technische universität dortmund

fakultät für mathematik

Prof. Dr. Ivan Veselić

Dr. Michele Serra

Synopse zur 5. Vorlesung – 06.09.2024

Vorkurs Mathematik

im Wintersemester 2024

Kapitel I – Methode der kleinsten Quadraten

Letztes mal haben wir festgestellt:

- Die erste Ableitung der Funktion $R \mapsto Q(R)$ ist Null im Punkt

$$R_0 = \frac{\sum_{j=1}^{N} U_j I_j}{\sum_{j=1}^{N} I_j^2}$$

- Dies ist der einzige Wert von R mit verschwindender Ableitung (Alle Umformungen waren Äquivalenzumformungen “ $\iff$ ”)

- In $R = R_0$ liegt ein Minimum vor, sogar ein striktes Minimum, da

$$\frac{d^2}{dR^2} Q(R_0) > 0$$

Minima und Maxima

Sei f eine Funktion mit Definitionsbereich $D = (a, b) \subset \mathbb{R}$ und Wertbereich $\mathbb{R}$:

$$f\colon D \to \mathbb{R}$$

- In einem Punkt $x_0 \in D$ liegt ein *lokales Minimum* von f vor, genau dann wenn es ein (kleines) $\epsilon > 0$ gibt, sodass für alle $x \in (x_0 - \epsilon, x_0 + \epsilon)$ gilt $f(x) \geq f(x_0)$.

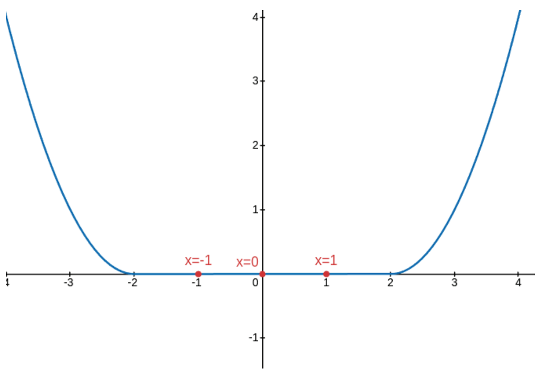
- In einem Punkt $x_0 \in D$ liegt ein *striktes lokales Minimum* von f vor, genau dann wenn es ein (kleines) $\epsilon > 0$ gibt, sodass für alle $x \in (x_0 - \epsilon, x_0 + \epsilon)$ mit $x \neq x_0$ gilt $f(x) > f(x_0)$.

- In einem Punkt $x_0 \in D$ liegt ein *lokales Maximum* von f vor, genau dann wenn es ein (kleines) $\epsilon > 0$ gibt, sodass für alle $x \in (x_0 - \epsilon, x_0 + \epsilon)$ gilt $f(x) \leq f(x_0)$.

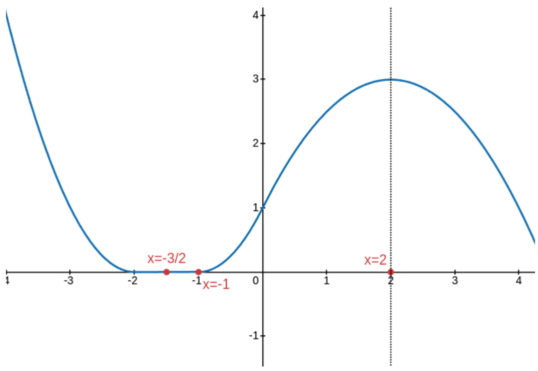
- In einem Punkt $x_0 \in D$ liegt ein *striktes lokales Maximum* von f vor, genau dann wenn es ein (kleines) $\epsilon > 0$ gibt, sodass für alle $x \in (x_0 - \epsilon, x_0 + \epsilon)$ mit $x \neq x_0$ gilt $f(x) < f(x_0)$.

- In einem Punkt $x_0 \in D$ liegt ein *globales Minimum* vor, genau dann wenn, für alle $x \in D$ gilt $f(x) \geq f(x_0)$.
- In einem Punkt $x_0 \in D$ liegt ein *striktes globales Minimum* vor, genau dann wenn, für alle $x \in D \setminus \{x_0\}$ gilt $f(x) > f(x_0)$.
- In einem Punkt $x_0 \in D$ liegt ein *globales Maximum* vor, genau dann wenn, für alle $x \in D$ gilt $f(x) \leq f(x_0)$.
- In einem Punkt $x_0 \in D$ liegt ein *striktes globales Maximum* vor, genau dann wenn, für alle $x \in D \setminus \{x_0\}$ gilt $f(x) < f(x_0)$.
- Wenn man bloß *Minimum* (bzw. *Maximum*) sagt ist *lokales Minimum* (bzw. *lokales Maximum*) gemeint.

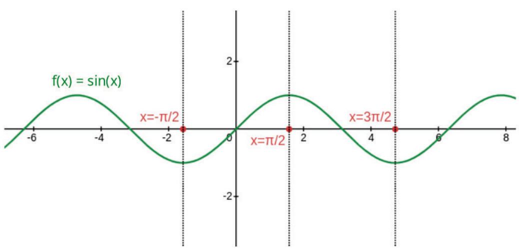
Warnung! Ein globales Minimum (bzw. Maximum) muss nicht existieren und, falls es eins gibt, muss nicht eindeutig sein.



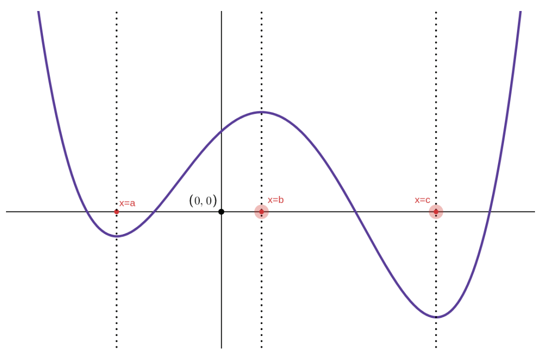
In (u.a.) $x = -1, x = 0, x = 1$ liegen globale Minima vor.



In $x = -1$ liegt ein lokales nicht-striktes Minimum vor. In $x = 2$ liegt ein lokales striktes Maximum vor.

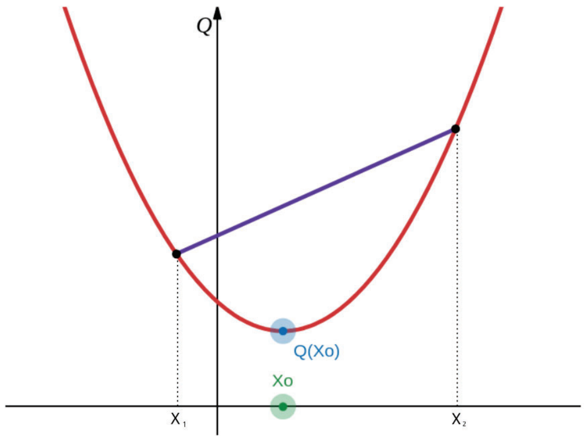


Für jedes $n \in \mathbb{Z}$ liegt in $x = \frac{\pi}{2} + 2n\pi$ ein globals striktes Maximum vor und in $x = \frac{\pi}{2} + 2n\pi$ ein globals striktes Minimum.



In $x = a$ liegt ein lokales striktes Minimum vor; in $x = b$ ein lokales striktes Maximum und in $x = c$ ein globales striktes Minimum.

- Da die 2. Ableitung von Q überall strikt positiv ist, kann man folgern, dass die Funktion Q folgende geometrische Form hat



- Wenn ich mich also von x_0 nach links oder rechts entferne, biegt sich die Kurve nach oben und dies schließt intuitiv weitere Minima aus

Konvexe Funktion

Eine Funktion $f\colon (a,b) \to \mathbb{R}$ heißt *konvex*, genau dann wenn für jedes Punktepaar $x_1, x_2 \in (a,b)$ mit $x_1 < x_2$ und für jedes $t \in [0,1]$ gilt

$$\underbrace{f(tx_1) + (1-t)x_2}_{\text{Wert der Funktion}} \leq \underbrace{tf(x_1) + (1-t)f(x_2)}_{\text{Wert der Sekante}} \tag{A}$$

Indem ich den Wert von t kontinuierlich von 0 bis 1 erhöhe, taste ich das gesamte Intervall von x_1 bis x_2 ab und an jeder Stelle liegt der Funktionsgraph unterhalb der Sekante. Mit der Notation des letzten Bilds, wo f die "rote" Funktion ist, dann ist die Sekante die Gerade durch die Punkte $(x_1, f(x_1))$ und $(x_2, f(x_2))$. Sie hat (siehe 3. Vorlesung) die Form

$$g(x) = \frac{f(x_1) - f(x_2)}{x_1 - x_2} x + f(x_1) \frac{f(x_1) - f(x_2)}{x_1 - x_2} x_1 \tag{B}$$

Bonusaufgabe: Überzeugen Sie sich von folgendem:
ist $f\colon (a,b) \to \mathbb{R}$ eine konvexe Funktion wie in (A), sind $x_1, x_2 \in (a,b)$ mit $x_1 < x_2$ und ist $g\colon \mathbb{R} \to \mathbb{R}$ eine Sekante wie in (B) dann, für alle $x \in (x_1, x_2)$ gilt $g(x) \geq f(x)$.

- Einsetzen des Wertes R_0 in die Geradengleichung ergibt

$$U = g(I) = \frac{\sum_{j=1}^{N} U_j I_j}{\sum_{j=1}^{N} I_j^2} \cdot I$$

  und dies ist die unseren Messdaten zugrundliegende beste Bestimmung des Wiederstandes.

Ergänzende Frage: Wie fehlerbehaftet ist unsere Messung? Das kann ich ablesen aus der Größe $Q(R_0)$ des quadratischen Fehlers Q an der Stelle R_0:

$$Q(R_0) = \sum_{k=1}^{N} (g(I_k) - U_k)^2 = \sum_{k=1}^{N} \left(\frac{\sum_{j=1}^{N} U_j I_j}{\sum_{j=1}^{N} I_j^2} \cdot I_k - U_k \right)^2$$

Wir betrachten wieder eine Situation, wo nur ein Parameter zu optimieren ist. Aber diesmal "der konstante Term".

Aufgabe 7 (Beste konstante Funktion)
Zu gegebenen Daten $(x_1, y1_), (x_2, y_2), \ldots, (x_N, y_N) \in \mathbb{R}^2$ finde unter der Schaar der waagerechten Geraden $g\colon \mathbb{R} \to \mathbb{R},\ g(x) = c$ (der Wert von a ist bei waagerechten Geraden Null) dijenige mit dem kleinsten quadratischen Fehler.
<u>Anders formuliert:</u> Bei welcher konstanten Funktion $g\colon \mathbb{R} \to \mathbb{R},\ g(x) = c$ ist der folgende quadratische Gesamtfehler am kleinsten?

$$Q = Q(c) = \sum_{j=1}^{N} (g(x_j) - y_j)^2 = \sum_{j=1}^{N} (c - y_j)^2$$

Wie bei Kurvendiskussion

- Berechne die 1. Ableitung:

$$\frac{d}{dc} Q(c) = \frac{d}{dc} \sum_{j=1}^{N} (c - y_j)^2 = \sum_{j=1}^{N} 2(c - y_j)$$

- Notwendige Bedingung für ein Minumum: $\frac{d}{dc} Q(c) = 0$. Wenn wir dies anwenden erhalten wir:

$$\begin{aligned} 0 &= 2\sum_{j=1}^{N} (c - y_j) = 2\sum_{j=1}^{N} c - 2\sum_{j=1}^{N} y_j \\ &\Updownarrow \\ 0 &= \sum_{j=1}^{N} c - \sum_{j=1}^{N} y_j \\ &\Updownarrow \\ N \cdot c &= \sum_{j=1}^{N} y_j \\ &\Updownarrow \\ c &= \frac{1}{N} \sum_{j=1}^{N} y_j \end{aligned}$$

Definition

Die Größe $\bar{y} := \frac{1}{N} \sum_{j=1}^{N} y_j$ heißt das *arithmetische Mittel* des Datensatzes $y_1, \ldots, y_N$.

- Teste die 2. Ableitung, um zu sehen ob tatsächlich ein Minimum vorliegt:

$$\frac{d^2}{dc^2} Q(c) = \ldots = \sum_{j=1}^{N} 2 = 2N > 0$$

Die 2. Ableitung ist positiv für alle Werte $c \in \mathbb{R}$, insbesondere für $c = \bar{y}$. In $c = \bar{y}$ liegt somit tatsächlich ein Minimum von Q vor.

- $c = \bar{y}$ ist das einzige globale Minimum und der Graph der Funktion Q ist nach oben gekrümmt, sie ist also konvex.

Reflektion

Der Wert $c = \bar{y}$ hängt nicht von den x-Werten der Daten $(x_1, y1_)$, $(x_2, y_2), \dots, (x_N, y_N)$ ab. Das ist aber natürlich. Beispielerweise hängt das Durchschnittsgewicht eine Population nicht davon ab, ob ich die Menschen alphabetisch oder nach Alter aufliste.

- Wie groß ist $Q(\bar{y})$? Also, wie groß ist der minimal erzielbare quadratische Fehler? Dies erfahren wir durch Einsetzen

$$Q(\bar{y}) = \sum_{j=1}^{N} (\bar{y} - y_j)^2 = \sum_{j=1}^{N} (y_j - \bar{y})^2 = \sum_{j=1}^{N} \left(y_j - \frac{1}{N} \sum_{j=1}^{N} y_j \right)^2$$

Um dem ganzen eine stochastische Interpretation zu geben, verstehen wir den Datensatz $(x_1, y1_)$, $(x_2, y_2), \dots, (x_N, y_N)$ als Zufallsexperiment bei dem jeder Datenpunk mit Wahrscheinlichkeit $1/N$ ausgewürfelt wird.

Einblick in der Stochastik

Annahme: die Werte $y_1, \dots, y_N$ werden jeweils mit Wahrscheinlichkeit $1/N$ angenommen. Die Größe

$$\mathrm{Var}(y) := \frac{1}{N} Q(\bar{y}) = \frac{1}{N} \sum_{j=1}^{N} (y_j - \bar{y})^2$$

heißt die *Varianz* des Datensatzes, und

$$\sigma := \sqrt{\mathrm{Var}(y)}$$

heißt die Standard-abweichung.

- Große Varianz $\Rightarrow$ Die Werte $(y_1, \dots, y_N)$ streuen stark.
- Kleine Varianz $\Rightarrow$ Die Werte $(y_1, \dots, y_N)$ streuen wenig.