Theoretical Exercises

Excercise 3.1: (Theoretical) Comparison of techniques to obtain data

Consider the ways of obtaining data discussed in the lecture (experiments/surveys, databases, APIs, scraping). Compare these techniques concerning the following points:

a) How fast do we get the data

b) How trustworthy is the data

c) How big is the effort obtaining the data

Practical Exercises

Excercise 3.2: (Practical) Using an API

Perform a research concerning freely available APIs. Choose your favorite one. Write a program, which calls the API with different inputs, takes the data and store the data in a CSV file.
Hint: An url can be retrieved with the help of the requests package in python.

Excercise 3.3: (Practical) Labeling

In the lecture we learned how data could be collected. Unfortunately, collected data is not always complete, e.g. sometimes we need to label the data. In the Ilias course room, a file called numbers.zip[a], containing several images of numbers, is given.

a) Download the data and extract it into a proper directory.

b) Write a program to label the data: The program shows an image, receives a class as input of the user, saves the result properly.

c) Count how many images of the different classes are given - compare the result with you colleagues!

Hint: There are several libraries in python to read images and show them, one is called PIL another opencv.

[a] Part of the MNIST dataset, taken from `https://www.kaggle.com/datasets/scolianni/mnistasjpg`