TP4: linear model and generalizations

Logistic regression

Car purchase data

We first consider the data in file 'car_income.txt'. Each row corresponds to a family, with 3 columns:

- purchase: binary indicator of whether the family plans on buying a new car this year
- income: yearly family income (in thousands of dollars)
- age: age of current car

Load the data and explore them briefly. Construct a logistic regression and interpret the output. Can you think of transformations of the data to make the model better?

Bank notes

We now consider the data set "bank" from package gclus, on properties of genuine and counterfeit bank notes. You can access the data with the commands

```
# install.packages("gclus")
library(gclus)
data(bank)
```

and the command help(bank) gives a description of the columns.

Load the data and explore them briefly. Implement a logistic regression; the variable to explain is Status. How do you interpret the output? Build a plot which explains this situation.

Diabetes in Pima Indians

Load the data Pima.tr from package MASS. Explore the data briefly.

```
library(MASS)
data(Pima.tr)
```

Build a logistic regression and refine it until you reach a satisfactory model.

We now consider the new data set Pima.te. Perform prediction for these data, and compute your error rate. Repeat for other models.