

Voluntary Homework 2

Data Science (Summer term 2020)

Exercise 1 (Regression (Theory)) 5 points

Consider a hypothetical dataset with number of working hours as input (X) and throughput as output (Y). A simple linear regression between X and Y is uniquely defined as follows:

$$\hat{y} = \beta_0 + \beta_1 x$$

Where β_0 is 25 and β_1 is 0.5.

- What is the corresponding predicted value for the input $x = 10$?
- Compute the residual corresponding to $x = 6$ $y = 30$.
- How does \hat{y} change with respect to x when x increases by 1 unit?
- Rewrite the original regression equation in terms of x where x is the number of working hours measured in seconds. Compare predicted values of both of these equations for $x = 5$.

Exercise 2 (Regression (Practical)) 5 points

Read the description of and get familiar with the "housing.csv" dataset. The task is to create a predictive model that can predict the price of a house (last column in the provided dataset) given some of its descriptors.

- Preprocess the data by imputing, scaling etc., wherever necessary.
- Fit the LinearRegression from sklearn and report your results (r2_score)
- Examine 5 largest weights (largest absolute value) of the model, excluding bias weight. Try to interpret every of these values with respect to the meaning of the column (feature) that this value corresponds to.