Exercise 9

Advanced Methods for Regression and Classification

December 13, 2018

1. *Classification trees*: function `rpart()` from the R package `rpart`

   Take the bank data set (see previous exercises). for classification trees. The goal is to predict if the client will subscribe a term deposit or not. This information is represented by the binary variable y (last one).

   (a) Select randomly a training set of a reasonable size and apply a tree T_0 (see `help(rpart)` or lecture notes).

   (b) Visualize the tree with the function `plot()` and `text()`, and interpret the results.

   (c) Predict the group membership for the test set (see `help(predict.rpart)` or lecture notes). How high is the resulting misclassification rate?

   (d) Show and interpret results of cross-validation obtained by using `printcp()` und `plotcp()`. What is the optimal complexity?

   (e) Prune the tree T_0 of the optimal complexity using `prune()`. Visualize und interpret the results.

   (f) Predict the group membership for the test set and calculate the resulting misclassification rate. Do we observe any improvement?

   (g) As in previous exercises, we see that the misclassification of the "yes" clients is very high. Use undersampling to reduce this error rate. Does the tree change?

   (h) Repeat 1.(g) hundred times. For simplicity, take every time the same complexity parameter as in 1.(g). This gives for each test set observation 100 predictions. Take the "majority vote" for the final group assignment. Did the error rate for the "yes" group improve?

2. Run the following code:

```
paste(intToUtf8(acos(log(1))*180/pi-13),
toupper(substr(month.name[2],2,2)),
paste(rep(intToUtf8(acos(exp(0)/2)*180/pi+2^4+3*2),2), collapse = intToUtf8(0)),
LETTERS[5^(3-1)], intToUtf8(atan(1/sqrt(3))*180/pi+2),
toupper(substr(month.abb[10],2,2)),
intToUtf8(acos(log(1))*180/pi-(2*3^2)),
toupper(substr(month.name[4],3,4)),
intToUtf8(acos(exp(0)/2)*180/pi+2^4+3*2+1),
intToUtf8(acos(exp(0)/2)*180/pi+2^4+2*4),
intToUtf8(acos(log(1))*180/pi-13),
LETTERS[median(0:2)],
intToUtf8(atan(1/sqrt(3))*180/pi*3-7),
sep = intToUtf8(0)
)
```

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_9.R* (no word document, no plots). Submit this file to Exercise 9 of our tuwel course (deadline December 19).