NEURAL NETWORKS – exercise 1

The aim of the exercise is to introduce/recall the tools and get acquainted with the data that we will use later in the course to evaluate neural networks as a machine learning method.

In this and subsequent exercises - 2 and 3 - you are allowed to use any libraries to operate on data (loading, exploratory analysis, preprocessing), obtaining model quality metrics and efficient matrix calculations. (However, you will have to implement the learning methods yourself!) A suggested, easy-to-use configuration is:

- Python
- Pandas for viewing data
- Numpy for matrix calculations
- Scikit-learn to calculate quality metrics

During the first class, you have time to familiarize yourself with the relevant libraries/remember them from the previous semester. Then, load and analyze the Heart Disease dataset:

https://archive.ics.uci.edu/dataset/45/heart+disease

Exploratory analysis should include answers to the following questions:

- Is the set balanced in terms of the number of samples per class?
- What are the means and deviations of numerical features?
- For numerical features: is their distribution approximately normal?
- For categorical features: is the distribution approximately uniform?
- Are there missing features and what strategy can we use to replace them?

As a result of the work by next week, the first part of the summary report should be created, which will include exercises 1-4 and the code transforming the data into a matrix of numerical features (examples × features).

The exercise is graded on a scale of 0-10 points.