

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 8, Block B

17.06.2024

Abgabe: 27.06.2024

#### Aufgabe 8.1: Metamodellierung (10 Punkte)

(a) In dieser Aufgabe soll zunächst die Funktion

$$f_1(x) = 10 - x \cos(x), \quad x \in [-10, 10]$$

betrachtet werden.

Verwenden Sie folgende Verfahren und Parameter, um Metamodelle zu erstellen. Tasten Sie dabei die Funktion  $f_1$  an 11 äquidistanten Stellen im Definitionsbereich ab (d.h. bei -10, -8, -6, usw. bis 10) und führen Sie eine Anpassung der Modelle mit folgenden Parametern durch:

(1) Polynomielle Modelle mit den Graden  $k = \{1, 2, 5, 10, 15\}$ . Verwenden Sie hierzu das Paket *sklearn*, siehe Präsenzblatt 7.

(2) Kriging mithilfe der Funktion `GaussianProcessRegressor` ebenfalls aus dem Paket *sklearn*. Testen Sie folgende Kernels: Radiale Basisfunktion (RBF), Matern mit  $\nu \in \{0.5, 1.5, 2.5\}$  und `ExpSineSquared` mit  $p \in \{4, 6, 8\}$ . Diese finden sich in `sklearn.gaussian_process.kernels`. Nutzen Sie ansonsten ausschließlich Default-Einstellungen.

(3) Symbolische Regression mithilfe der Funktion `SymbolicRegressor` aus dem Paket *gplearn*. Setzen Sie hier die Populationsgröße fest auf 500, Turniergröße auf 10 und die Anzahl an Generationen auf 25. Das Stoppkriterium darf auf 0.01 erhöht werden, um Rechenaufwand zu sparen. Erklären Sie, welche jeweilige Funktion diese Parameter haben. Optimiert werden soll nur die Menge an zur Verfügung stehenden Funktionen. Testen Sie folgende Kombinationen:

$$\{+, -\}, \{\cdot, \div\}, \{+, \cdot\}, \{+, -, \cdot, \div\}, \{\sin, \cos\}, \{+, \cdot, \sin\}, \{-, \cdot, \cos\}, \{+, -, \cdot, \div, \sin, \cos\}$$

Beachten Sie, dass es sich bei (3) um ein stochastisches Verfahren handelt. Wiederholen Sie die Auswertung jeder Kombination daher 10-mal und entscheiden Sie mithilfe eines aussagekräftigen Boxplots, welche Kombination zu den besten Ergebnissen führt und erläutern Sie mögliche Gründe dafür. Lassen Sie sich dafür auch die erzeugten Funktionen anzeigen, indem Sie `print()` auf dem von `SymbolicRegressor` zurückgegebenen Objekt aufrufen.

Hinweis: Setzen Sie bei `SymbolicRegressor` den Parameter `n_jobs=-1`, um eine Parallelisierung der Fitnessfunktionsauswertungen zu aktivieren. Zudem kann es dem Verständnis helfen, während der Arbeit an der Implementierung `verbose=1` zu setzen.

Bewerten Sie die Güte der Modelle zu (1) und (2) mithilfe der mittleren absoluten Abweichung sowie der mittleren quadratischen Abweichung, indem Sie die erstellten Vorhersagemodelle an 201 äquidistanten Stellen im Definitionsbereich anwenden und mit den tatsächlichen Werten von  $f_1$  vergleichen. Tun Sie selbiges für das beste gefundene Modell der symbolischen Regression. Interpretieren Sie die Ergebnisse.

Visualisieren Sie die besten erstellten Modelle der einzelnen Verfahren in einer Grafik. Schauen Sie sich zudem die weniger guten Modelle an. Erstellen und interpretieren Sie weitere Grafiken, an denen

Ihnen Besonderheiten auffallen. Begründen Sie die Besonderheiten mit Bezug auf die jeweils zugrundeliegenden Parameter.

(b) Wiederholen Sie abschließend die Erstellung der Modelle und Parametersuche für die Funktion

$$f_2(x, y) = 5 \min(x, y) - 3x, \quad x \in [0, 100], \quad y \in [30, 70].$$

Tasten Sie die Funktion  $f_2$  in der Dimension  $x$  an 101 äquidistanten Stellen im Definitionsbereich ab. In der Dimension  $y$  sollen 100 zufällige Abtaststellen aus der Normalverteilung  $\mathcal{N}(50, 7)$  gezogen werden. Für die volle Punktzahl sollten Sie Ihre Funktionen zur Modellerstellung und -anpassung aus Aufgabenteil (a) hier wiederverwenden können.

Beim Kriging kann es für  $f_2$  passieren, dass einzelne Kernel nicht anwendbar sind und die Anpassung mit einer Fehlermeldung abbricht. Fangen Sie die Fehler im Code und berichten Sie, um welche Kernel es sich handelt.

Bei (3) darf hier auf die Suche der besten Kombination verzichtet werden. Wählen Sie eine geeignete Funktionsmenge und begründen Sie Ihre Wahl. Erhöhen Sie ggf. die Anzahl an Generationen und die Populationsgröße, um ein besseres Modell zu erhalten.

Bestimmen Sie die Güte der Modelle analog zu Aufgabenteil (a) auf einem äquidistanten Raster von  $101 \times 41$  über dem Definitionsbereich. Visualisieren Sie das beste Modell pro Verfahren in einem 3D-Plot und interpretieren Sie die Ergebnisse.