Final Project

17.8.2021

The purpose of the final project is to apply some of the methods that we covered this semester to a dataset of your choice. The questions are open-ended, and you are required to decide both *what* to investigate, and *how* to do so. The project will be graded as a whole, and not on a per-question basis, so feel free to put an emphasis on the aspects you find more interesting.

The project should be submitted either as a Jupyter ntoebook, or a PDF/DOC file. If submitting a notebook, please attach any data files necessary to run it.

1. [Data] choose a dataset for the project. You are free to use whatever data you find interesting, as long as it has reasonable content that allows to answer the following meaningfully:
2. [Descriptive statistics] display the content of the dataset with a short descriptive analysis (plots, tables, etc.)
3. [Hypothesis testing] formulate 1-2 hypotheses that arise from the data, and go through all the stages of the appropriate hypothesis test. Use a significance level of your choice (state it clearly!) What is your conclusion?
4. [Regression] apply 2 of the regression methods that we covered. Discuss the results.
5. [Clustering] apply one of clustering methods that we covered to the data. Discuss the results (did you find meaningful clusters? why did you choose this method?)
6. [Dimensionality reduction] apply PCA to some aspect of the data and visualize the results. Did you find what you expected?
7. [XAI] Train a classifier of your choice to predict some aspect of your data. Use an explanation method to give explanations for individual predictions in the test set. Are you satisfied by the explanations you got?