technische universität dortmund

fakultät für mathematik

Prof. Dr. Ivan Veselić Dr. Michele Serra

Synopse zur 7. Vorlesung – 10.09.2024

Vorkurs Mathematik

im Wintersemester 2024

Kapitel I – Methode der kleinsten Quadraten

Auch die Varianz kann man graphisch mit Hilfe des Prinzips der kleinsten Quadraten darstellen. Dies haben bereits in Aufg. 5 in der x-y-Ebene dargestellt. Hier ist es noch einfacher, da wir nur Punkte auf der t-Achse betrachten.

Frage: Wie ändert sich der Ausdruck

$$f(\bar{t}) = \frac{1}{N}\sum_{j=1}^{N}(t_j - \bar{t})^2 \quad \rightsquigarrow \quad f(z) = \frac{1}{N}\sum_{j=1}^{N}(t_j - z)^2$$

wenn ich den Bezugswert $\bar{t}$ durch ein beliebiges $z \in \mathbb{R}$ ersetze?

Behauptung: Der Wert wird größer, d.h., $f(z) \geq \mathrm{Var}(t)$ für alle $z \in \mathbb{R}$.

Beweis: Für jede reelle Zahl $z \in \mathbb{R}$ gilt offensichtlich

$$(t - z)^2 = (t - \bar{t} + \bar{t} - z)^2 = (t - \bar{t})^2 + 2(t - \bar{t})(\bar{t} - z) + (\bar{t} - z)^2$$

Analog für jeden Index j:

$$(t_j - z)^2 = (t_j - \bar{t} + \bar{t} - z)^2 = (t_j - \bar{t})^2 + 2(t - \bar{t})(\bar{t} - z) + (\bar{t} - z)^2$$

Wir addieren diese N Gleichungen und Multiplizieren mit $1/N$

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N}(t_j - z)^2 &= \frac{1}{N}\sum_{j=1}^{N}\left[(t_j - \bar{t})^2 + 2(t_j - \bar{t})(\bar{t} - z) + (\bar{t} - z)^2\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}(t_j - \bar{t})^2 + 2\frac{1}{N}\sum_{j=1}^{N}(t_j - \bar{t})(\bar{t} - z) + \frac{1}{N}\sum_{j=1}^{N}(\bar{t} - z)^2 \\
&= \mathrm{Var}(t) + 2(\bar{t} - z)\frac{1}{N}\sum_{j=1}^{N}(t_j - \bar{t}) + \frac{1}{N}N(\bar{t} - z)^2
\end{aligned}$$

Der mittlere Summand in der letzten Summe verschwindet, weil:

$$\frac{1}{N}\sum_{j=1}^{N}(t_j - \bar{t}) = \frac{1}{N}\sum_{j=1}^{N}t_j - \frac{1}{N}\sum_{j=1}^{N}\bar{t} = \bar{t} - \frac{N}{N}\bar{t} = 0$$

also bekommen wir

$$\mathrm{Var}(t) + N(\bar{t} - z)^2 \begin{cases} \geq \mathrm{Var}(t) \\ > \mathrm{Var}(t) \text{ falls } z \neq \bar{t} \end{cases}$$

Bemerkung: Ein alternativer Beweis befindet sich in Aufgabe 7!

Zurück zu Aufgabe 5

Falls wir das Skalarprodukt mit

$$\langle \vec{r}, \vec{t} \rangle := \sum_{j=1}^{N} r_j t_j$$

bezeichnen, so kann man $\bar{m}$ schreiben als

$$\bar{m} = \frac{1}{N} \langle \vec{r}, \vec{t} \rangle$$

Kovarianz

Die Kovarianz der Datensätze

$$(t_1, \ldots, t_N), \; (r_1, \ldots, l_N)$$

ist durch

$$\mathrm{Cov}(r,t) := \frac{1}{N} \sum_{j=1}^{N} r_j t_j - \left(\frac{1}{N} \sum_{j=1}^{N} r_j \right) \left(\frac{1}{N} \sum_{j=1}^{N} t_j \right) = \bar{m} - \bar{r}\bar{t}$$

gegeben.
Intuitive Interpretation:

- $\mathrm{Cov}(x,y)$ positiv: Die Datensätze $(x_1, \ldots, x_N)$ und $(y_1, \ldots, y_N)$ tendieren in die selbe Richtung, z.B. Körpergewicht und Größe bei Menschen
- $\mathrm{Cov}(x,y)$ negativ: Die Datensätze $(x_1, \ldots, x_N)$ und $(y_1, \ldots, y_N)$ tendieren in entgegengesetzte Richtungen, z.B. Lebenserwartung und Anzahl gerauchter Zigaretten

Wir haben also für den besten Wert von a aus dem System (7) den Ausdruck ($\mathbf{a_0}$) gefunden.

$$a = \frac{\bar{m} - \bar{r}\bar{t}}{\bar{k} - \bar{t}^2} = \frac{\mathrm{Cov}(r,t)}{\mathrm{Var}(t)}.$$

Exkurs zum Thema "Fallunterscheidung"

Erinnern wir uns an das Gleichungssystem

$$\begin{array}{ccccc} 0 & = & a\bar{t} & +c & -\bar{r} \\ 0 & = & a\bar{k} & +c\bar{t} & -\bar{m} \end{array} \tag{7}$$

damals haben wir die erste Gl. mit $\bar{t}$ multipliziert. Das können wir nur machen, ohne Informationen zu verlieren, wenn $\bar{t} \neq 0$. Falls aufgrund der Problemsdarstellung der Fall $\bar{t} = 0$ nicht ausgeschlossen ist, muss ich diesen separat behandeln:
die Gleichungen werden

$$\left.\begin{array}{cccccccc} 0 & = & +c & -\bar{r} & \Longleftrightarrow & c = & \bar{r} \\ 0 & = & a\bar{k} & -\bar{m} & \Longleftrightarrow & a = & \frac{\bar{m}}{\bar{k}} \end{array}\right\} (*)$$

($\bar{k} = \frac{1}{N} \sum_{j=1}^{N} t_j^2$ ist als Summe von Quadraten immer > 0)

Andererseits ergibt sich $(*)$ aus $(\mathbf{a_0})$ und $(\mathbf{c_0})$ wenn man $\bar{t}=0$ einsetzt

$$\begin{aligned} a &= \tfrac{\bar{m}-\bar{r}\bar{t}}{\bar{k}-\bar{t}^2} &\overset{\bar{t}=0}{=} \tfrac{\bar{m}}{\bar{k}} \\ c &= -\tfrac{\bar{m}-\bar{r}\bar{t}}{\bar{k}-\bar{t}^2}\bar{t}+\bar{r} &\overset{\bar{t}=0}{=} \bar{r} \end{aligned}$$

Durch Fallunterscheidung in der Rechnung/Beweisführung haben wir gezeigt, dass die Forderung $\frac{d}{da}Q=0$ und $\frac{d}{dc}Q=0$ zum System (7) führen, unabhängig vom Wert von $\bar{t}\in\mathbb{R}$. Das ist eine notwendige aber noch nicht hinreichende Bedingung für ein Minimum.

Ist in diesem Punkt (a_0,c_0) die 2. Ableitung positiv? Was bedeutet dies eigentlich?
Es gibt ja 4 Ableitungen, die wir in einer 2×2-Matrix darstellen können:

$$M(a,c)=\begin{pmatrix} \frac{d}{dc}\left(\frac{d}{dc}Q\right)(a,c) & \frac{d}{dc}\left(\frac{d}{da}Q\right)(a,c) \\ \frac{d}{da}\left(\frac{d}{dc}Q\right)(a,c) & \frac{d}{da}\left(\frac{d}{da}Q\right)(a,c) \end{pmatrix} = \begin{pmatrix} \underbrace{2\sum_{j=1}^{N}1}_{=2N>0} & 2\sum_{j=1}^{N}t_j \\ 2\sum_{j=1}^{N}t_j & \underbrace{2\sum_{j=1}^{N}t_j^2}_{>0} \end{pmatrix}$$

Determinante einer 2×2-Matrix

Die *Determinante einer 2×2-Matrix*

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ist gegeben durch $\det(A)=ad-bc$

Spezialfall des Hurwitz-Kriteriums

Ist $\frac{d}{da}Q(a_0,c_0)=0$ und $\frac{d}{dc}Q(a_0,c_0)=0$, so liegt im Punkt (a_0,c_0) ein (lokales) Minimum der Funktion $Q\colon\mathbb{R}^2\to\mathbb{R}$ vor falls, $\frac{d}{dc}\left(\frac{d}{dc}Q\right)(a_0,c_0)>0$ und $\det(M(a_0,c_0))>0$.

- Die Bedingung $\frac{d}{dc}\left(\frac{d}{dc}Q\right)(a_0,c_0)>0$ ist bereits festgestellt, da wir in der Definition von M sehen dass der erste Eintrag immer positiv, sogar unabhängig vom Wert von a und c, ist.
- Zur 2. Bedingung: Wir berechnen die Determinante:

$$\begin{aligned} \det(M) &= 2N\cdot 2\sum_{j=1}^{N}t_j^2-4\left(\sum_{j=1}^{N}t_j\right)^2 \\ &= 4N^2\left[\frac{1}{N}\sum_{j=1}^{N}t_j^2-\left(\frac{1}{N}\sum_{j=1}^{N}t_j\right)^2\right] \\ &= 4N^2(\bar{k}-(\bar{t})^2) \\ &= 4N^2\mathrm{Var}(t)>0 \end{aligned}$$

Beide Bedingungen des Hurwitz-Kriterium sind erfüllt und somit können wir schließen, dass in (a_0, c_0) tatsächlich ein Minimum von Q vorliegt.