

## Linear Model

### 1 Aufgaben

Die erste Aufgabe können Sie im Kopf bzw. mit Stift und Papier bearbeiten. Für die darauf folgenden verwenden Sie bitte die von uns bereitgestellten Python-Templates.

#### 1.1 Einfaches 2D-Beispiel

In Abb. 1 ist eine Punktwolke dargestellt. Außerdem sind zwei Geraden eingezeichnet, welche die Ein- und Ausgabedaten linear annähern. Sie sollen nun mithilfe der Fehlerquadratsumme beurteilen, welche Gerade die bessere Näherung liefert.

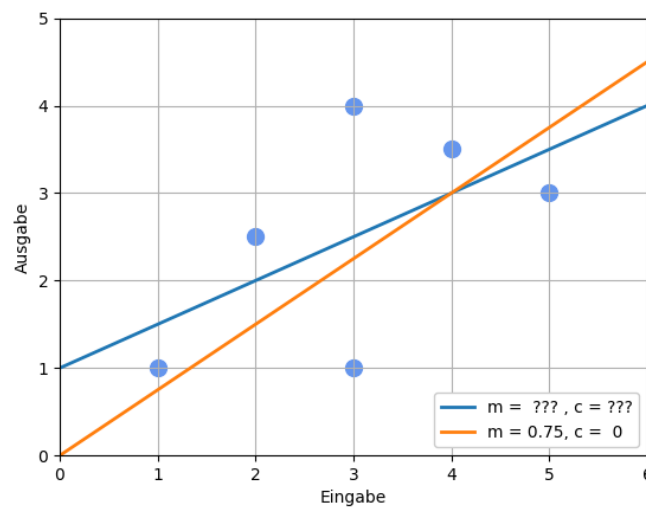


Abbildung 1: Punktwolke mit bekannten Ein- und Ausgaben, für die zwei Ausgleichsgeraden vorgeschlagen wurden.

- Berechnen Sie den Mittelwert der Ausgabedaten.
- Berechnen Sie die Parameter  $m$  und  $c$  der blauen Ausgleichsgerade.
- Berechnen Sie die Fehlerquadratsumme beider Geraden.

- (d) Eine der beiden Geraden wurde durch eine lineare Regression berechnet. Welche ist es?

## 1.2 Erste lineare Regression

Verwenden Sie für diese Aufgabe die Datei `lm_error.py`. Beim ersten Ausführen (ohne Änderungen) sollte ein Fenster ähnlich Abb. 2a erscheinen.

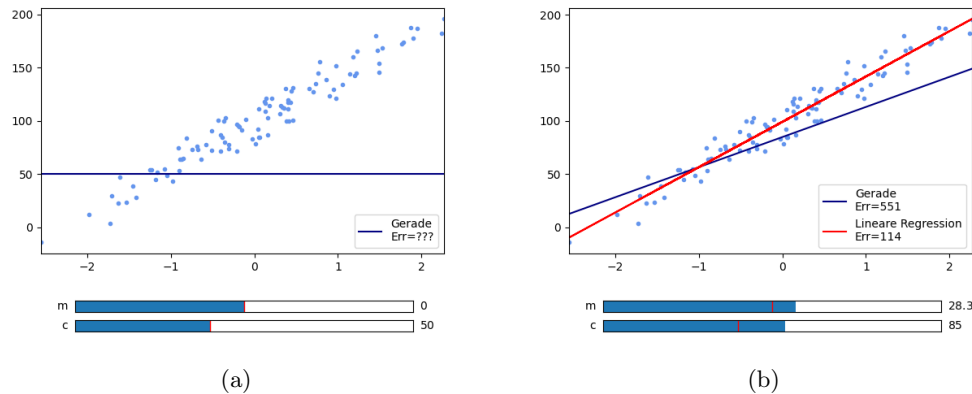


Abbildung 2

Die dargestellten Koordinaten (X, Y) liegen in den Arrays `X` und `Y` vor. Es ist eine Gerade eingezeichnet, welche über die beiden Slider `m` und `c` kontrolliert werden kann. Diese beiden Parameter entsprechen den Parameter des linearen Modells. Die vorgegebene Funktion `update` wird aufgerufen, sobald Sie den Wert eines Sliders ändern.

- Lesen und verstehen Sie den vorgegebenen Code. Benutzen Sie den Debug Modus.
- Erweitern Sie die Funktion `update` um eine Berechnung der Fehlerquadratsumme wie in Aufgabe 1.1 und erweitern Sie die Legende um diesen Wert.
- Versuchen Sie nun händisch die optimalen Parameter `m` und `c` zu finden. (Sie können diese Werte im Code hinterlegen, damit sie beim erneuten Ausführen darauf zurückgreifen können.)
- Führen Sie nun mittels der scikit-learn-Klasse `LinearRegression` eine lineare Regression des synthetischen Datensatzes durch und plotten Sie die so berechnete Ausgleichsgerade.
- Verwenden Sie die scikit-learn-Funktion `mean_squared_error` um die Fehlerquadratsumme der Regression zu bestimmen und erweitern Sie die Legende um diesen Wert.

### 1.3 Immobilienbewertung

Verwenden Sie für diese Aufgabe die Datei `house_data.py`. Das Ziel ist es, den Preis einer Immobilie basierend auf ihren Eigenschaften vorherzusagen. Gegeben sind vier verschiedene Parameter (Verkaufsdatum, Alter der Immobilie, Abstand zur nächsten Bahnhaltestelle und Anzahl der Supermärkte in der Umgebung) sowie der Immobilienpreis.

Datensatz:

<https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set>

- (a) Verwenden Sie nur einen Parameter für die Vorhersage. Berechnen Sie den  $R^2$ -Wert des Modells.
- (b) Welcher Parameter eignet sich am besten?
- (c) Verwenden Sie nun alle Parameter für das Lineare Modell. Ist die Vorhersage nun besser?