

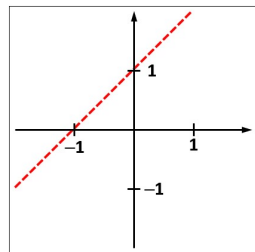
Methods & Algorithms FS 2023

Aufgabenblatt 1: Lineare Regression

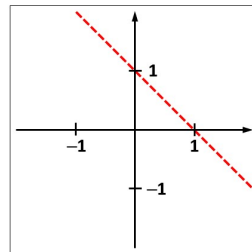
Die Bearbeitung der Aufgaben ist freiwillig; es erfolgt keine Bewertung.

Aufgabe 1

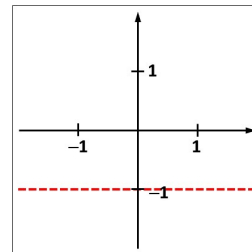
- a) Gegeben sind drei verschiedene Regressionsgeraden (rot-gestrichelte Linie). Bestimmen Sie für jede der Geraden A, B und C die beiden Parameter β_0 und β_1 durch Ablesen aus der Grafik.



A

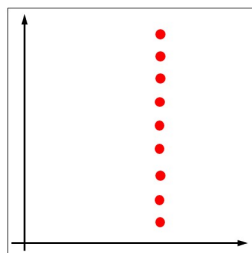


B

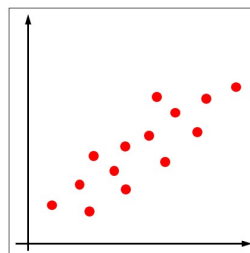


C

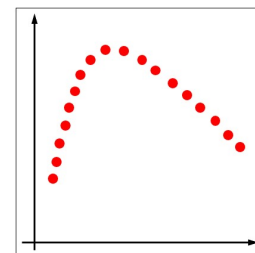
- b) Gegeben sind folgende Datenerhebungen A, B und C mit den zugehörigen Datenpunkten. Entscheiden Sie für jede Erhebung, ob eine Analyse mittels linearer Regression möglich ist und begründen Sie kurz Ihre Antwort, falls dies nicht geht.



A



B



C

Aufgabe 2

Der „California Housing“-Datensatz aus dem Paket `sklearn.datasets` beinhaltet verschiedene Informationen aus 20.640 Haushalten in Kalifornien. Laden Sie diesen Datensatz und nutzen Sie daraus die Spalten 2 (durchschnittliche Anzahl an Räumen) und 3

(durchschnittliche Anzahl an Schlafzimmern) für eine Regressionsanalyse. Gehen Sie dazu wie folgt vor:

```
from sklearn import datasets  
data = datasets.fetch_california_housing()  
data_x = data.data[:,2]  
data_y = data.data[:,3]
```

Teilen Sie die Daten im Verhältnis 80:20 in Trainings- und Testdaten auf und führen Sie damit eine Regressionsanalyse durch, um festzustellen, ob es einen linearen Zusammenhang zwischen Anzahl an Räumen und Anzahl an Schlafzimmern gibt. Bestimmen Sie für den Testdatensatz zudem den MSE sowie R^2 -Wert und plotten Sie Ihre Ergebnisse in einem Diagramm.