

Grundlagen des maschinellen Lernens (2023/24)

Aufgabenblatt 1

Abgabetermin: Freitag, 27. Oktober 2023

Die Abgabe erfolgt durch Hochladen der Lösungen in ILIAS.

Aufgabe 1

Beweisen Sie die folgende, auf den Vorlesungsfolien postulierte Identität:

$$\frac{\sum_{i=1}^N (x_i y_i - \bar{x} \bar{y})}{\sum_{i=1}^N (x_i^2 - \bar{x}^2)} = \frac{\sum_{i=1}^N (x_i - \bar{x}) y_i}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Dabei sind $(x_1, y_1), \dots, (x_N, y_N)$ beliebige Paare reeller Zahlen, und \bar{x} und \bar{y} bezeichnen die jeweiligen arithmetischen Mittel $\frac{1}{N} \sum_{i=1}^N x_i$ und $\frac{1}{N} \sum_{i=1}^N y_i$.

Erstellen Sie bitte Ihre Lösung für die Abgabe mit Hilfe von LaTeX (wobei Sie das LaTeX-Sourcefile des Aufgabenblattes zur Unterstützung heranziehen können) und reichen Sie das erzeugte PDF-File ein.

Aufgabe 2

Nimmt man einen der beiden obigen gleichwertigen Ausdrücke als w_1 und setzt w_0 gleich $\bar{y} - w_1 \bar{x}$, dann legen w_0 und w_1 , wie in der Vorlesung gezeigt, die Regressionsgerade für die gegebenen Eingabe-Ausgabe-Paare fest. Implementieren Sie diesen Ansatz zur Berechnung der Regressionsgerade in Python.

Definieren Sie dazu eine Funktion `SimpleRegression`, der der gegebene Datensatz als Liste von Zweiertupel übergeben wird und die als Ausgabe ein Zweiertupel bestehend aus Achsenabstand w_0 und Steigung w_1 der Regressionsgerade liefert.

Wenden Sie Ihre Funktion auf den in Abschnitt 2.1 des Jupyter-Notebooks zu Kapitel 4 von Gérons Buch beschriebenen, mit Hilfe von Zufallszahlen erzeugten Datensatz an. Erzeugen Sie eine graphische Darstellung des Datensatzes sowie der generierten Regressionsgeraden. (Dazu können Sie sich an besagtem Abschnitt im Notebook orientieren, in dem die `matplotlib`-Library für diese Zwecke verwendet wird.) Betten Sie das alles in ein eigenes Jupyter-Notebook ein und laden die das entsprechende ipynb-File in ILIAS hoch.