Statistical Data Mining Project

Due: April 27th, 2018

1 The Data

The data file **bank.csv**, which is available on Blackboard in the 'Project' folder, contains data related with the direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls with the aim of assessing if the client would subscribe to a bank term deposit or not.

The data file contains the following variables:

1. age
2. job: type of job ('admin.','blue-collar','entrepreneur','housemaid','management','retired', self-employed','services','student','technician','unemployed','unknown')
3. marital : marital status ('divorced','married','single')
4. education ('primary','secondary','tertiary','unknown')
5. default: has credit in default? ('no','yes')
6. balance: average yearly balance, in euros
7. housing: has housing loan? ('no','yes')
8. loan: has personal loan? ('no','yes') related with the last contact of the current campaign:
9. day: last contact day of the month

10. month: last contact month of year ('jan', 'feb', 'mar', , 'nov', 'dec')

11. campaign: number of contacts performed during this campaign and for this client

12. pdays: number of days that passed by after the client was last contacted from a previous campaign (-1 means client was not previously contacted)

13. previous: number of contacts performed before this campaign and for this client

14. poutcome: outcome of the previous marketing campaign ('failure','other','success', 'unknown')

15. Output variable (desired target): y - has the client subscribed a term deposit? ('yes','no')

2 Citation

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A DataDriven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:2231, June 2014

3 Question of Interest

Using this data can you accurately predict if a client will subscribe to a term deposit?

4 Report

Choose a classification approach and write a report answering the question above. When writing your report, you can assume that the reader has taken introductory statistics, so they are familiar with statistical concepts and terminology. The report should include the following:

1. **Title page** - This should include your project title, your name and your student number

2. **Abstract** - The abstract should concisely summarize the information in the report

3. **Introduction** - The introduction should provide any relevant background information to the reader. The introduction should also be used to introduce the data you are analyzing as well as the question(s) you hope to answer and why the question(s) is worth answering.

4. **Methods** - The methods section should describe your analysis approach.

5. **Results** - The results section should present the results of the analysis only, any discussion of the results should be deferred to the discussion section.

6. **Discussion** - The discussion should provide an interpretation of the results that you presented in the results section. The discussion should not discuss results that have not been previously mentioned.

7. **Conclusion** - The conclusion should summarize the important points made in the discussion. The conclusion should not present any new information and should only be based on points that have been previously discussed.

8. **References** - The references section should contain citations for any source that was referenced throughout the report (including R packages). You may use any reference style of your choosing.

9. **Appendix** - The appendix should contain your R code

The body of the report (i.e., excluding the title page, abstract, references and appendix) should not exceed 6 pages including tables and figures. Please submit the report document and the appendix as **a single PDF or word file**.