

Fuzzy data analysis, practical assignment: Classification of bank marketing data set

Practical assignment is classification task where purpose is to predict whether product would be subscribed or not with direct marketing campaign. Data is given in .csv file (or two where smaller one is 10 % of larger one and it can be used to e.g. test model). More information about the variables can be find in bank-names.txt file. Purpose is to get as good results with the data set as possible by using at least one method from the fuzzy data analysis course. Idea here is to try to mimic 'real world' situation where you (as data analyst) need to make decisions on how to preprocess the data first, which preprocessing tools to use and try, what methods to apply, how to evaluate your model(s), how to possibly change your model after reviewing results, etc. Finally after you are happy with your model you need to write a report from what you did. In the report present also those results which you have tried, but which have not necessarily improved the performance measure values which you have chosen. Notice that already in preprocessing the data into numerical form you make several decisions, which effect the outcome. E.g. if you process attribute no 3 {m,d,s} (m=married,d=divorced,s=single) values into (1,2,3) or whether you make three variables from this one: 3a:m present or not by {1,0}, 3b: d:{1,0} and 3c: s:{1,0}. In first choice attribute values may not follow ordering $1 < 2 < 3$ where as in second choice you end up making three variables from which all may not be important and you may end up creating 'noise' variables (-> in such case one could e.g. apply feature selection methods to find most important ones, i.e. as done in exercise 5). Another example from decisions needed to make already at the first stage is how to deal with missing values, do you i.e. remove those samples or try to apply imputation methods (see i.e. knnimpute from matlab).

In the report there should at least be, cover page, introduction, explanation of how you have preprocessed the data, explanation of the methods applied, results and conclusions. Sent your work to pasi.luukka@lut.fi (practical assignment report and the matlab codes which you used). Also be prepared to explain what you have done if asked. Deadline for returning practical assignment is 3.3.2019.