

Aufgabenblatt 8

Abgabe: 30.11.2020

Aufgabe 1.

Implementiere die lineare Regression in Python. Verwende dafür einen geeigneten Datensatz, z.B. das Verhältnis von Salzgehalt und Temperatur in <https://www.kaggle.com/sohier/calcofi>. Gehe dabei wie folgt vor:

1. Downloade den Datensatz.
2. Lade ihn in Python und importiere die entsprechenden Spalten. (z.B. mit der Bibliothek `pandas`)
3. „Reinige“ ihn, d.h. entferne doppelte und fehlerhafte Zeilen.
4. Führe die Regression durch. (z.B. mit `sklearn`, `numpy`, `scipy.optimize.curve_fit`)
5. Plote dir das Ergebnis, d.h. die Datenpunkte und die Regressionsgerade. (z.B. mit `matplotlib`)

Wie gut ist die Vorhersage? (Welche Norm eignet sich hierfür?)

Zusatz:

- a) Splitte den Datensatz in Trainings- und Testdaten (z.B. 80% zu 20%). „Trainiere“ die Regressionsgerade auf den Trainingsdaten und schaue wie gut die Vorhersage auf den Testdaten ist. Warum ist das sinnvoll?
- b) Überlege dir mehrere Modelle für die Regression. Welches eignet sich am besten?