Exercise 7

Advanced Methods for Regression and Classification

December 6, 2018

1. See last exercise: use `smooth.splines()` to fit smoothing splines

   (a) Fit a model to the training data; select the degrees of freedom (`df`) by internal cross-validation (`cv=TRUE`). Visualize the fitted values in the plot of the data. Predict the response for the test data, compute the resulting MSE, and compare with the results from last exercise.

   (b) Do the same as in (a), but provide values for `df`, e.g. in a grid from 2 to 20. Visualize the resulting MSE's (again based on the test data), and look if the "optimal" value for `df` from (a) (using the training data) is reasonable according to this evaluation.

2. At the TUWEL course you can find the data set *starsdata.csv*. Load these data with `read.csv()`. The data originate from a Hertzsprung-Russell diagram (see, e.g. Wikipedia), and they relate to the surface temperature of stars and their light intensity. Use for the following tasks the light intensity (`light`) as response variable, and the temperature (`temp`) as explanatory variable. Plot the data.

   (a) Compute smoothing splines (for the complete data) using internal cross-validation for the `df`. Visualize the fit in the plot.

   (b) Fit a linear model on a "reasonable" number of natural cubic splines (`ns()` from package `splines`), and visualize the fit in the plot.

   (c) Similar as 2.(c), but use polynomial regression instead. This means that you need to use `poly(x,3)` instead of `ns()`.

   (d) Fit a classical linear model, and visualize the fitted regression line.

   (e) Fit a robust (against data outliers) linear model, using the function `lmrob()` from the package `robustbase`, and visualize the fitted regression line.

   What do we learn from this exercise?

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_7.R* (no word document, no plots). Submit this file to Exercise 7 of our tuwel course (deadline December 5).