Von der Mathematischen Biologie zur Systembiologie

(Vorlesung Prof. Dr. J. Timmer)

Aufgabenzettel Nr. 5

Aufgabe 9 (Übung): Lineare Regression

In dieser Aufgabe sollen Sie eine *linare Regression* programmieren.

$$\mathbb{R} \ni x \mapsto y(x) = a \cdot x + b + \varepsilon(x) \in \mathbb{R}, \qquad \varepsilon(x) \propto N(0, \sigma) \tag{1}$$

i.) Generieren Sie einen Vektor $\overline{y}$ der Länge $n = 21$ mit Zufallsdaten gemäß der Gleichung (1). Dabei sei $a = 2$, $b = 3$ und $\sigma = 0,5$ und $x \in [0, 10]$.

ii.) Nachdem Sie nun einen Datensatz $\overline{y} = \{\overline{y}_1, \overline{y}_2, ..., \overline{y}_n\}$ generiert haben, sollen Sie mit Hilfe einer von Ihnen programmierten Funktion die Parameter a und b *schätzen*. Eine effektive Methode ist es die Summe der Fehlerquadrate

$$f^2 : \mathbb{R}^2 \ni (a, b) \mapsto f^2(a, b) = \sum_{i=1}^{n} (\overline{y}_i - (a \cdot x_i + b))^2 \in \mathbb{R}^2_+ \tag{2}$$

zu minimieren. Im weiteren gehen sie folgendermaßen vor:

(a) Programmieren Sie die Funktion f^2 . Diese hat die Argumente (a, b). Der Rückgabewert ist f^2.

(b) Programmieren Sie nun die Gradientenfunktion

$$\nabla f^2 : \mathbb{R}^2 \ni (a, b) \mapsto (\nabla f^2)(a, b) = \left(\frac{\partial f^2}{\partial a}, \frac{\partial f^2}{\partial b} \right) \in \mathbb{R}^2. \tag{3}$$

(c) Programmieren Sie nun mit Hilfe der in den vorherigen Aufgabenteilen erarbeiteten Funktionen eine Rekursion, die die Fehlerquadrate minimiert. Dies geschieht indem Sie eine weitere Funktion programmieren, die die Parameter in Richtung des negativen Gradienten optimiert:

$$(a, b)^{(\ell+1)} = (a, b)^{(\ell)} - \lambda (\nabla f^2) \left((a, b)^{(\ell)} \right). \tag{4}$$

Dabei ist $(a, b)^{(0)} = (10, 20)$ der Startwert, und λ ein positiver Parameter, der die Schrittweite bestimmt.

(d) Stellen Sie nun der Verlauf von f^2 in Abhängigkeit von ℓ für verschiedene Werte von λ dar. Was fällt auf, und was heißt dies für den Optimierer?

Aufgabe 10 (Übung): Lokale Minima der f^2-Funktion

Betrachten Sie nun die Funktionen

$$\mathbb{R} \ni t \mapsto s(t) = \sin(c \cdot t) + \varepsilon \in \mathbb{R}, \qquad \varepsilon \propto N(0, \sigma = 0,2), \quad k = \frac{2 \cdot \pi}{10}. \tag{5}$$

i.) Simulieren Sie analog zur vorherigen Aufgabe einen Datensatz $\overline{s}$ auf dem Intervall $[0, 10]$ zur Funktion (5).

ii.) Betrachten Sie nun analog zur vorherigen Aufgabe die f^2-Funtion in Abhängigkeit des Parameters c. Stellen Sie das Ergebnis in einem (c, f^2)-Diagramm dar.

iii.) Was bedeutet das Ergebnis für die Parameterschätzung?