Exercises “Methoden der Datenanalyse”

Chapter 3: Advanced Regression Modeling

General Remarks

- **Provide a PDF with an analytical report (Text and Figures!), which also contains the most important R outputs and your conclusions from it.**
- **Perform the analysis with R (give the code used separately from the report)**
- **Solutions may be worked out in groups of up to 3 students.**

- Use one of the following datasets and develop various predictive models for the continuous target variable.

  Your final report should contain descriptive statistics and exploratory data visualizations.

  Apply the whole range of regression-methods we discussed in the classroom. Try also to build derived predictors (grouping of continuous predictors, ratios of raw data).

  Apply data splitting or cross-validation techniques to decide for the best model which you would propose.

  Check also for outliers within the data and appropriateness of underlying modelling assumptions.

  Data Sets:

Name of Data Set	**Library**	**Dependent Variable**
College	ISLR	Apps
Boston	MASS	medv
Property Values	XLSX on Moodle	taxable_value
Bike Sharing Dataset Data Set	http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset	cnt
Online News Popularity	http://archive.ics.uci.edu/ml/datasets/Online+News+Popularity	Shares
Marketing Dataset	R-library: ElemStatLearn	Income

Chapter 4: Advanced Classification Modeling

General Remarks

- **Provide a PDF with an analytical report (Text and Figures!), which also contains the most important R outputs and your conclusions from it.**
- **Perform the analysis with R (give the code used separately from the report)**
- **Solutions may be worked out in groups of up to 3 students.**

- Use one of the following datasets and develop various predictive classification models for the categorical target variable.

  Your final report should contain descriptive statistics and exploratory data visualizations.

  Apply the whole range of classification-methods we discussed in the classroom. Try also to build derived predictors (grouping of continuous predictors, ratios of raw data).

  Apply data splitting or cross-validation techniques to decide for the best model which you would propose.

  Check also for outliers within the data and appropriateness of underlying modelling assumptions.

  Data Sets:

Name of Data Set	**Library**	**Dependent Variable**
Insurance Company Benchmark	http://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+%28COIL+2000%29	CARAVAN
Wine Quality	http://archive.ics.uci.edu/ml/datasets/Wine+Quality	A grouping of Quality Score
SPAM dataset	R-library: kernlab or ElemStatLearn Also: Spambasedata.txt on Moodle	Last Column of Data Frame
Credit Card Default	R-library: ISLR	default
Handwritten Digit Recognition Data	zip.test, zip.train R-library: ElemStatLearn	digit
Bank Marketing Data Set	http://archive.ics.uci.edu/ml/datasets/Bank+Marketing	Y client subscribed a term deposit