Stochastic Methods for Material Science (Winter term 2022)

Programming Project – Part 2

(Deadline for submission: February 12th 2023 at 11:59 pm)

Notice: You have to submit an R script including comments and the commands for solving the subsequent tasks. The evaluation and interpretation of the results and data analysis in R can be done via comments within the R script.

1. **Task**
   In the data set *timber.txt*, the experimentally determined values of the stiffness modulus (*Rigid*), elastic modulus (*Elast*) and density (*Dens*) of air-dried wood are available for 49 samples of different wood species.

   a) Investigate which of the three variables has the highest linear relationship to the two other ones, i.e., which of the following three models is most suitable for explaining data set *timber.txt*:

   $$\begin{aligned} Rigid &= \theta_0 + \theta_1\, Elast + \theta_2\ Dens \\ Elast &= \theta_0 + \theta_1\, Rigid + \theta_2\ Dens \\ Dens &= \theta_0 + \theta_1\, Rigid + \theta_2\ Elast \end{aligned}$$

   Consider exclusively the most suitable model in the following.

   b) Using a residual analysis, check whether there are any abnormalities in the regressed model chosen in a). Are there indications that the usual model assumptions are not fulfilled or are there influential points in the data set? If so, remove the outliers and influential data points and rerun the regression including a new residual analysis.

   c) Investigate whether the model is significantly improved by adding an interaction term between the two explanatory variables.

   d) A 50th wood sample was measured and the following values were observed:

   Rigid: 2078, *Elast:* 237.5, emphDens: 70.8.

   Does the new data point fit the regressed model, i.e., is it in the appropriate 95% range?

2. **Task**
   A noisy signal curve is given as output of a spectroscopic analysis in the file *signal.txt*. A Lorentz model curve is to be fitted to this curve,

   $$L(x; a, x_0) = \frac{a}{1 + (x - x_0)^2},$$

   with parameters $a, x_0 \in \mathbb{R}$.

   a) Determine the parameters by nonlinear regression. Choose appropriate initial values. Graphically output the estimated signal curve and data points.

   b) Examine whether the maximum signal strength (amplitude) of the estimated signal curve is significantly above the critical value $L_c = 3$.