

Tutorial Business Analytics

Tutorial 4

Exercise 4.1

Note: Use R to solve this exercise(Exercise 4.1_R-template.R).

Load the training data ("*admit-train.csv*") into R. Proceed by typing *names(train)* to print the attribute names to the console. The attribute "*admit*" indicates whether a student has been admitted to a Master's Course. The independent attributes "*gre*" and "*gpa*" are numerical and contain the results of certain exams. The attribute "*rank*" is of ordinal scale and represents the reputation rank of the student's current university. The smaller the rank, the higher is the university's reputation. The functions *summary()*, *table()*, *sd()*, *hist()*, *plot()* etc. provide you with several statistics about the attributes.

Due to the fact that the dependent attribute "*admit*" is binary, you have decided to use a logistic regression model. Use below-mentioned commands to create a logit-model from the training data and to obtain the results.

```
mylogit = glm(admit~gre+gpa+as.factor(rank), data=train, family=binomial(link="logit"))
summary(mylogit)
```

- a) Which attributes are statistically significant regarding a significance level of 5%?
- b) Interpret the coefficients.
- c) Test the significance of the attribute "*rank*" using a Wald-Test. In order to do that, install the package *aod* (RStudio: Tools -> Install Packages or *install.packages("aod")*). Then enter the following commands:

```
library(aod)
```

```
wald.test(b=coef(mylogit), Sigma=vcov(mylogit), Terms=4:6)
```

- d) In order to gain a better understanding of the model, have a look at the predicted probabilities of some observations. Adjust only one parameter and keep the others constant. For example keep "*gre*" and "*gpa*" constant (using their mean/average) and vary "*rank*". Can you draw any conclusions? (You can use the *predict()* function to predict based on the model)
- e) Find the McFadden ratio and interpret the results.

```
McFadden <- 1 - (mylogit$deviance / mylogit$null.deviance)
```
- f) Load the data record "*admit-test.csv*" and predict the probability. Construct a confusion matrix.
- g) Find the logit model's error rate.

Exercise 4.2

You are provided the following numbers from the result of a Poisson Regression model.

Variable	Estimate	Std. Error
Intercept	1.5499	0.0503
age	-0.0047	0.0009

- a) According to the model above, what qualitative effect does a change in the independent variable *age* (+1) have on the dependent variable *dv*.
- b) According to the model above, what quantitative effect (on the incidence rate and log-incidence rate) does a change in the independent variable *age* (+1) have on the dependent variable *dv*.