Master BeNeFri in Computer Science

Course: Statistical Learning Methods
Spring 2016

Exercise #5: Multiple regression with R

1. In the folder Exercise#4, we have the file EducationBis.txt containing the data of Exercise #1 with modifications and without the outliers. Apply the `lm()` function to all the data and interpret the most important values you can find in the output you obtain with R.

2. Download from the ILIAS website the dataset `Computer` dataset (filename: ComputerData.txt) (see Exercise #4). The performance of the system (response) is indicated by the variable `PRP`. You must ignore the variable `ERP`. Remove the variables that cannot be viewed as good predictors. You can use all the variables you want to build a multiple regression model. Which variables do you use? Does your model explain something? Explain the model building strategy you have applied. Interpret the most important values of the final model you obtain with R.

3. Visualize graphically the relationship found by you model (by considering the the most important variable). Do you see outlier(s) or hard observations to predict with your model?

4. Same questions (#2 and #3) with the Cars2Data.txt (see Exercise #4). The performance of the system (response) is indicated by the variable `mpg` (miles per gallon).