

Optimierung der Trainingsdatenverteilung für regressionsbasiertes maschinelles Lernen

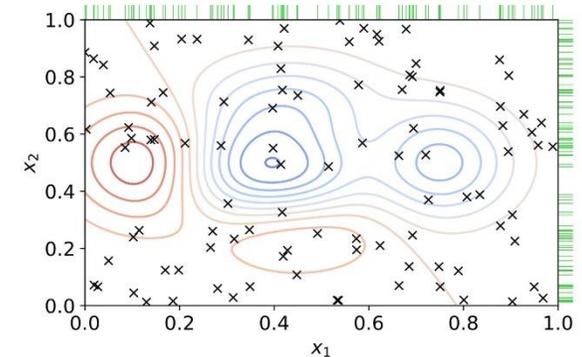
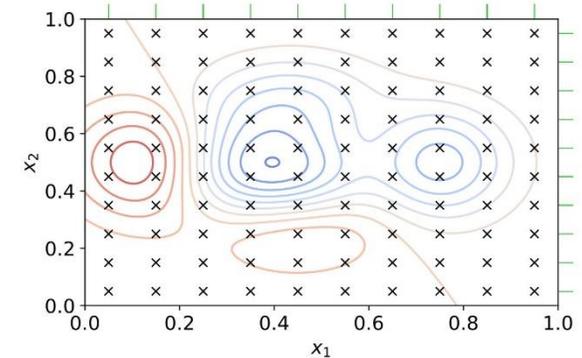
Methoden zur Auswahl der Datenpunkte beim Erlernen oder Approximieren einer Zielfunktion

Proseminar (1 Person, Homeoffice)

In verschiedenen Anwendungsfällen ist es notwendig, eine unter Umständen komplexe und hochdimensionale Funktion anhand von Messdaten zu erlernen. Typische Anwendungen sind die Erstellung einer einfachen Ersatzfunktion für einen komplizierten Zusammenhang oder die Notwendigkeit in dem hochdimensionalen Funktionsraum Optimierungen durchzuführen (z.B. bei der Hyperparameteroptimierung im Bereich des Maschinellen Lernens). Zum Erlernen dieser (Ersatz)-Funktion stehen aus der Wissenschaft verschiedene Methoden zur Verfügung, wie die unbekannte Zielfunktion auf Basis der Messdaten erlernt werden kann (Regressionen, Neuronale Netze, Gauß-Prozesse, ...).

Im Rahmen dieses Proseminars soll der Frage nachgegangen werden, wie die Datenpunkte, mit denen die Zielfunktion erlernt werden soll, optimal ausgewählt werden sollen. Die Datenpunkte könnten beispielsweise geordnet in einem Raster oder ungeordnet zufällig ausgewählt werden (siehe rechts). Im Rahmen dieser Arbeit soll eine wissenschaftliche Recherche über verschiedene Methoden der Datenpunktauswahl durchgeführt werden. Hierbei sollen die untersuchten Methoden insbesondere auf ihre Vor- und Nachteile im Hinblick auf Dateneffizienz und der dazu in Frage kommenden Methoden zum Erlernen der Zielfunktion gelegt werden.

Zum Erlernen des wissenschaftlichen Arbeitens soll anschließend eine schriftliche Zusammenfassung der Ergebnisse im Rahmen eines kurzen Berichts (ca. 10 Seiten) verfasst und ein kurzer wissenschaftlicher Vortrag (ca. 10 Minuten) vorgetragen werden.



Quelle: Alexander Elvers - Wikipedia

M.Sc. Felix Häusser



Tel.: 0391 / 67-58776

E-Mail: felix.haeusser@iat.tu-darmstadt.de

Web: <https://www.ccps.tu-darmstadt.de/ccps/>