Data Mining

Teacher: Annamaria Guolo

Practical assessment: September 6, 2018

INSTRUCTIONS: The examination takes 2 hours and 30 minutes. Write a brief report explaining the data analysis performed and the results. Provide a printed copy of the report. Remember to write your i) name and surname, ii) enrolment number, iii) date. Reports without the last three information will not be evaluated.

In the following you can find a description of the data and a brief guide for the analysis to be carried out. The use of the material of the course (slides, notes) is allowed. Internet is not allowed.

Dataset `houses`: data refer to sales prices of houses sold in one city of Canada

- `price`: Sale price of a house

- `lotsize`: lot size of a property in square feet

- `bedrooms`: Number of bedrooms

- `bathrooms`: Number of bathrooms

- `stories`: Number of stories excluding basement

- `driveway`: Does the house have a driveway? no/yes

- `recreation`: Does the house have a recreational room? no/yes

- `fullbase`: Does the house have a full finished basement? no/yes

- `gasheat`: Does the house use gas for hot water heating? no/yes

- `aircon`: Is there central air conditioning? no/yes

- `garage`: Number of garages

- `prefer`: Is the house located in the preferred neighborhood of the city? no/yes

The aim of the analysis is the evaluation of the association of the sales price to the characteristics of the house.

1. Consider the dataset composed by `price`, `lotsize`, `stories`, `prefer`, `aircon`. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from `R` and the graphical evaluation fo the model/models that are considered most useful in order to explain the analysis and the results.

2. Consider all the variables in the dataset. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from `R` and the graphical evaluation fo the model/models that are considered most useful in order to explain the analysis and the results.

If needed, please report the seed used in your analyses.