Prof. Marc Alexa, Ugo Finnendahl,
Dimitrios Bogiokas, Gabriel Dogadov

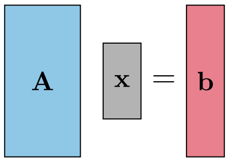
Wissenschaftliches Rechnen - Übung 2.2

Lineare Ausgleichsrechnung

20.11.2023 bis 24.11.2023

Aufgabe 1: Ausgleichsrechnung und lineare Regression

Ein überbestimmtes lineares Gleichungssystem $\mathbf{Ax} = \mathbf{b}$ mit $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{b} \in \mathbb{R}^n$ sieht wie folgt aus:



Dabei hat die Matrix $\mathbf{A}$ vollen Rang und es gibt mehr Gleichungen als Unbekannte. Zwar können überbestimmte lineare Gleichungssysteme genau eine oder keine Lösung haben, jedoch haben diese in der Praxis fast ausschließlich keine Lösung.

1. Warum haben überbestimmte Gleichungssysteme $\mathbf{Ax} = \mathbf{b}$ in der Praxis meist keine Lösung?

   Lösung

   Dies hat numerische und praktische Gründe:

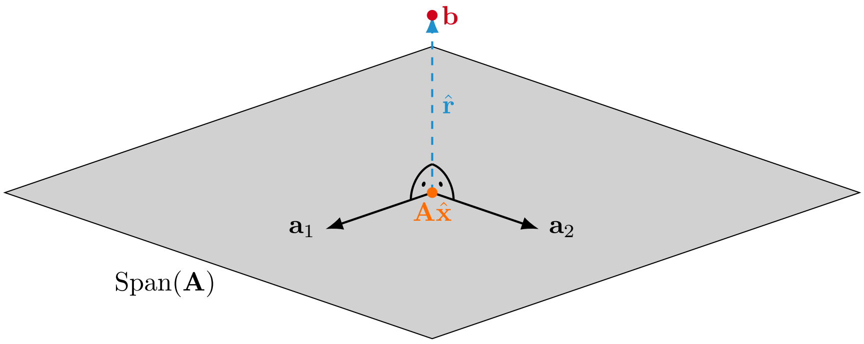
   - Die verwendeten Modelle, die auf stetigen (kontinuierlichen) Funktionen basieren, müssen diskretisiert werden.
   - Messungen haben in der realen Welt nur eine beschränkte Genauigkeit.
   - Die bei der Darstellung und Berechnung verwendeten Gleitkommazahlen erzeugen numerische Ungenauigkeiten.

   Lösung Ende

2. Wie kann man eine approximative Lösung des überbestimmten Gleichungssystems berechnen? Welche geometrische Interpretation hat die Lösung?

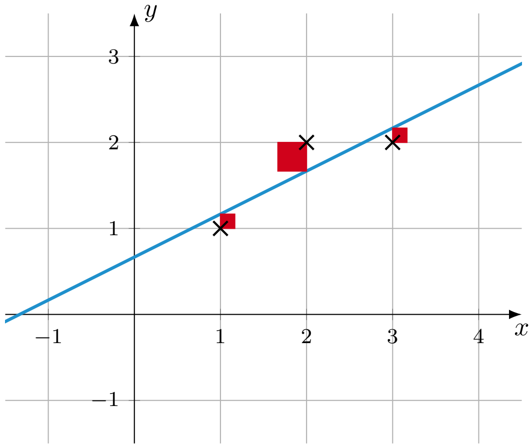
   Lösung

   Eine approximative Lösung kann durch die Anwendung der Normalengleichung bestimmt werden. Statt $\mathbf{Ax} = \mathbf{b}$ löst man das LGS $\mathbf{A}^\mathsf{T}\mathbf{Ax} = \mathbf{A}^\mathsf{T}\mathbf{b}$. Die Lösung der Normalengleichung $\hat{\mathbf{x}}$ minimiert die Zielfunktion $\|\mathbf{r}\|^2 = \|\mathbf{Ax} - \mathbf{b}\|^2$, wobei $\mathbf{r}$ als Residuum betitelt wird. Dessen (euklidische) Norm wird minimiert, indem man fordert, dass es orthogonal zu allen Spaltenvektoren von $\mathbf{A}$ steht, also ein Normalenvektor des Spanns von $\mathbf{A}$ ist. Der Vektor $\mathbf{b}$ wird also orthogonal auf den Spaltenraum von $\mathbf{A}$ projiziert, sodass die Lösung der Normalengleichung $\hat{\mathbf{x}}$ dafür sorgt, dass $\mathbf{A}\hat{\mathbf{x}}$ nächstmöglich an $\mathbf{b}$ ist.



Lösung Ende

3. Gegeben seien drei Datenpunkte $(x_i, y_i) \in \mathbb{R}^2$ mit den Werten $(1,1),(2,2)$ und $(3,2)$. Diese sind nochmal in der Abbildung dargestellt.



Im Folgenden möchten wir die Punkte mithilfe einer Geraden $f(x) = ax + b$ approximieren.

a) Modellieren Sie das Regressionsproblem mithilfe eines überbestimmten LGS $\mathbf{Ax} = \mathbf{b}$.

b) Warum können Sie sich sicher sein, dass das LGS keine Lösung hat, auch wenn Sie es nicht explizit lösen?

c) Lösen Sie das überbestimmte Gleichungssystem approximativ und zeichnen Sie die resultierende Funktion in die Abbildung oben ein.

d) Zeichnen Sie die Fehler ein, die durch das Ausgleichsproblem $\min \|\mathbf{Ax} - \mathbf{b}\|^2$ minimiert werden.

e) Nun möchten wir statt einer Geraden ein quadratisches Polynom $g(x) = ax^2 + bx + c$ verwenden. Wie sieht das lineare Gleichungssystem dazu aus?

f) Das Gleichungssystem aus der vorherigen Aufgabe ist jedoch eindeutig lösbar. Wie verhält sich das berechnete quadratische Polynom zu den Datenpunkten im Vergleich zur Geraden?

∗ Ist es auch möglich $\{\exp(x), \exp(2x)\}$ als Basisfunktionen für das Regressionsproblem zu verwenden? Falls ja, geben Sie das zugehörige Gleichungssystem an.

Lösung

a) Eine Möglichkeit für das LGS sieht wie folgt aus:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Sämtliche Permutationen der Zeilen und Spalten des LGS sind ebenfalls zulässig. Bei Spaltenpermutationen ist zu beachten, dass die Variablen im Losungsvektor ebenfalls permutiert werden müssen.

b) Das LGS fordert für alle drei Messpunkte (x_i, y_i) mit $i \in \{1,2,3\}$, dass $f(x_i) = y_i$. Das LGS hat also genau dann eine Lösung, wenn es eine Gerade $f(x) = ax + b$ gibt, welche alle drei Messpunkte trifft. Da dies nicht der Fall ist, hat das LGS keine Lösung.

c) Wir wenden die Normalengleichung an:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$
$$\Leftrightarrow \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

Da die Systemmatrix eine 2×2-Matrix ist, kann sie sehr einfach invertiert werden:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Somit gilt für die Lösung:

$$\begin{aligned} \begin{bmatrix} b \\ a \end{bmatrix} &= \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 1 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 2/3 \\ 1/2 \end{bmatrix} \end{aligned}$$

Alternativ kann dies auch mit einem Online-Rechner oder numpy gelöst werden. Also lautet unsere Funktion $f(x) = \frac{1}{2}x + \frac{2}{3}$.

d) Siehe Abbildung. Wider allen Erwartungen aus der Lösung der Teilaufgabe b) sind die Fehler nicht orthogonal zur Funktion ausgerichtet. Hierbei sollte beachtet werden, in welchem Raum das Residuum orthogonal ist, nämlich im $\mathbb{R}^3$, in welchem das zugehörige, überbestimmte LGS mit den Spaltenvektoren $(1, 1, 1)^\mathsf{T}$ und $(1, 2, 3)^\mathsf{T}$ sich abspielt.

e) Eine Möglichkeit für das LGS sieht wie folgt aus:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Erneut sind auch hier wieder Permutationen dieser Lösung zulässig.

f) Da das LGS eindeutig lösbar ist, gibt es genau ein Polynom, welches die drei Stützpunkte interpoliert. Dieses finden wir auch durch die Lösung des linearen Gleichungssystems. Da das vorherige LGS nur approximativ gelöst wurde, hat die Lösung dessen nur eine approximative Funktion ergeben.

$*$ Man kann alles Mögliche als Basisfunktionen wählen. Jedoch sollte man, damit die Normalengleichung eine eindeutige Lösung ergibt, darauf achten, dass die Basisfunktionen an den Stützstellen ausgewertet linear unabhängige Spalten der Systemmatrix ergeben. In den vorherigen Aufgaben waren es $\{1, x\}$ bzw. $\{1, x, x^2\}$, aber Exponentialfunktionen funktionieren analog:

$$\begin{bmatrix} \exp(1) & \exp(2) \\ \exp(2) & \exp(4) \\ \exp(3) & \exp(6) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Die resultierende Funktion $h : \mathbb{R} \to \mathbb{R}$ hat die Form $h(x) = a\exp(x) + b\exp(2x)$ und kann analog bestimmt werden.

Lösung Ende

Aufgabe 2: Cholesky-Zerlegung

Die Cholesky-Zerlegung einer Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ ist eine Faktorisierung von $\mathbf{A}$ in eine untere Dreiecksmatrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ und ihre Transponierte:

$$\begin{bmatrix} a_{1,1} & \cdot & \cdot & \cdot \\ a_{2,1} & a_{2,2} & \cdot & \cdot \\ \vdots & \vdots & \ddots & \cdot \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix} = \begin{bmatrix} \ell_{1,1} & 0 & \dots & 0 \\ \ell_{2,1} & \ell_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{n,1} & \ell_{n,2} & \dots & \ell_{n,n} \end{bmatrix} \begin{bmatrix} \ell_{1,1} & \ell_{2,1} & \dots & \ell_{n,1} \\ 0 & \ell_{2,2} & \dots & \ell_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \ell_{n,n} \end{bmatrix}$$

Eine solche Matrix $\mathbf{A}$ muss **symmetrisch** und (mindestens) **positiv (semi)definit** sein. Positiv (semi)definit bedeutet, dass $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0$ ($\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \geq 0$) für alle $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Dabei ist $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}$ die quadratische Form von $\mathbf{A}$, ein Polynom zweiten Grades in n Variablen, das nur quadratische Terme enthält.

1. Zeigen Sie, dass $\mathbf{A}$ symmetrisch und positiv semidefinit sein muss, um eine Cholesky-Zerlegung $\mathbf{A} = \mathbf{L}\mathbf{L}^\mathsf{T}$ zu besitzen.

Lösung

- Symmetrisch:

$$\mathbf{A}^\mathsf{T} = (\mathbf{L}\mathbf{L}^\mathsf{T})^\mathsf{T} = \mathbf{L}^{\mathsf{T}^\mathsf{T}}\mathbf{L}^\mathsf{T} = \mathbf{L}\mathbf{L}^\mathsf{T} = \mathbf{A}$$

- Positiv semidefinit:

$$\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{L}^\mathsf{T}\mathbf{x} = (\mathbf{L}^\mathsf{T}\mathbf{x})^\mathsf{T}(\mathbf{L}^\mathsf{T}\mathbf{x}) = \|\mathbf{L}^\mathsf{T}\mathbf{x}\|^2 \geq 0$$

Hinweis: Man kann die Matrix $\mathbf{L}$ durch eine beliebige Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ oder ihre Transponierte $\mathbf{A}^\mathsf{T}$ ersetzen und damit zeigen, dass für jede Matrix die Gram-Matrizen $\mathbf{A}\mathbf{A}^\mathsf{T}$ sowie $\mathbf{A}^\mathsf{T}\mathbf{A}$ symmetrisch und positiv semi-definit sind.

Lösung Ende

2. Gegeben ist folgende symmetrische Matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 5 & -7 \\ 3 & -7 & 14 \end{bmatrix}$$

Wie kann man überprüfen, ob $\mathbf{A}$ positiv definit ist?

Lösung

Dazu gibt es mehrere Methoden, von denen wir drei vorstellen. Diese funktionieren alle nur mit symmetrischen Matrizen.

- Eigenwertkriterium: Wir berechnen die Eigenwerte jener Matrix. Genau dann, wenn alle Eigenwerte größer als null sind, ist sie positiv definit. Da dies aufwendig ist und Eigenwerte in den nächsten Wochen behandelt werden, werden diese hier nicht berechnet.
- Hauptminorenkriterium: Wir betrachten die führenden Hauptminoren, also die Determinanten der Untermatrizen, die beim Element oben links anfangen. Genau dann, wenn alle führenden Hauptminoren positiv sind, ist die Matrix positiv definit.

$$\det 1 = 1 > 0$$
$$\det \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix} = 6 > 0$$
$$\det \mathbf{A} = 4 > 0$$

  Also ist $\mathbf{A}$ positiv definit.

- Cholesky-Zerlegung: Die in der Praxis am häufigsten angewandte Methode ist die Berechnung einer Cholesky-Zerlegung. Falls die Berechnung dieser nicht fehlschlagt und das Ergebnis (die Matrix $\mathbf{L}$) keine Null auf der Diagonalen hat, ist $\mathbf{A}$ positiv definit.

Lösung Ende

3. Berechnen Sie die Cholesky-Zerlegung der Matrix aus der vorherigen Teilaufgabe. In welchen Schritten kann die Berechnung schiefgehen, falls $\mathbf{A}$ nicht positiv definit ist?

Lösung

Die Cholesky-Zerlegung kann mit folgenden Formeln berechnet werden:

$$\ell_{i,j} = \begin{cases} \sqrt{a_{i,i} - \sum\limits_{k=1}^{i-1} \ell_{i,k}^2} & \text{falls } i = j \\ \frac{1}{\ell_{j,j}} \left(a_{i,j} - \sum\limits_{k=1}^{j-1} \ell_{i,k}\ell_{j,k} \right) & \text{sonst} \end{cases}.$$

An den Formeln erkennt man, wo etwas schiefgehen kann:

- Bei der Berechnung der Diagonaleinträge kann das Argument der Wurzel negativ werden. Das ist genau dann der Fall, wenn die Untermatrix von $\mathbf{A}$ beginnend oben links bis zum Element $a_{i,i}$ nicht positiv semidefinit ist.
- Die Berechnung der Nichtdiagonaleinträge geht schief, falls das zuvor berechnete Element der Diagonalen in der gleichen Spalte 0 ist. Deshalb fordert der Algorithmus, dass $\mathbf{A}$ sogar positiv definit sein muss.

Für die gegebene Matrix sähe die Berechnung wie folgt aus:

$$\begin{aligned} \ell_{1,1} &= \sqrt{a_{1,1}} = \sqrt{1} = 1 \\ \ell_{2,1} &= \frac{1}{\ell_{1,1}} a_{2,1} = \frac{1}{1}(-1) = -1 \\ \ell_{2,2} &= \sqrt{a_{2,2} - \ell_{2,1}^2} = \sqrt{5 - (-1)^2} = 2 \\ \ell_{3,1} &= \frac{1}{\ell_{1,1}} a_{3,1} = \frac{1}{1} 3 = 3 \\ \ell_{3,2} &= \frac{1}{\ell_{2,2}} (a_{3,2} - \ell_{2,1}\ell_{3,1}) = \frac{1}{2}(-7 - (-1) \cdot 3) = -2 \\ \ell_{3,3} &= \sqrt{a_{3,3} - \ell_{3,1}^2 - \ell_{3,2}^2} = \sqrt{14 - 3^2 - (-2)^2} = 1 \end{aligned}$$

Also ist:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 3 & -2 & 1 \end{bmatrix}.$$

Man sieht nun, dass $\mathbf{A}$ positiv definit ist, da $\mathbf{L}$ nur positive Diagonaleinträge hat.

Lösung Ende

4. Wie kann man die Cholesky-Zerlegung dazu nutzen, um lineare Gleichungssysteme $\mathbf{Ax} = \mathbf{b}$ zu lösen, falls $\mathbf{A}$ symmetrisch und positiv definit ist? Welche Vorteile hat dies gegenüber der Gauß-Elimination?

Lösung

Falls wir $\mathbf{A}$ in $\mathbf{A} = \mathbf{LL}^\mathsf{T}$ zerlegen, so ergibt sich das LGS

$$\mathbf{L}\underbrace{\mathbf{L}^\mathsf{T}\mathbf{x}}_{\mathbf{y}} = \mathbf{b},$$

welches sich in folgende zwei LGS aufteilen lässt:

$$\mathbf{Ly} = \mathbf{b} \quad \text{und} \quad \mathbf{L}^\mathsf{T}\mathbf{x} = \mathbf{y}.$$

Das erste LGS lasst sich durch Vorwärtseinsetzen lösen und das zweite durch Rückwärtseinsetzen. Beide Verfahren haben eine Laufzeit von $\mathcal{O}(n^2)$. Zwar erfordert die Berechnung der Cholesky-Zerlegung eine asymptotische Laufzeit wie bei der Gauß-Elimination von $\mathcal{O}(n^3)$, jedoch ist sie in der Praxis nur halb so groß wie bei der Gauß-Elimination. Zusätzlich erfordert die Lösung mehrerer LGS $\mathbf{Ax}_i = \mathbf{b}_i$ mit identischer Systemmatrix $\mathbf{A}$ nur die einmalige Berechnung der Zerlegung. Ferner ist die Cholesky-Zerlegung numerisch stabiler als die Gauß-Elimination.

Hinweis: Die Cholesky-Zerlegung eignet sich besonders zum Lösen von Ausgleichsproblemen. Die Gram-Matrix $\mathbf{B}^\mathsf{T}\mathbf{B}$ mit $\mathbf{B}^{n\times m}$ ist immer symmetrisch und positiv semidefinit (dasselbe gilt auch für $\mathbf{B}\mathbf{B}^\mathsf{T}$). Falls $\mathbf{B}$ aber vollen Rang hat und mehr Zeilen als Spalten $(n > m)$, so ist $\mathbf{B}^\mathsf{T}\mathbf{B}$ garantiert positiv definit. Somit lasst sich das lineare Ausgleichsproblem eindeutig mit der Cholesky-Zerlegung lösen.

Lösung Ende

5. Neben den Eigenschaften positiv (semi)definit gibt es weitere, analog definierte Eigenschaften: Eine quadratische Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist

$$\left.\begin{array}{ll} \text{negativ definit,} & \text{falls } \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} < 0 \\ \text{negativ semidefinit,} & \text{falls } \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \leq 0 \\ \text{indefinit,} & \text{sonst} \end{array}\right\} \text{ für alle } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}.$$

Wie kann man für eine symmetrische Matrix überprüfen, welche von diesen Eigenschaften sie hat?

Lösung

- Das Eigenwertkriterium funktioniert analog für alle anderen Eigenschaften: Sei $\mathbf{A} \in \mathbb{R}^{n\times n}$ eine **symmetrische** Matrix. Dann ist $\mathbf{A}$

positiv definit,	gdw. alle Eigenwerte größer als 0 sind.
positiv semidefinit,	gdw. alle Eigenwerte größer gleich 0 sind.
negativ definit,	gdw. alle Eigenwerte kleiner als 0 sind.
negativ semidefinit,	gdw. alle Eigenwerte kleiner gleich 0 sind.
indefinit,	sonst.

- Das Hauptminorenkriterium funktioniert jedoch anders:
  1) Matrix $\mathbf{A}$ ist genau dann positiv definit, wenn alle führenden Hauptminoren positiv sind.
  2) Matrix $\mathbf{A}$ ist genau dann negativ definit, wenn alle führenden Hauptminoren abwechselnd negativ und positiv sind (wobei der erste führende Hauptminor negativ ist).
  3) Für Semidefinitheit ist das Kriterium zu kompliziert.
- Man kann daran erkennen, dass eine Matrix nicht positiv definit ist, dass sie einen nicht-positiven Eintrag auf ihrer Diagonalen hat (analog nicht positiv semidefinit mit einem negativen Diagonaleintrag). Analog ist eine Matrix nicht negativ (semi)definit, falls sie einen nicht-negativen (positiven) Eintrag auf ihrer Diagonalen hat. Falls sie sowohl einen positiven als auch negativen Eintrag auf der Diagonalen hat, ist sie indefinit (nicht notwendig, aber hinreichend).

Lösung Ende