

Exercise 5 on Machine Learning WS 2023/24  
Prof. Dr. Dominik Heider  
M. Sc. Arsam (Mohammad) Tajabadi  
Submission: Until 29.11.2023, 23:59 on Ilias

---

**Task 1. Regression (5 points)**

Use the Wilt dataset for this task. A description is available at:  
<https://archive.ics.uci.edu/ml/datasets/Wilt>. The data set is divided into a training data set and a test data set. Use the DataFrame from pandas as the data structure.

- a) Learn a function via regression that can determine (i) the attribute *SD\_Pan* using the other numeric attributes. Use the training data to do this. (2 points)
- a) Learn a function that predicts the attribute *Mean\_Red* based on the attribute *Mean\_Green*. Create a scatterplot that maps *Mean\_Green* to *Mean\_Red* and add the regression function (line plot) to it. (3 points)

**Task 2. Logistic regression (5 points)**

For this task, use the dataset *examresults.csv*. It is intended to check whether passing an exam is dependent on the time spent preparing for the exam.

- a) Calculate and plot a logistic regression for the data set. (3 points)
- b) Based on the estimated values, calculate (by hand) the chance of passing an exam if (i) one hour of work is invested and (ii) 10 more working hours are invested. (2 points)