

Lab 4

Statistical learning - EEBL

2023-10-19

Classification: logistic regression

How do you estimate a logistic regression model? (theory review)

Consider the *German credit* dataset, where the variable *Group* takes values 1 for credit-worthy and 2 for non credit-worthy individuals.

- how many individuals are worth of credit?
- after estimating the logistic regression model that includes all the variables in the dataset and more:
 - (i) what is the baseline category for the variable **StatusCAccount**?
 - (ii) interpret the coefficient of the variable **Duration**;
 - (iii) what is the residual deviance?
 - (iv) what is the in-sample misclassification rate of the full model?
- split the dataset into training and test sample and perform subset selection on the variables of the full model according to BIC:
 - (i) which variables are retained in the reduced model?
 - (ii) what is the residual deviance of the reduced model?
 - (iii) looking at the *true positive* rate on the test sample. Does the selected model improve with respect to the full one?