Technische Universität Dortmund
Fakultät Statistik
Lehrstuhl Computergestützte Statistik

Uwe Ligges
Marieke Stolte
Luca Sauer
Rudi Zulauf

Übung zur Vorlesung

Computergestützte Statistik

Wintersemester 2022/2023

Übungsblatt Nr. 12

Abgabe ist Dienstag, der 17.01.2023 bis 08:00 Uhr im Moodle

Aufgabe 1 **(4 Punkte)**

Implementieren Sie einen (μ, λ) Evolutionären Algorithmus (3 Punkte). Gehen Sie wie folgt vor:

1. Ihre Funktion soll die Signatur `myEvolution(f, lower, upper, mu, lambda, K)` haben. Dabei ist f die auf dem d-dimensionalem Interval [lower, upper] zu optimierende Funktion und K die Iterationsanzahl. Die weiteren Parameter sind Kontrollparameter des EA.
2. Die initiale Population der Größe μ ziehen Sie gleichverteilt aus [lower, upper].
3. In der t-ten Iteration wählen Sie zufällig λ Paare von Elternindividuen $\beta^{(1)}$ und $\beta^{(2)}$ aus, um einen Nachkommen zu erzeugen. Verwenden Sie dazu uniforme Rekombination sowie eine Gauß-Mutation mit $\sigma_t = 1 - 0.9 \cdot t/K$ (vgl. Folie 556).
4. Entfernen Sie alle Individuen außer den μ besten Nachkommen aus der Population.

Testen Sie Ihren EA auf $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$ (1 Punkt). Wählen Sie dabei für die weiteren Kontrollparameter: $\mu = 10 \cdot d$, $\lambda = 15 \cdot d$, $K = 1000 \cdot d$ für $d = 1, 2, 3$. Denken Sie daran, dass Ihr EA ein stochastisches Suchverfahren ist, daher sollten Sie mehr als eine Wiederholung betrachten. Was fällt auf, wenn Sie die Lösungen für die verschiedenen Dimensionen vergleichen und wie erklären Sie dieses Verhalten?

Hinweis: Das Testen dauert *lange*. Denken Sie daher daran, den entsprechenden Aufruf auszukommentieren und die Ergebnisse mit abzugeben (im PDF oder als Kommentare)!

Aufgabe 2 **(4 Punkte)**

Betrachten Sie die folgende Datensituation:

5.98	0.43	4.13	0.98	3.10	1.59	0.48	0.04	-0.60	4.18	2.09	3.92	1.78	5.59	4.71
-1.04	-0.39	0.42	-2.45	1.01	1.95	-0.14	0.79	0.29	-1.20	2.87	3.76	0.88	-2.31	0.00

Nehmen Sie an, dass diese Daten wie in Beispiel 3.5 aus einem Gauß'schen Mischungsmodell mit zwei Normalverteilungen stammen. Bestimmen Sie die Parameter $(\mu_1, \sigma_1, \mu_2, \sigma_2, \pi)$, die sich am besten an die Datenpunkte anpassen, verwenden Sie dazu den vorgestellten EM-Algorithmus.

a) (2 Punkte) Implementieren Sie den EM-Algorithmus, wählen Sie dabei zufällige Startwerte für die einzelnen Parameter. Eingabewert Ihrer Funktion soll lediglich der Beobachtungsvektor y sein. Rückgabewert sollen neben den geschätzten Parametern zum einen die finalen responsibilites und zum anderen der Verlauf der Likelihood über die Zeit sein.

b) (0.5 Punkte) Rufen Sie Ihre Funktion mehrfach auf und betrachten Sie die Schätzer. Welche erachten Sie als am realistischen? Warum?

c) (0.5 Punkte) Wie viele Iterationen braucht Ihr Algorithmus im Mittel bis zur Konvergenz?

d) (1 Punkt) Visualisieren Sie die Konvergenz, indem Sie (für einen speziellen Optimierungsverlauf) die Log-Likelihood gegen die Iterationszahl plotten. Geben außerdem Sie in einem weiteren Plot zu jedem Datenpunkt (y-Wert) die zugehörigen responsibilities an. Interpretieren Sie die Ergebnisse.

Aufgabe 3 **(4 Punkte)**

a) Schreiben Sie eine Funktion zur n-fach ($n \in \mathbb{N}$ gerade) wiederholten Invertierung einer Matrix $\mathbf{A}$, Sie dürfen dabei gerne auf die Funktion `solve` zurückgreifen. Ihre Funktion soll 1. die mittlere Differenz zwischen n-facher Inversen und der Ursprungsmatrix, sowie 2. die Spektral- und 3. die Frobeniusnorm zurückgeben. Sie dürfen auch `eigen` verwenden, jedoch keine Funktionen, die Ihnen direkt die Normen ausgeben.

b) Wenden Sie Ihre Funktion mit $n = 2$ auf Matrizen der Art:

$$\mathbf{A} = a + \begin{pmatrix} 9 & 9 & 8 \\ 9 & 8 & 8 \\ 8 & 8 & 7 \end{pmatrix}$$

an mit $a \in \{0, 10, 20, ..., 90, 190, 290, ..., 990, 1990, 2990, ..., 9990, ..., 9999990\}$. Visualisieren Sie Fehler und Normen. Was fällt Ihnen auf?