

Übungsblatt: Lineare Verfahren

In dieser Übung werden Sie selbst das Verfahren der *linearen Regression* implementieren. Für diesen Zweck wird Ihnen eine Python-Datei mit vorgefertigten Hilfsmethoden und Methodenköpfen, die Sie implementieren müssen, zur Verfügung gestellt.

In dieser Übung werden Sie mit einem Datenset über Hauspreise in Boston arbeiten. Dieses liegt im Ordner *data* als csv-Datei und wird mittels *Pandas* importiert. Das Datenset enthält 13 Features sowie die zugehörigen Hauspreise (der Name dieser Spalte ist *target*).

Mittels der Methode *get_linear_regression_training_set_from_df* können Sie sich ein Trainingsset für beliebige Features ausgeben lassen.

Aufgabe 1

In dieser Aufgabe sollen Sie sich zunächst die Datenpunkte für verschiedene Features in Relation zu den Hauspreisen ansehen. Zu diesem Zweck können Sie die Methode *plot_features_from_df* nutzen.

Schauen Sie sich die Features 2, 3, 4, 5 und 9 an.

Welches dieser Features lässt sich am besten mit einem linearen Modell darstellen?

Geben Sie Ihre Antwort als String in der Methode *question_1* zurück. Ihre Antwort muss der Name eines Features in derselben Schreibweise wie im Datenset sein.

Welches dieser Features lässt sich am schlechtesten mit einem linearen Modell darstellen?

Geben Sie Ihre Antwort als String in der Methode *question_2* zurück. Ihre Antwort muss der Name eines Features in derselben Schreibweise wie im Datenset sein.

Aufgabe 2

Implementieren Sie nun die Methoden *closed_form_solution* und *mse*.

In der Methode *closed_form_solution* sollen die optimalen Model-Parameter für das übergebene Datenset ermittelt werden. Nutzen Sie dazu die aus der Vorlesung bekannte Formel, um die Modell-Parameter direkt zu berechnen.

In der Methode *mse* sollen Sie die Kostenfunktion *Mean-Squared-Error* implementieren. Mittels der Kostenfunktion kann das Model evaluiert werden.

Interlude

Die Methoden *closed_form_solution* und *mse* werden genutzt, um die in Aufgabe 1 betrachteten Features einzeln zu evaluieren. Entspricht das Verhältnis der *MSEs* zueinander dem was Sie in Aufgabe 1 erwartet haben?

Aufgabe 3

Berechnen Sie die optimalen Model Parameter für die Features 2, 3, 4, 5 und 9 zusammengekommen.

Welches dieser Features hat den größten Einfluss auf die Vorhersage?

Geben Sie Ihre Antwort als String in der Methode *question_4* zurück. Ihre Antwort muss der Name eines Features in derselben Schreibweise wie im Datenset sein.

Welches dieser Features hat den geringsten Einfluss auf die Vorhersage?

Geben Sie Ihre Antwort als String in der Methode *question_4* zurück. Ihre Antwort muss der Name eines Features in derselben Schreibweise wie im Datenset sein.

Optional

Nutzen Sie die Methoden, um verschiedene Features und Kombinationen von Features zu evaluieren.