Einführung in Operations Research

Lineare Ausgleichsprobleme

Prof. Dr. Thomas Slawig

CAU Kiel
Institut für Informatik

Inhalt

Lernziele

- Die Form eines Linearen Ausgleichsproblems kennen.
- Entscheiden können, ob ein beschriebenes Problem ein lineares oder ein nichtlineares Ausgleichsproblem ist.
- Daten und Modellparameter definieren können.
- Notwendige Bedingung für Minimum kennen.
- Beide Lösungsmethoden kennen und anwenden können.

Inhalt

Lineares Ausgleichsproblem

- Problemdefinition

$$\min_{x \in \mathbb{R}^n} \|Ax - z\|^2$$

mit $A \in \mathbb{R}^{m \times n}, z \in \mathbb{R}^m$, aber jetzt mit $m > n$ (überbestimmtes Problem).

- Dabei ist

$$\|x\| = \sqrt{\sum_{k=1}^{n} x_k^2},$$

also

$$\|Ax - z\|^2 = \sum_{k=1}^{m} ((Ax - z)_k)^2 = \sum_{k=1}^{m} \left(\sum_{j=1}^{n} A_{kj} x_j - z_k \right)^2.$$

Inhalt

Motivation

- Gegeben: Datenpunkte

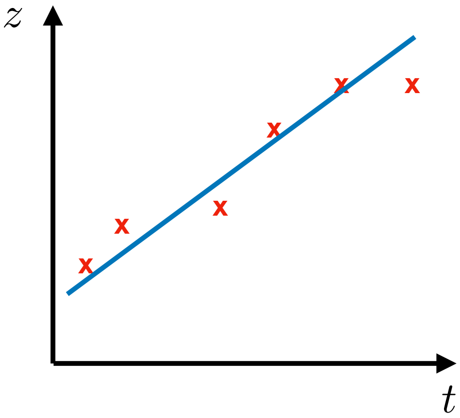
$$(t_k, z_k)_{k=1,\ldots,m},\ t_k, z_k \in \mathbb{R}.$$

- Gesucht: Ausgleichsgerade

$$y(t) = at + b$$

mit der Eigenschaft

$$y(t_k) = at_k + b = z_k, \quad k = 1, \ldots, m.$$



Motivation

- Gesucht: Ausgleichsgerade mit der Eigenschaft

$$y(t_k) = at_k + b = z_k, \quad k = 1, \ldots, m.$$

- Das ist ein lineares Gleichungssystem für die Unbekannten $x = (a, b) \in \mathbb{R}^2$:

$$\underbrace{\begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}}_{=A \in \mathbb{R}^{m \times 2}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{=x \in \mathbb{R}^2} = \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}}_{=z \in \mathbb{R}^m} \quad \Longleftrightarrow \quad Ax = z.$$

Überbestimmtes System

- Das lineare Gleichungssystem

$$Ax = z, \quad A \in \mathbb{R}^{m \times n},$$

ist für $m > n$ überbestimmt, wenn die t_i paarweise verschieden sind.

- Dann gibt es keine Lösung.
- Was können wir tun?
- Das Minimum bestimmen:

$$\min_{x \in \mathbb{R}^n} \|Ax - z\|^2$$

Inhalt

Unterschied: lineares ↔ nichtlineares Ausgleichsproblem

- Gegeben wie oben: Datenpunkte

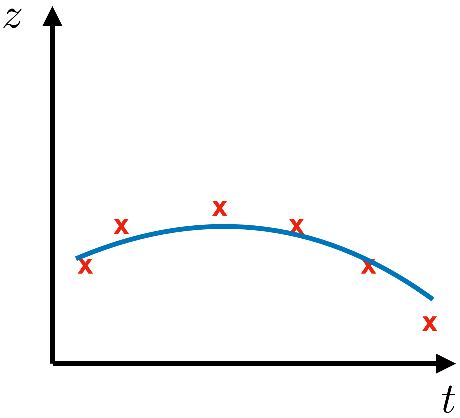
$$(t_k, z_k)_{k=1,\dots,m},\ t_k, z_k \in \mathbb{R}.$$

- Gesucht: quadratische Funktion

$$y(t) = at^2 + bt + c$$

mit

$$y(t_k) = at_k^2 + bt_k + c = z_k,$$
$$k = 1, \dots, m.$$



Unterschied: lineares ↔ nichtlineares Ausgleichsproblem

- Gesucht: quadratische Funktion mit

$$y(t_k) = at_k^2 + bt_k + c = z_k, \quad k = 1, \dots, m.$$

- Das ergibt wieder ein lineares Gleichungssystem für $x = (a, b, c) \in \mathbb{R}^3$:

$$\underbrace{\begin{bmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_m^2 & t_m & 1 \end{bmatrix}}_{=A \in \mathbb{R}^{m \times 3}} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{=x \in \mathbb{R}^3} = \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}}_{=z \in \mathbb{R}^m} \quad \Longleftrightarrow \quad Ax = z.$$

- Wieder *lineares* Ausgleichsproblem: Funktion y ist linear in den gesuchten Parametern $x = (a, b, c)$. Sie ist nichtlinear (quadratisch) bezüglich t (Daten).

Unterschied: lineares ↔ nichtlineares Ausgleichsproblem

- Analog: Gesucht:

$$\text{Polynom:} \quad y(t) = \sum_{k=0}^{m} a_k t^k, \quad x = (a_k)_{k=0}^m$$

$$\text{Trigonometrisches Polynom:} \quad y(t) = \sum_{k=1}^{m} a_k \sin(kt), \quad x = (a_k)_{k=1}^m .$$

- Gegeben beliebige Funktion (Modell) y, die linear bezüglich der Parameter x ist, d. h.

$$\left. \begin{aligned} y(\alpha x, t) &= \alpha y(x, t), \\ y(x + \hat{x}, t) &= y(x, t) + y(\hat{x}, t) \end{aligned} \right\} \quad \forall t, \alpha \in \mathbb{R}, x, \hat{x} \in \mathbb{R}^m$$

- Beispiel: y = Gewinn zur Zeit t, abhängig von Umsatz x, Annahme: lineare Abhängigkeit zwischen Umsatz x und Gewinn y (realistisch?)
- Wieder *lineare* Ausgleichsprobleme.

Unterschied: lineares $\leftrightarrow$ nichtlineares Ausgleichsproblem

- Gesucht: Funktion

$$y(t) = ae^{bt}$$

mit

$$y(t_k) = ae^{bt_k} = z_k, \quad k = 1, \ldots, m.$$

- Das lässt sich *nicht* als lineares Gleichungssystem für $x = (a, b) \in \mathbb{R}^2$ schreiben.
- Funktion y ist zwar linear bezüglich a, aber nichtlinear bezüglich b,
- $\rightsquigarrow$ kein *lineares, sondern nichtlineares* Ausgleichsproblem.
- Analog: Funktion (Modell) $y = y(x, t)$, die *nichtlinear* bezüglich der Parameter x ist.
- $\rightsquigarrow$ kein *lineares, sondern nichtlineares* Ausgleichsproblem.

Inhalt

Bedingungen für Minimum

- Vgl. $n = 1$: x^* Minimum. Es gilt

$$f(x^*) = \min_{x \in \mathbb{R}} f(x) \quad \Longrightarrow \quad f'(x^*) = \frac{df}{dx}(x^*) = 0.$$

- Verallgemeinerung $n > 1$: x^* Minimum. Es gilt

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \quad \Longrightarrow \quad \nabla f(x^*) = \left(\frac{\partial f}{\partial x_k}(x^*)\right)_{k=1}^{n} = 0.$$

Bedingungen für Minimum

- Es gilt für

$$f(x) := \frac{1}{2}\|Ax - z\|^2 = \frac{1}{2}\sum_{i=1}^{m} (Ax - z)_i^{\,2} = \frac{1}{2}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j - z_i\right)^2$$

- Ableitung:

$$\begin{aligned}
\frac{\partial f}{\partial x_k}(x) &= \frac{1}{2}\sum_{i=1}^{m}\frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n} A_{ij}x_j - z_i\right)^2 = \frac{1}{2}\sum_{i=1}^{m} 2\left(\sum_{j=1}^{n} A_{ij}x_j - z_i\right) A_{ik} \\
&= \sum_{i=1}^{m} A_{ik}(Ax - z)_i = \sum_{i=1}^{m}\left(A^\top\right)_{ki}(Ax - z)_i \\
&= \left(A^\top (Ax - z)\right)_k, \qquad k = 1, \ldots, n.
\end{aligned}$$

Inhalt

Bedingungen für Minimum: Lösungsmethode Normalengleichung

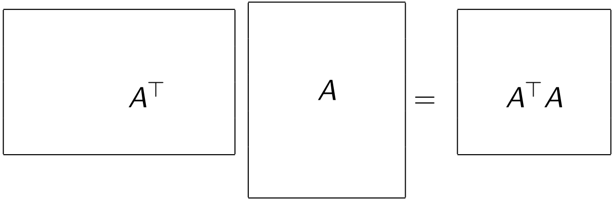
- Ableitung:

$$\frac{\partial f}{\partial x_k}(x) = \left(A^\top (Ax - z)\right)_k, \quad k = 1, \dots, n.$$

- Bedingung für Minimum:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_k}(x)\right)_{k=1}^{n} = A^\top (Ax - z) = 0 \iff A^\top Ax = A^\top z.$$

  Diese Gleichung heißt *Normalengleichung*.
- Lösung mit Gauß-Algorithmus ($A^\top A \in \mathbb{R}^{n \times n}$).



Geometrische Interpretation: Begriff Normalengleichung

- Bedingung für Minimum in x^* (Normalengleichung):

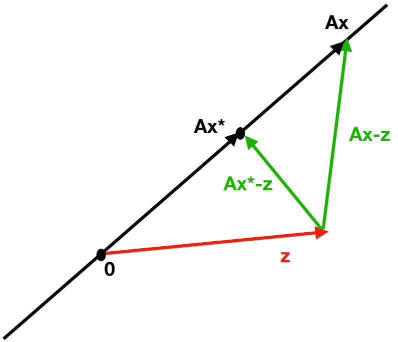
$$A^\top(Ax^* - z) = 0.$$

- $\{Ax \in \mathbb{R}^m : x \in \mathbb{R}^n\}$ ist linearer Unterraum des $\mathbb{R}^m$.
- x^* Minimum von $\|Ax - z\|$, wenn

$$Ax \perp (Ax^* - z) \quad \forall x \in \mathbb{R}^n,$$

- insbesondere für $x = e_i$ (Standardbasisvektoren): Es gilt $Ae_i = A_{*i}$ und

$$\begin{aligned} & A_{*i} \perp (Ax^* - z) && \forall i = 1, \dots, n \\ \Leftrightarrow \quad & A_{*i}^\top(Ax^* - z) = 0 && \forall i = 1, \dots, n \\ \Leftrightarrow \quad & A^\top(Ax^* - z) = 0. \end{aligned}$$



Beispiel: Ausgleichsprobleme

- Ausgleichsgerade:

$$A = \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix} \quad \Rightarrow \quad A^\top A = \begin{bmatrix} t_1 & \cdots & t_m \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^m t_i^2 & \sum\limits_{i=1}^m t_i \\ \sum\limits_{i=1}^m t_i & m \end{bmatrix}$$

Beispiel: Ausgleichsprobleme

- Polynom vom Grad k:

$$A^\top A = \begin{bmatrix} t_1^k & \cdots & t_m^k \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} t_1^k & \cdots & 1 \\ \vdots & & \vdots \\ t_m^k & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m t_i^{2k} & \cdots & \sum_{i=1}^m t_i \\ \vdots & & \vdots \\ \sum_{i=1}^m t_i & \cdots & m \end{bmatrix}$$

- Einträge in der Matrix haben oft sehr unterschiedliche Größenordnung:

$$m \ll \sum_{i=1}^m t_i^{2k}.$$

Beispiel: Ausgleichsprobleme

- Polynom vom Grad $k = 3$, Messpunkte $t = 1, \dots, m = 100$:

$$A^\top A = \begin{bmatrix} 1.4791e+13 & 1.7171e+11 & 2.0503e+09 & 2.5502e+07 \\ 1.7171e+11 & 2.0503e+09 & 2.5502e+07 & 3.3835e+05 \\ 2.0503e+09 & 2.5502e+07 & 3.3835e+05 & 5.0500e+03 \\ 2.5502e+07 & 3.3835e+05 & 5.0500e+03 & 1.0000e+02 \end{bmatrix}$$

- $\rightsquigarrow$ Problem wird sehr anfällig für kleine Ungenauigkeiten in z_i (Messwerte).

Inhalt

Lösungsmethode: Singulärwertzerlegung (Singular Value Decomposition)

Satz (Singulärwertzerlegung)

Für jede Matrix $A \in \mathbb{R}^{m\times n}$ gibt es orthogonale Matrizen $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ und $\Sigma \in \mathbb{R}^{m\times n}$ mit

$$\boxed{\quad A \quad} = \boxed{\quad U \quad} \underbrace{\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ 0 & \cdots & 0 \end{pmatrix}}_{=\Sigma} \boxed{\quad V \quad}$$

mit $\sigma_i > 0, i = 1, \ldots, n$.

Orthogonale Matrizen

Definition

Eine Matrix $U \in \mathbb{R}^{n \times n}$ heißt **orthogonal**, wenn ihre Zeilen und Spalten ein Orthonormalsystem bilden.

Lemma

Für eine orthogonale Matrix $U \in \mathbb{R}^{n \times n}$ gilt $U^\top U = UU^\top = I$, also $U^{-1} = U^\top$.

Beweis.

Für $E = (E_{ij})_{i,j=1}^n := AB$ gilt nach den Regeln der Matrixmultiplikation $E_{ij} = A_{i*}^\top B_{*j}$. Für $A = U^\top$ ist $A_{i*} = U_{*i}$, also

$$E_{ij} = A_{i*}^\top B_{*j} = U_{*i}^\top U_{*j} = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

da U orthogonale Spalten hat. Also ist E die Einheitsmatrix. $\square$

Orthogonale Matrizen

Lemma

Für eine orthogonale Matrix $U \in \mathbb{R}^{n\times n}$ *gilt*

$$\left.\begin{array}{rcl} (Ux)^\top Uy & = & x^\top y \\ \|Ux\| & = & \|x\| \end{array}\right\} \quad \forall x, y \in \mathbb{R}^n,$$

d. h. sie ist längen- und winkelerhaltend. Wenn U *orthogonal ist, dann ist es auch* $U^\top$*.*

Beweis.

Folgt aus $(Ux)^\top Uy = x^\top U^\top Uy = x^\top U^{-1} Uy = x^\top y$. Der Winkel zwischen zwei Vektoren wird über das Skalarprodukt definiert. Die Aussage über $U^\top$ folgt direkt aus der Definition. □

Definition

Die σ_i heißen Singulärwerte.

Beispiele

Beispiel

Drehungen und Spiegelungen als Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^n$ werden durch orthogonale Matrizen dargestellt:

- Drehung in $\mathbb{R}^2$ um Winkel α:

$$U = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix},$$

$$UU^\top = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} \cos^2\alpha + \sin^2\alpha & 0 \\ 0 & \cos^2\alpha + \sin^2\alpha \end{bmatrix} = I$$

- Spiegelung in $\mathbb{R}^2$ um eine (hier x_1-) Achse:

$$U = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad U = U^\top, \quad UU = I.$$

Lösung des Ausgleichsproblems mit SVD

- Benutze U orthogonal:

$$\|Ax - z\| = \|U\Sigma Vx - z\| = \|U^\top U\Sigma Vx - U^\top z\| = \|\Sigma Vx - U^\top z\|.$$

- Setze $\hat{x} = Vx, \hat{z} = U^\top z$.

$$\|Ax - z\| = \|\Sigma\hat{x} - \hat{z}\|.$$

- Struktur von Σ:

$$\Sigma\hat{x} = \begin{bmatrix} \sigma_1\hat{x}_1 \\ \vdots \\ \sigma_n\hat{x}_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \hat{z} = \begin{bmatrix} \hat{z}_1 \\ \vdots \\ \hat{z}_n \\ \hat{z}_{n+1} \\ \vdots \\ \hat{z}_m \end{bmatrix} \Rightarrow \Sigma\hat{x} - \hat{z} = \begin{bmatrix} \sigma_1\hat{x}_1 - \hat{z}_1 \\ \vdots \\ \sigma_n\hat{x}_n - \hat{z}_n \\ -\hat{z}_{n+1} \\ \vdots \\ -\hat{z}_m \end{bmatrix}$$

Lösung des Ausgleichsproblems mit SVD

- Es gilt:

$$\min \|Ax - z\| \Longleftrightarrow \min \|Ax - z\|^2.$$

- Wir benutzen

$$\|Ax - z\|^2 = \|\Sigma\hat{x} - \hat{z}\|^2 = \left\| \begin{bmatrix} \sigma_1\hat{x}_1 - \hat{z}_1 \\ \vdots \\ \sigma_n\hat{x}_n - \hat{z}_n \\ -\hat{z}_{n+1} \\ \vdots \\ -\hat{z}_m \end{bmatrix} \right\|^2 = \underbrace{\sum_{i=1}^{n}(\sigma_i\hat{x}_i - \hat{z}_i)^2}_{\text{minimal , wenn } \hat{x}_i = \frac{\hat{z}_i}{\sigma_i}} + \underbrace{\sum_{i=n+1}^{m} \hat{z}_i^2}_{\text{konstant für beliebiges } x}$$

Algorithmus: Lösung des Ausgleichsproblems mit SVD

- Berechne Singulärwertzerlegung von $A \rightsquigarrow U, V, \sigma := (\sigma_i)_{i=1}^n$
- Berechne
$$\hat{z} := U^\top z.$$
- Berechne
$$\hat{x}_i := \frac{\hat{z}_i}{\sigma_i}, \quad i = 1, \dots, n.$$
- Berechne Lösung
$$x := V^\top \hat{x}$$
- ... und ggfs. Wert der Kostenfunktion:
$$\|Ax - z\|^2 = \sum_{i=n+1}^{m} \hat{z}_i^2.$$

Lernziele

- Die Form eines Linearen Ausgleichsproblems kennen.
- Entscheiden können, ob ein beschriebenes Problem ein lineares oder ein nichtlineares Ausgleichsproblem ist.
- Daten und Modellparameter definieren können.
- Notwendige Bedingung für Minimum kennen.
- Beide Lösungsmethoden kennen und anwenden können.