

# Problem3

November 22, 2021

DAT405 Assignment 2

Noah Lanai - 9808252192 - 8h work

Theodor Jendle - 9901134339 - 8h work

Carl Hjalmarsson - 9305198930 - 8h work

When training on a data set, the used model runs a risk of getting overly trained i.e. overfitted for that particular data set. This means that when using the model on another data set it might not produce as accurate results. Splitting up the data set into two or three sub sets; training, test (and validation), we can use the test data to measure the performance of our obtained models, e.g. some polynomial regression by computing the error between true values and predicted values of the test data. The validation set is used to make sure the obtained model is not overfitted, where we keep fitting the model until a certain level of error is achieved. This is done by measuring the error of the prediction vs. the real value of the validation set.