Group Project: Tasks and Description

Lernen von Daten und Lineare Algebra

Team sizes: 2-4 persons
Deadline: Thursday 09.04.2020 @ 23:59 pm

Description

Organize yourself into groups consisting of 2-4 persons (contact Alice if you haven't done so far). Take the provided Jupyter Notebook from the Moodle page and investigate the provided Oktoberfest dataset. The goals of this project are that you come in touch with basic machine learning tasks and practice Linear Algebra with real-world data. There is no wrong and right, we just want you to play with the learned concepts of the lecture.

Tasks

The Jupyter notebook takes you step by step through the tasks. Your project consists of several minor tasks.

1. Data Analysis
    a. Look into the data matrix A. Which dimension does A have? Which features describe the data?
    b. Plot with Seaborn various data dependencies (e.g beer price vs. beer consumption) and depict the most interesting ones.
    c. (Fun Question): What could be the reason for the jump in the chicken price between 1999 and 2000?
    d. Factorize A with SVD.
        i.   Plot the diagonal S matrix. Can you estimate the rank with the singular values?
        ii. Perform a low-rank approximation of A
2. Learn from Data
    a. Choose a variable which you want to predict (e.g. predict the beer consumption based on all other features). Solve the least-squares problem for this variable.
    b. Drop some of the features and try to predict the variable you have chosen before. What do you observe in the least-squares error?
    c. What is the test error? What is the training error? What deviations do you observe between training and test set error?
    d. Compare the results obtained from solving the least-squares problem with the unnormalized dataset A with the results obtained from the normalized dataset B.

Final Submission

Craft a report as PDF File and upload it on Moodle. The report should consist of at most three pages. The report should contain the answers to all the questions given in Tasks.