

Exercise 5

Advanced Methods for Regression and Classification

November 22, 2018

Load the data OJ from the package ISLR. Our goal is to use logistic regression to predict the grouping variable `Purchase`.

Use the function `glm(...,family=binomial)` in the following for logistic regression.

1. *Univariate case:* Use only the variable `LoyalCH` to predict the class of the response, and all available observations.
 - (a) Plot the predictor variable versus the response variable. Extract the estimated coefficients from the model, and show the regression line on the scale of the linear predictor.
 - (b) Plot the predictor variable versus the response variable. Extract the estimated coefficients from the model, and show the logit regression function (predicted probabilities on logit scale).
2. *Multivariate case:* Use all available variables in the data frame to predict the class of the response. Prepare the categorical explanatory variables as appropriate binary variables. Select randomly a training set of about 2/3 of the observations, build the classification model, predict the group membership for the (remaining) test data and compute the misclassification rate.
 - (a) Which variables have significant contribution to the classification problem (`summary` of result object)? What is the misclassification rate?
 - (b) Use stepwise variable selection (`step()`). Which variables have significant contribution to the classification problem? What is the misclassification rate?
 - (c) Compare both models using `anova(model1,model2,test="Chisq")`. What do you conclude?
3. Take the bank data set (see last exercise). Logistic regression will lead to a high misclassification rate of the “yes” clients, which is undesirable. How can you modify the method (approach) to reduce these misclassifications?

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer.5.R* (no word document, no plots). Submit this file to Exercise 5 of our tuwel course (deadline November 21).