Práctica 2: Bank Direct Marketing

1 Files included
Bank.arff – Bank marketing campaign data

2 The assignment
The Bank of Gloucester in the UK wants to run a marketing campaign in order to attract new deposits. Given the current financial crisis there are huge pressures for banks to increase their financial assets, and at the same time to reduce their costs. In order to improve efficiency resulting in lesser contacts in marketing campaigns, the Bank of Gloucester is planning to use *directed marketing* campaigns for attracting new deposits.

The Bank of Gloucester has hired you as a consultant to give them advice about strategies to conduct the directed marketing campaign for their new deposits. They have provided you with data of past campaigns. These data were collected during 17 campaigns that occurred between May 2008 and November 2010 consisting of a total of 45211 contacts (for this práctica only 10% of the data is provided). The marketing campaign consisted of offering an attractive long-term deposit application with good interest rates to potential clients. For each contact, 16 attributes were generated (see Table bellow) together with the contact result, i.e. success or failure.

Name	Description and Values
Personal Client Information	
Age	Age at the contact date (Numeric ≥18)
Marital status	Married, single, divorced, widowed, separated (Nominal)
Job	Services, management, blue-collar, admin, technician, entrepreneur, student, housemaid, retired, unemployed, self-employed
Education	unknown, secondary, primary, tertiary (Nominal)
Bank Client Information	
Annual balance	Average yearly balance in *euros* (Numeric)
Credit	Yes or No (Nominal)
Personal loan	Yes or No (Nominal)
Housing loan	Yes or No (Nominal)
Last Contact Information	
Contact communication type	Unknown, telephone, cellular (Nominal)
Day	Last contact day of the month (Numeric)
Month	Last contact month of year (Nominal)
Duration	Of the contact in seconds (Numeric)
Other attributes	
Campaign	Number of contacts performed during this campaign and for this client (Numeric)
Prev days	Number of days that passed by after the client was last contacted from a previous campaign (Numeric, -1 means client was not previously contacted)
Previous	Number of contacts performed before this campaign and for this client (Numeric)
Poutcome	Outcome of the previous marketing campaign (Nominal: unknown, other, failure, success)

Your job is to provide the Bank of Gloucester with a predictive model for contact success/failure based on the attributes contained in the data set, and to analyze the resulting model in order to understand which factors are the most influential for the campaign success. In other words, by analyzing the model, you are expected to advise the Bank of Gloucester on which of the collected attributes are the most relevant inputs to influence success of the campaign. The bank will use this information for designing and implementing their directed marketing campaign.

3 Questions

1. What is the expected classification accuracy of the default classifier (one which chooses the most common class)?
2. Evaluate the accuracy (measured as correctly classified instances percentage) of the models produced by the classification algorithms you know (logistic regression, k-NN, and decision trees) using the training set and 10-fold cross validation. Explain the difference between the two accuracies.
3. Using 10-fold cross validation which of the classification algorithms you know has better accuracy?
4. What are the 5 attributes most influential attributes for the campaign success?
5. What is your prediction for two people with the following characteristics for Age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, poutcome:

   Person1: 66, retired, married, primary, no, 206, no, no, cellular, 9, feb, 479, 1, -1, 0, unknown
   Person2: 54, technician, married, tertiary, no, 876, no, no, cellular, 27, oct, 269, 3, 541, 3, success

Submitting your answer
The práctica can be solved in teams of two people (1 submission per team). Submission is through the Aula Global. Submissions should contain the code of the files you modified and the plots your programs generated. Deadline is the beginning of the next práctica. Late submissions will be penalized.