

Übungen zur Vorlesung

Praktische Optimierung, SoSe 2024

Prof. Dr. Günter Rudolph, Dr. Marco Pleines

<http://ls11-www.cs.tu-dortmund.de/people/rudolph/teaching/lectures/P0KS/SS2024/lecture.jsp>

Blatt Präsenz 8

24./27.06.2024

Abgabe: keine

Latin Hypercube Sampling und Kriging

In dieser Aufgabe sollen unterschiedliche Möglichkeiten für das initiale Design beim Kriging verglichen werden. Dazu werden verschiedene Varianten des *Latin Hypercube Sampling* (LHS) betrachtet. Erklären Sie LHS kurz in eigenen Worten.

Nutzen Sie die Implementierungen `qmc.LatinHypercube` aus dem Paket `scipy.stats`. Implementieren Sie die vier Varianten `centered`, `scramble`, `random-cd` und `lloyd`. Achten Sie darauf, dass die gezogenen Samples den gesamten Definitionsbereich abdecken.

Verwenden Sie folgende Zielfunktion:

$$f(x, y) = x^2 + y^2 + 10 \cos(xy), \quad x, y \in [-10, 10].$$

Für das Kriging-Modell soll die Funktion `GaussianProcessRegressor` aus dem Paket `sklearn.gaussian_process` ausschließlich mit Defaultwerten verwendet werden (obwohl hier eine Optimierung der Parameter prinzipiell sinnvoll sein könnte).

Führen Sie nun die folgenden Schritte aus:

1. Erzeugen Sie ein initiales Design mit 49 Punkten für x und y mithilfe eines der vier verschiedenen Typen von LHS.
2. Erstellen Sie auf Basis des Designs einen Datensatz, der für das Kriging verwendet werden kann.
3. Bilden Sie ein Kriging-Modell.
4. Nutzen Sie das Kriging-Modell, um eine Vorhersage auf einem Gitter mit jeweils 201 äquidistanten Punkten pro Dimension zu erstellen. Berechnen Sie aus den Vorhersagen den mittleren absoluten Abstand zur Wahrheit.

Für jeden der vier verschiedenen LHS-Typen soll das Vorgehen mindestens 20 mal wiederholt werden. Stellen Sie die Ergebnisse als Boxplots für die vier verschiedenen Typen dar. Welches Verfahren erzielt die besten Ergebnisse? Vergleichen Sie Ihre Ergebnisse mit den Ergebnissen, die für ein 7×7 -Gitter, eine gleichverteilte Stichprobe und eine zentrierte normalverteilte Stichprobe mit Standardabweichung $\sigma = 10$ (in allen Dimensionen) erzielt werden.

Messen Sie auch die Zeit, die im Durchschnitt pro Wiederholung benötigt wird und interpretieren Sie die Ergebnisse.