DAIS2022: Assignment #2
Plotting, Correlation, Regression

Kyra Ahrens, Burhan Hafez, Kerim Erekmen
{kyra.ahrens,burhan.hafez}@uni-hamburg.de
{kerim.erekmen}@informatik.uni-hamburg.de

University of Hamburg — 14.04.2022
Due: 28.04.2022 8 p.m.

Introduction

In this assignment, we want to get familiar with data exploration. We will have a look at how to create plots with Python, how to calculate correlations and regressions using the *NumPy* library.

The assignment has 100 points in total, distributed over the tasks. You need to get at least 50 points to pass this assignment.

1 Plotting and Correlations: 35 points

This task introduces the plotting library *matplotlib* and its applications in regards to correlations. Please do all the tasks specified in the attached Jupyter notebook *PlottingCorrelation.ipynb*. For a closer lookup you can visit the link `https://matplotlib.org/stable/tutorials/index`.

2 Plotting gone wrong: 10 points

This task is modelled as part of a moodle quiz. Please solve questions 1-3 of the quiz for this task. You will find it on the moodle page of the course or by clicking this link:
`https://lernen.min.uni-hamburg.de/mod/quiz/view.php?id=87082`.

3 Statistical Independence: 20 points

Another technique in data mining is to test for independence between variables using statistical tests. An example is the χ^2 test (cf. Lecture 3), which you are now asked to perform on an example.
The example is the second part of the aforementioned moodle quiz. Please solve questions 4-11 of the quiz for this task. You will find it on the moodle page of the course or by clicking this link:
`https://lernen.min.uni-hamburg.de/mod/quiz/view.php?id=87082`.
If you want to learn more about the the χ^2 test, follow the link: `https://www.researchgate.net/publication/277935900_Chi-square_test_and_its_application_in_hypothesis_testing`.

4 Regression: 35 points

This task builds upon the knowledge with plotting and correlations to perform a regression. Please do all the tasks specified in the attached Jupyter notebook *Regression.ipynb*.

5 Next Assignment

For the next tutorial, prepare the following topics:

- Decision Trees: Entropy, Classification with decision trees
  "Data Mining, Practical Machine Learning Tools and Techniques", Witten/Frank/Hall/Pal, ***Chapter 6***
  "Machine Learning, Tom Mitchell", ***Chapter 3***
- Neural Networks: Perceptrons, Multi-Layer perceptrons (MLP), Neural network training, Backpropagation algorithm
  "Neural Networks - A Systematic Introduction", Raul Rojas, ***Chapters 3 (until 3.3.4), 4 (until 4.2.5), 7 (until 7.3.4)****, freely available* `https://page.mi.fu-berlin.de/rojas/neural/`