Machine Learning
Exercise 1: Introduction to linear regression with scikit-learn and scipy

Prof. Dr. Thomas Kopinski

August 9, 2022

Abstract

In this week's exercise you will become familiar with using the python module scikit-learn to perform a linear regression on a given dataset. In addition you will use the more flexible curve_fit function from the scipy package. The data can be checked out from this courses's Git repository.

Task 1: Getting familiar with scikit-learn

- Additional information about implementing linear regression in python with the help of scikit-learn can be found here

- The repository for this course can be found here

Task 2: Linear regression with scikit-learn

- import the necessary modules

- load the file with the name "simple_regression.csv" from the data subfolder with numpy or pandas

- instantiate a LinearRegression model from the scikit-learn library

- fit a linear model $f(x) = y = mx + b$ to the given data (reshaping of the data might be necessary)

- display the values of the coefficients and plot the data and fitted function

- interpret the results

Task 3: Linear regression with scipy

- import the necessary modules

- load the file with the name "simple_regression.csv" from the data subfolder with numpy or pandas

- define a model function that returns a linear mapping of the variable x

- fit that linear model to the data with scipy.optimize.curve_fit

- display the values of the coefficients and covariance matrix

- what is the meaning of the covariance matrices entries?

- plot the data and fitted function

- interpret the results and compare it to the scikit-learn version

- can this method be used in a more general way?