

Python Programming

Exercise 11: Data Visualization

Prof. Dr. Thomas Kopinski

December 9, 2021

Abstract

This exercise will focus on displaying information in a concise way.

Task 1: Plotting and explaining

Download the Boston housing data set from [here](#)
Use the packages seaborn and matplotlib.

- The link mentions what each of the attributes holds - read about the data first
- Prepare the data, load it into a dataframe and use the describe() function to start exploring the data
- Using boxplots on each attribute: Explain the distribution and variation in the data as well as potential outliers
- Using distribution plots from seaborn: what kind of distribution does the data in each variable hold? Is the data skewed in any way?
- Use a heatmap to display the correlation for the whole attribute matrix - what does it show? Which attributes correlate most and why?
- Use scatterplots to display the data points for all combinations of attributes - can you plot a regression line here? (cf. code example given in lecture)
- Bonus: Can you use the regression line to predict values for data in the data set?