

Data Science 2

Übungsblatt 1

Aufgabe 1 (Auto MPG Datensatz)

Der Auto/Verbrauchs-Datensatz ist in der Datei **auto.csv** enthalten. Sie finden die Datei unter der URL:

`https://data.hsbo.de/auto.csv`

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?