_2(x_i) f_1(x_i) & \cdots & \sum_{i=1}^{n} f_2(x_i) f_m(x_i) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} f_m(x_i) f_1(x_i) & \cdots & \sum_{i=1}^{n} f_m(x_i) f_m(x_i) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \vec{b}$$

$$\text{mit} \qquad \vec{b} = \begin{pmatrix} \sum_{i=1}^{n} f_1(x_i) y_i \\ \sum_{i=1}^{n} f_2(x_i) y_i \\ \vdots \\ \sum_{i=1}^{n} f_m(x_i) y_i \end{pmatrix}$$

# Matrizenform

Summation über Datenpunkte kann als Matrixmultiplikation dargestellt werden!

$$\text{Mit } F = \begin{pmatrix} f_1(x_1) & f_1(x_2) & \cdots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \cdots & f_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(x_1) & f_m(x_2) & \cdots & f_m(x_n) \end{pmatrix}$$

wird das Gl.system für die Fitparameter $\vec{c}$ zu

$$F \cdot F^{\mathsf{T}} \cdot \vec{c} = F \cdot \vec{y}$$

mit den vorgegebenen $y$-Werten als Spaltenvektor

# Implementierung mit Ti-nspire

Durch die 9 Punkte

| $x_i$ | $y_i$ |
|---|---|
| 1 | 5.13 |
| 2 | 4.28 |
| 3 | 1.07 |

| $x_i$ | $y_i$ |
|---|---|
| 4 | $-3.05$ |
| 5 | $-4.75$ |
| 6 | $-2.71$ |

| $x_i$ | $y_i$ |
|---|---|
| 7 | $-0.18$ |
| 8 | 0.52 |
| 9 | $-2.13$ |

soll eine Funktion der Form

$$y = f(x) = c_1 \cdot \sin(x) + c_2 \cdot x + c_3 \cdot 1$$

gelegt werden (Sinus plus Gerade)

# Arbeiten mit Listen

Basisfunktionen sind

$$f_1(x) = \sin(x), \qquad f_2(x) = x, \qquad f_3(x) = 1$$

# Arbeiten mit Listen

Basisfunktionen sind

$$f_1(x) = \sin(x), \qquad f_2(x) = x, \qquad f_3(x) = 1$$

- bilden $x$ als Liste (geschweifte Klammern) und $y$ als Spaltenvektor

# Arbeiten mit Listen

Basisfunktionen sind

$$f_1(x) = \sin(x), \qquad f_2(x) = x, \qquad f_3(x) = 1$$

- bilden $x$ als Liste (geschweifte Klammern) und $y$ als Spaltenvektor
- wenden Basisfunktionen auf Liste $x$ an, gibt drei Listen $f_1$, $f_2$, $f_3$ von Funktionswerten

# Arbeiten mit Listen

Basisfunktionen sind

$$f_1(x) = \sin(x), \qquad f_2(x) = x, \qquad f_3(x) = 1$$

- bilden $x$ als Liste (geschweifte Klammern) und $y$ als Spaltenvektor
- wenden Basisfunktionen auf Liste $x$ an, gibt drei Listen $f_1$, $f_2$, $f_3$ von Funktionswerten
- hängen Listen zusammen mit augment

# Arbeiten mit Listen

Basisfunktionen sind

$$f_1(x) = \sin(x), \qquad f_2(x) = x, \qquad f_3(x) = 1$$

- bilden $x$ als Liste (geschweifte Klammern) und $y$ als Spaltenvektor
- wenden Basisfunktionen auf Liste $x$ an, gibt drei Listen $f_1$, $f_2$, $f_3$ von Funktionswerten
- hängen Listen zusammen mit augment
- verwandeln lange Liste in Matrix $F$ mit list▷mat

# Fitparameter bestimmen

- bilden $M = F \cdot F^{\mathsf{T}}$ und $\vec{b} = F \cdot \vec{y}$

# Fitparameter bestimmen

- bilden $M = F \cdot F^{\mathsf{T}}$ und $\vec{b} = F \cdot \vec{y}$
- dann sind die Fitparameter

$$\vec{c} = M^{-1} \cdot \vec{b}$$

# Fitparameter bestimmen

- bilden $M = F \cdot F^{\mathsf{T}}$ und $\vec{b} = F \cdot \vec{y}$
- dann sind die Fitparameter

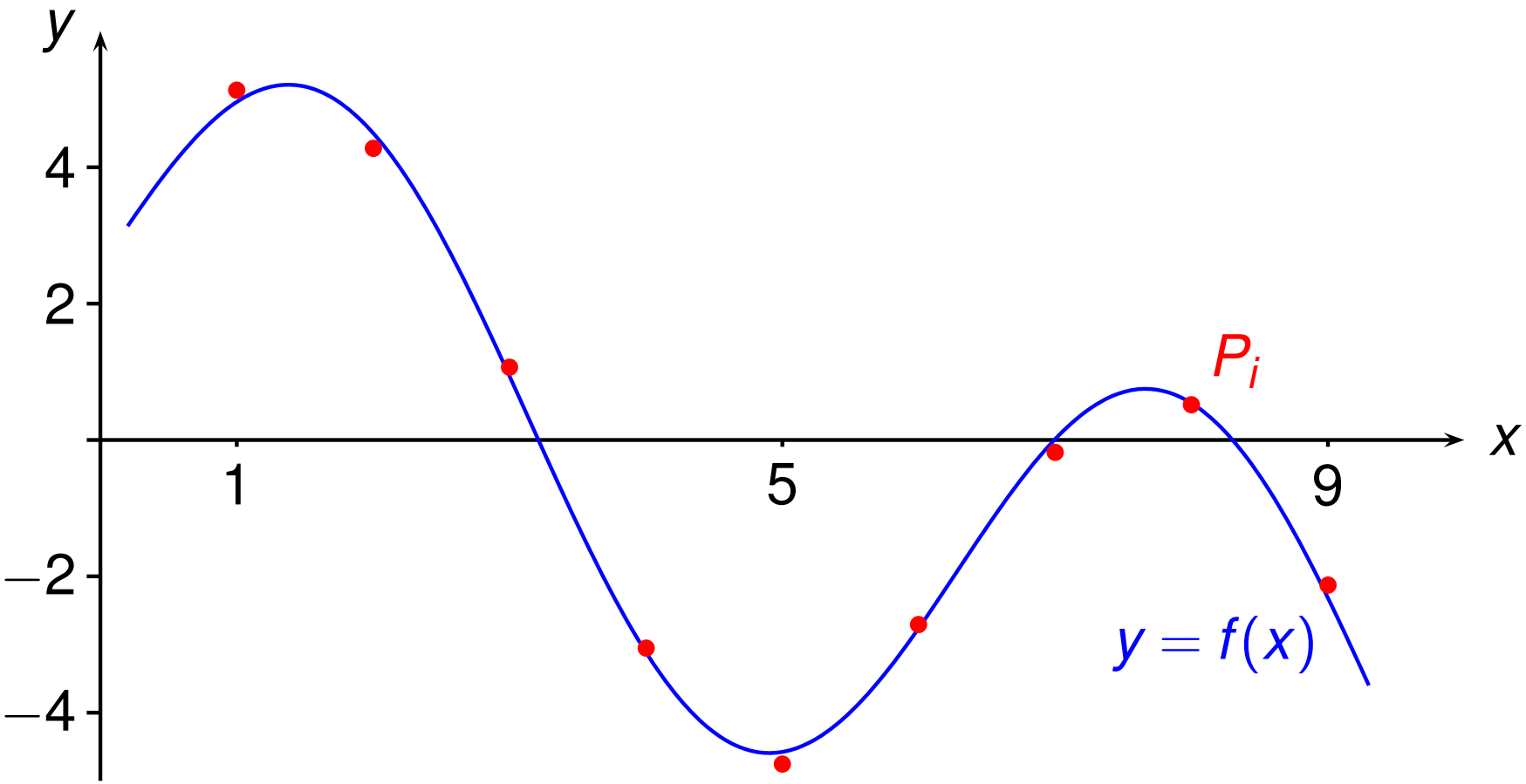
$$\vec{c} = M^{-1} \cdot \vec{b}$$

- in unserem Beispiel:

$$c_1 = 3.71815, \qquad c_2 = -0.70998, \qquad c_3 = 2.53991$$

$$f(x) = 3.71815 \sin(x) - 0.70998x + 2.53991$$

# Grafik



# Fazit

Ich verwende solche Beispiele im Unterricht, um

- verschiedene Themen zu kombinieren (mehrdim. Diff.rechnung und lineare Algebra)
- klarzumachen, dass
  - lineare Regression nicht nur Geraden liefern muss
  - wenn immer möglich lineare Ausgleichungsrechnung vorzuziehen ist gegenüber nicht-linearem Fitten

# Fazit

Ich verwende solche Beispiele im Unterricht, um

- verschiedene Themen zu kombinieren (mehrdim. Diff.rechnung und lineare Algebra)
- klarzumachen, dass
  - lineare Regression nicht nur Geraden liefern muss
  - wenn immer möglich lineare Ausgleichungsrechnung vorzuziehen ist gegenüber nicht-linearem Fitten
  - ein paar zusätzliche Abstraktionsstufen, einmalig durchgeführt, zu erheblich geringerem Arbeitsaufwand in der Anwendung führen: „solve“ ist selten die bestmögliche Antwort. . .