

Applied Statistical Methods - Exercise 4

Peter von Rohr

2022-03-16

Problem 1: Overfitting

Use the extended dataset on **Body Weight** of animals and fit all the variables and the factor **breed**. Compare the result with a regression that uses only **Breast Circumference** or with the linear model that only uses the factor **Breed**. The data set is available from: https://charlotte-ngs.github.io/asmss2022/data/asm_bw_flem.csv

Problem 2: Plotting

The first step before doing any analysis should always be to plot the data which helps to visualise the internal structure of a dataset. A very instructive plot is the so-called **pairs-plot**. This plot can be done using the function `pairs()`. The task of this problem is to create a **pairs-plot** for the extended dataset on **Body Weight** of animals. The input to the function `pairs()` must be all numeric. This means that the column containing the **Breed** in our dataset must be converted to a datatype called **factor**. This can be done using the function `as.factor()`.

Results of linear models can also be plotted. In such plots, we are mainly interested in the behavior of the residuals. Hence, fit a linear regression model between **Body Weight** and **Breast Circumference** and plot the resulting linear model object.