

Data Science 2

Sommersemester 2021

Übungsblatt 1

Hinweis: Da die Übung in dieser Woche noch nicht vollumfänglich stattfindet, ist dieses Übungsblatt als Wochenaufgabe vorgesehen. Sie können das Blatt unter der Woche bearbeiten, die Besprechung der Lösung findet in der Übung in der nächsten Woche am 13.4.2021 statt.

Aufgabe 1 (Auto MPG Datensatz)

Der Auto/Verbrauchs-Datensatz ist in der Datei `auto-mpg.csv` enthalten. Sie finden die Datei im Verzeichnis `Vorlesung/data`.

1. Laden Sie den Datensatz in einen DataFrame. Welche Attribute enthält der Datensatz? Welche Datentypen haben die Attribute?
2. Trainieren Sie ein lineares Regressionsmodell auf den Merkmalen `weight` und `displacement`, das den `mpg`-Wert vorhersagt. Bestimmen Sie den Trainingsfehler ihres Modells.
3. Erzeugen Sie einen Trainings- und Test-Datensatz und trainieren Sie das Modell auf 80% der Daten.
Ermitteln Sie den `mean-squared-error` und den `mean-absolute-error` auf dem Testdatensatz.
4. Vergleichen Sie den ermittelten `mean-absolute-error` auf den Testdaten mit dem Mittelwert und der Standardabweichung von `mpg` auf den Testdaten.
Erstellen Sie eine Vorhersage, die jedem Beispiel der Testdaten den `mpg`-Mittelwert zuweist. Wie groß ist der `mean-absolute-error` auf dieser Vorhersage?
5. Probieren Sie, das Vorhersagemodell auf den Trainings- und Testdaten zu verbessern, in dem Sie weitere Merkmale für das Training und die Vorhersage hinzunehmen. Welche Attribute eignen sich dafür?